Single Sensor Speech Enhancement for Cyclo-Stationary Noise Employing Discrete Modulation Transforms

Version 2.2

Omry Sendik & Avi Shoham

Table Of Contents

Figure List

Table List

Version History

Version Number	Date (YYYY/MM/DD)	Creator	Description
0.1	2008/09/13	Omry Sendik	Preliminary Draft
1.1	2008/10/01	Omry Sendik	First Version with an Inclusive Theory Chapter
1.2	2008/11/01	Omry Sendik	Begun Documenting the DMT Performance Evaluation
1.3	2009/03/28	Omry Sendik	Documentation
1.4	2009/04/15	Omry Sendik	Documentation
1.5	2009/04/15	Avi Shoham	Documentation
1.6	2009/06/15	Avi Shoham	Fixing Notations and Documentation
1.7	2009/06/22	Omry Sendik	Documentation
1.8	2009/06/28	Avi Shoham	Documentation
1.9	2009/06/30	Omry Sendik	Documentation
2.0	2009/06/30	Omry Sendik	First comprehensive version
2.1	2009/09/08	Omry Sendik	Merged comments
2.2	2009/09/08	Avi Shoham	Merged comments

Glossary

Term	Abbreviation/Explanation
LTI	Linear Time Invariant
CT	Continuous Time
DT	Discrete Time
CF	Continuous Frequency
DF	Discrete Frequency
DTFT	Discrete Time Fourier Transform (Continuous Frequency)
IDTFT	Inverse Discrete Time Fourier Transform (Continuous Frequency)
STFT	Short Time Fourier Transform
ISTFT	Inverse Short Time Fourier Transform
DTSTFT	Discrete Time Short Time Fourier Transform
IDTSTFT	Inverse Discrete Time Short Time Fourier Transform
CTMT	Continuous Time Modulation Transform
ICTMT	Inverse Continuous Time Modulation Transform
IDTMT	Inverse Discrete Time Modulation Transform
AM	Amplitude Modulation
FM	Frequency Modulation
Phoneme	The smallest language unit that distinguishes meaning
Formant	A peak in the frequency spectrum of a sound caused by acoustic resonance
LPF	Low Pass Filter
BPF	Band Pass Filter
BW	Band Width
OLA	Overlap Addition
FBS	Filter Bank Summation
LPF	Low Pass Filter
AWGN	Additive White Gaussian Noise
CSN	Cyclo-Stationary Noise
JF	Joint Frequency
MLE	Maximum Likelihood Estimator
CSN	Cyclo-Stationary Noise

1 Preface

A classical problem in signal processing is the problem of finding the ideal way to separate a certain signal from a mixture (an observation) that includes it and other signals. This type of problem can wear different forms, depending on the types of signals and observations involved. This problem can be described using the "Cocktail Party Effect". The "Cocktail Party Effect" can be thought of as two related, but different, problems. The primary problem of interest has traditionally been that of recognition. Hence, how do humans handle the task of segmenting speech? An immediate outcome question is whether it is feasible to imitate the human ears' ability to do so using a computer. The second problem, which is the problem this work deals with, is the development of a "toolbox" which will enhance the ability of a listener to separate mixed audio signals, or "clean" a certain signal from a mixture that contains it.

A variety of applications drive this subject of research in humans as well as in computer implementations. Medical applications, such as the ability to diagnose, in an early stage, pathologies in fetuses, observe difficulties due to the fact that the EKG sensor contains a mixture of heartbeat signals. Moreover, speech discrimination using hearing aids becomes difficult when multiple speakers constitute the input signal to the device. These are only several examples, many other examples exist.

The study of separating mixed sources observed in an array of sensors has been a classical and difficult signal processing problem. This phenomenon is still very much a subject of research (where it is typically referred to as source separation or blind source separation). The ability to separate mixed sources observed in a single source is of a few orders of magnitude more complex.

A problem which is very similar to source separation, which we will focus on throughout this work, is the problem of speech enhancement. The idea of the problem is quite similar – how to best recover a speech signal, out of a mixture that includes it and other (unwanted) signals. Figure 1 illustrates the problem.

This paper suggests solutions for the specific problem of enhancing a single speech signal out of a single mixture (or observation) where it is interfered by Cyclo-Stationary Noise (an exact mathematical definition of this type of noise will be given later). The motivation for dealing particularly with Cyclo-Stationary noises is the fact that it is a common mathematical model for periodic noises, thus being useful when dealing with engine and other electronic noises. This also suggests that successful enhancement will have many useful applications (some of which will be mentioned later). It should be noted that enhancement algorithms, that attempt to solve this problem, already exist. As two examples we can mention the algorithms of OM-LSA and spectral subtraction ([3] -[4] -). However, all of these algorithms are not necessarily optimal when cyclo-stationary noise is assumed, because they assume that the noise is quasi-stationary (which is untrue in our case), and because they do not try to take advantage of the AM properties that the Cyclo-Stationary noise possesses. We will suggest, throughout this paper, an alternative approach, that will use these special properties for the enhancement process. This approach will involve the representation of all signals in a new domain, named the “Joint Frequency” domain [8] -.

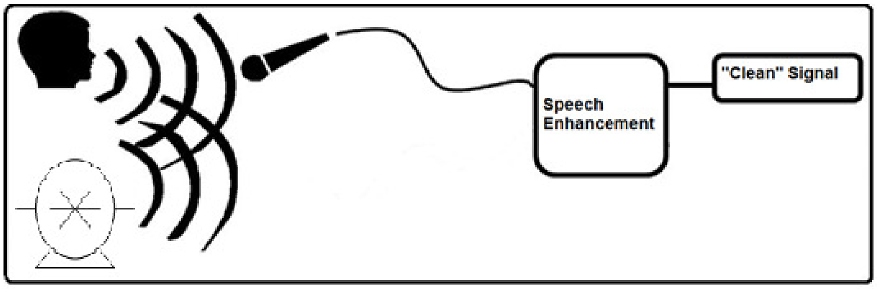


Figure 1 – Speech Enhancement Illustration

This paper is organized as follows. In chapter 2 we describe the mathematical background needed in order to understand the proposed solutions. In chapter 3 we depict the performance and achievements of the theoretical tools developed in chapter 2. Finally in chapter 4 we present a comparative view of the practice held out in chapter 3 and suggest our conclusive viewpoint.

2 Theoretical Background

The Project's Notations

Digital Signal Processing (DSP) involves usage of a variety of "mathematical tools". Different authors tend to use different notations which may cause the absence of clarity. We hereby define the notations we will be using throughout this paper.

When dealing with transformations whose domains are of entirely continuous regimes (CT, CF) the naming for the transformation will not point out the regimes. e.g. – the explicit naming for the Fourier Transform could have been the "CT CF Fourier Transform".

When dealing with transformations whose domains are of entirely discrete regimes (DT, DF) the naming for the transformation will include a 'D' prefix. e.g. - the DFT.

Transformations of different types will always comprise of DT and CF regimes. Regimes of CT and DF are of no interest to the scope of this paper.

Operator Name	Analytical Term
The Fourier Transform	$X(\omega)=\mathfrak{F}\{x(t)\}\triangleq\int_{-\infty}^{\infty}x(t)e^{-j\omega t}dt$
The Inverse Fourier Transform	$x(t)=IF\{X(\omega)\}\triangleq\frac{1}{2\pi}\int_{-\infty}^{\infty}X(\omega)e^{j\omega t}d\omega$
The DT Fourier Transform	$X(\omega)=DTFT\{x(n)\}\triangleq\sum_{n=-\infty}^{\infty}x(n)e^{-j\omega n}$
The DT Inverse Fourier Transform	$x(n)=IDTFT\{X(\omega)\}\triangleq\frac{1}{2\pi}\int_{-\pi}^{\pi}X(\omega)e^{j\omega n}d\omega$
The Discrete Fourier Transform	$X(k)=DFT\{x(n)\}\triangleq\sum_{n}x(n)e^{-j2\pi nk/K}$
The Inverse Discrete Fourier Transform	$x(n)=IDFT\{X(k)\}\triangleq\frac{1}{N}\sum_{k}X(k)e^{j2\pi nk/K}$
The Short Time Fourier Transform	$X(t,\omega)=STFT\{x(t)\}\triangleq\int_{-\infty}^{\infty}w(t-\tau)x(\tau)e^{-j\omega\tau}d\tau$
The Inverse Short Time Fourier Transform	$x(t)=ISTFT\{X(t,\omega)\}\triangleq\frac{1}{2\pi w(0)}\int_{-\pi}^{\pi}x(t,\omega)e^{j\omega\tau}d\omega$
The Discrete Short Time Fourier Transform	$X(n,k)=DSTFT\{x(n)\}\triangleq\sum_{m=1}^{K}x(m)w(n-m)e^{-j2\pi mk/K}$
The Discrete Inverse Short Time Fourier Transform	$x(n)=IDSTFT\{X(n,k)\}\triangleq\frac{1}{Nw(0)}\sum_{k=1}^{K}X(n,k)e^{j2\pi nk/K}$

The DT Short Time Fourier Transform	$X\left(n,\omega\right)=DTSTFT\left\{x\left(n\right)\right\}\triangleq\sum_{m=-\infty}^{\infty}w\left(n-m\right)x\left(m\right)e^{-j\omega m}$
The DT Inverse Short Time Fourier Transform	$x\left(n\right)=IDTSTFT\left\{X\left(n,\omega\right)\right\}\triangleq\frac{1}{2\pi w(0)}\int_{-\pi}^{\pi}X\left(n,\omega\right)e^{j\omega n}d\omega$
The Envelope Detector	$m\left(t\right)\triangleq\mathfrak{D}\left\{x\left(t\right)\right\}$
The Discrete Envelope Detector	$m\left(n\right)\triangleq\mathfrak{D}\left\{x\left(n\right)\right\}$
The Complementary Detector	$c\left(t\right)\triangleq\mathfrak{D}^{c}\left\{x\left(t\right)\right\}$
The Discrete Complementary Detector	$c\left(n\right)\triangleq\mathfrak{D}^{c}\left\{x\left(n\right)\right\}$
The Modulation Transform	$X\left(H,w\right)=\mathrm{M}\left\{x\left(t\right)\right\}\triangleq\mathrm{F}\left\{\mathrm{D}\left\{STFT\left\{x\left(t\right)\right\}\right\}\right\}$ $=\int\mathrm{D}\left\{\int x\left(t\right)w\left(\tau-t\right)e^{-j\omega t}dt\right\}e^{-jH\tau}d\tau$
The Inverse Modulation Transform	$x\left(t\right)=\mathfrak{M}^{-1}\left\{X\left(H,\omega\right)\right\}\triangleq ISTFT\left\{\mathfrak{F}^{-1}\left\{X\left(H,\omega\right)C\left(\tau,\omega\right)\right\}\right\}$ $=\frac{1}{2\pi}\int\int\left[\frac{1}{2\pi}\int X\left(H,\omega\right)e^{jH\tau}dH\right]C\left(\tau,\omega\right)e^{j\omega t}d\tau d\omega$
The DT Modulation Transform	$X\left(\eta,\omega\right)=DTMT\left\{x\left(t\right)\right\}\triangleq DTFT\left\{\mathrm{D}\left\{DTSTFT\left\{x\left(n\right)\right\}\right\}\right\}$ $=\sum_{m}\mathrm{D}\left\{\sum_{n}x\left(n\right)w\left(m-n\right)e^{-j\omega n}\right\}e^{-j\eta m}$
The DT Inverse Modulation Transform	$x\left(n\right)=IDTMT\left\{X\left(\eta,\omega\right)\right\}\triangleq IDTSTFT\left\{IDTFT\left\{X\left(\eta,\omega\right)C\left(m,\omega\right)\right\}\right\}$ $=\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{m}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}X\left(\eta,\omega\right)e^{j\eta m}d\eta\right]C\left(m,\omega\right)e^{j\omega n}d\omega$
The Short Time Modulation Transform	$X\left(T,H,\omega\right)=STMT\left\{x\left(t\right)\right\}\triangleq STFT\left\{\mathrm{D}\left\{STFT\left\{x\left(t\right)\right\}\right\}\right\}$ $=\int\mathrm{D}\left\{\int x\left(t\right)w\left(\tau-t\right)e^{-j\omega t}dt\right\}v\left(T-\tau\right)e^{-jH\tau}d\tau$
The Inverse Short Time Modulation Transform	$x\left(t\right)=ISTMT\left\{X\left(T,H,\omega\right)\right\}\triangleq ISTFT\left\{ISTFT\left\{X\left(T,H,\omega\right)\right\}C\left(\tau,\omega\right)\right\}$ $=\frac{1}{2\pi}\int\int\left[\frac{1}{2\pi}\int X\left(T,H,\omega\right)e^{jH\tau}dTdH\right]C\left(\tau,\omega\right)e^{j\omega t}d\tau d\omega$
The DT Short Time Modulation Transform	$X\left(l,\eta,\omega\right)=DTSTMT\left\{x\left(t\right)\right\}\triangleq DTSTFT\left\{\mathfrak{D}\left\{DTSTFT\left\{x\left(n\right)\right\}\right\}\right\}$ $=\sum_{m}\mathfrak{D}\left\{\sum_{n}x\left(n\right)w\left(m-n\right)e^{-j\omega n}\right\}v\left(l-m\right)e^{-j\eta m}$
The DT Inverse Short Time Modulation Transform	$x\left(n\right)=IDTSTMT\left\{X\left(l,\eta,\omega\right)\right\}\triangleq IDTSTFT\left\{IDTSTFT\left\{X\left(l,\eta,\omega\right)\right\}C\left(m,\omega\right)\right\}$ $=\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{m}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{l}X\left(l,\eta,\omega\right)e^{j\eta m}d\eta\right]C\left(m,\omega\right)e^{j\omega n}d\omega$

The Discrete Modulation Transform	$X\left(i,k\right)=DMT\left\{x\left(n\right)\right\}\triangleq DFT\left\{\mathfrak{D}\left\{DSTFT\left\{x\left(n\right)\right\}\right\}\right\}$ $=\sum_{m}\mathfrak{D}\left\{\sum_{n}x\left(n\right)w\left(m-n\right)e^{-j2\pi nk/K}\right\}e^{-j2\pi mi/I}$
The Discrete Short Time Modulation Transform	$X\left(l,i,k\right)=DSTMT\left\{x\left(n\right)\right\}\triangleq DSTFT\left\{\mathfrak{D}\left\{DSTFT\left\{x\left(n\right)\right\}\right\}\right\}$ $=\sum_{m}\mathfrak{D}\left\{\sum_{n}x\left(n\right)w\left(m-n\right)e^{-j2\pi nk/K}\right\}v\left(l-m\right)e^{-j2\pi mi/I}$

Table 1 - Paper Notations

Table 2 shows the various units of variables of various regimes.

Domain	Definition Region
CT Domain	$t\left[\sec\right]$
DT Domain	$n\left[samples\right]$
CF Domain	$\omega\left[rad/\sec\right]$
DF Domain	$\omega\left[samples/\sec\right]$

Table 2 – Units of the various variables per regime

The STFT

Introduction and Motivational Background

The ubiquitous Fourier Transform has proven itself and its usefulness in the analysis and synthesis of various signals in the Time and Frequency domains. However, applying the Fourier Transform on a relatively long signal causes it to lose its ability to present the various properties of a signal in the other domain (Time/ Frequency). Applying a Fourier transform to an audio signal, which is a few milliseconds long, will yield a vector or function which presents the signal's present spectral content. If, for example, the signal was composed of two pure sines heard one after the other, the Fourier transform lacks the ability to easily indicate which sine was heard when. We may then consider the Short Time Fourier Transform, whose central target is aimed at gaining the ability to convey some of the signal's time properties, as well as its frequency properties, both at once. e.g. - one would like to have the ability to indicate which one of the two sines in the abovementioned audio signal was heard first. Figure 2 shows an audio signal which is composed of two sines played on a time axis, whereas Figure 3 shows the Fourier Transform of the audio stream. One can immediately notice the two spectral components appearing in Figure 3 indicating about the presence of the two sines. But what if we would like to know which of the spectral components was played first? We would then turn to a "tool" whose output is the plot presented in Figure 4. We can easily notice that in various parts of the plot we can distinguish between different spectral components. The presence of spectral components is indicated by the color strength of a horizontal row. The vertical axis indicates the current time stamp. This "tool" is called the STFT.

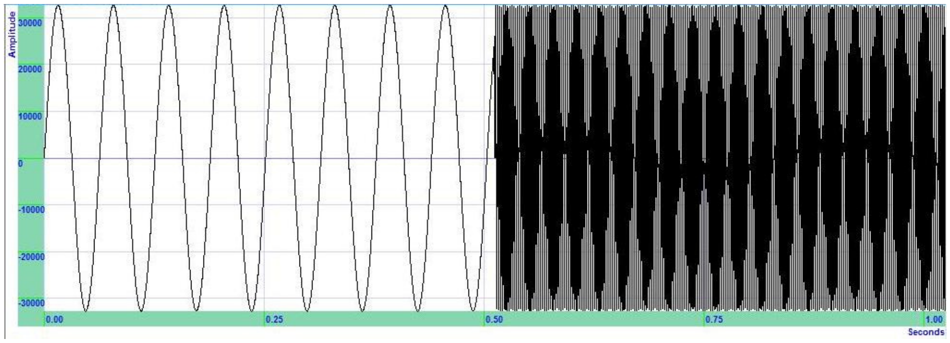


Figure 2 – Waveform of the an audio signal

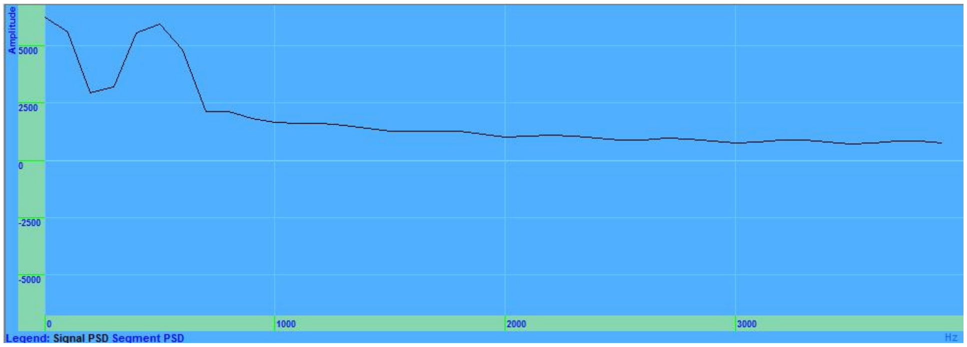


Figure 3 – Fourier Transform of the an audio signal

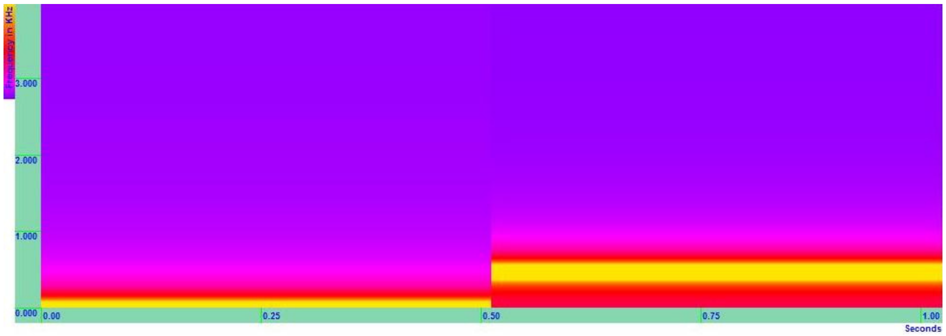


Figure 4 – Fourier Transform of an audio signal

Definition

As stated in section 2.2.1, the STFT is a two dimensional representation of a signal, the time and frequency dimensions. We define the STFT as –

$$X\left(n,\omega\right)=\sum_{m=-\infty}^{\infty}\underbrace{w\left(n-m\right)}_{Window}\underbrace{x\left(m\right)e^{-j\omega m}}_{Fourier\ Transform}$$

The right part of the expression under the summation is the ubiquitous DTFT. In order to enable the ability to present time properties while simultaneously presenting frequency properties we firstly multiply the signal by a "window" sequence which determines the portion of the input signal that will be transformed at a particular time index n.

The STFT is clearly a function of two variables, ω which is continuous, and the time index n which is discrete.

If we perform a variable transform of $n-m \to m'$ we get –

$$X(n,\omega)=\sum_{m'=-\infty}^{\infty} w(m')x(n-m')e^{-j\omega(n-m')}=e^{-j\omega n}\sum_{m'=-\infty}^{\infty} x(n-m')w(m')e^{j\omega m'} \Rightarrow$$

$$\tilde{X}(n,\omega)=\sum_{m'=-\infty}^{\infty} x(n-m')w(m')e^{j\omega m'}$$

Now by holding ω constant we notice that $\tilde{X}(n,\omega)$ is in the form of a convolution. Hence, we can consider the STFT to be a linear filtering system. Section 0 will deepen these insights into solid properties of the STFT.

By analogy to the comparison between the DFT and the DTFT we deduce that the DSTFT is given by

$$X(n,k)=DSTFT\{x(n)\}\triangleq\sum_{m} x(m)w(n-m)e^{-j2\pi mk/K}$$

Where, $k=0,\ldots,K-1$

Properties

Fourier Transform Interpretation

As stated in section 0 the STFT can be interpreted as a Fourier Transform of a signal multiplied by a window. The STFT is a function of the time index n which takes on all integer values so as to "slide" the window, $w(n-m)$, along the input signal $x(m)$. This is depicted in Figure 5 which shows an input signal and $w(n-m)$ as functions of m for various values of n (various locations of the window).

As in the case of discrete-time signals, the STFT is periodic in ω with a 2π period.

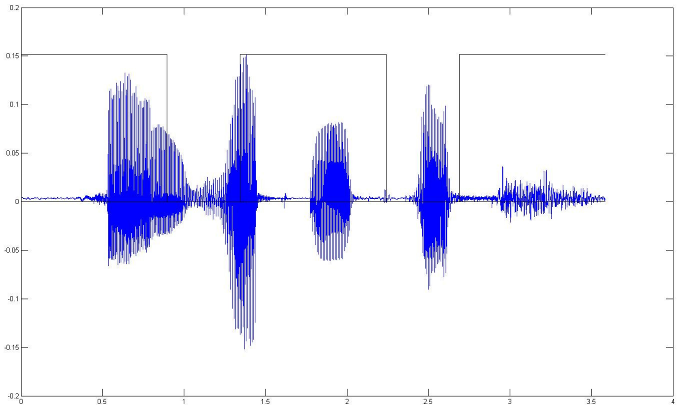


Figure 5 – Fourier Transform Interpretation of the STFT

The fact that for a given value of the index n we are handed with the "regular" Fourier Transform can be used in order to deduce that the ISTFT is given by –

$$w(n-m)x(m)=\frac{1}{2\pi}\int_{-\pi}^{\pi}X(n,\omega)e^{j\omega m}d\omega$$

Now, in order to obtain the section of signal which has undergone the STFT we set $n=m$ and by this get (assuming $w(0)\neq 0$) –

$$x(m)=\frac{1}{2\pi w(0)}\int_{-\pi}^{\pi}X(m,\omega)e^{j\omega m}d\omega$$

We now have the original segment which we have transformed.

Linear Filtering Interpretation

The equation for the STFT as presented above clearly exhibits the fact that for each value of ω, we can consider it to be a convolution of $w(n)$ with the sequence $x(n)e^{-j\omega n}$ -

$$X(n,\omega)=\sum_{m=-\infty}^{\infty}w(n-m)x(m)e^{-j\omega m}=\sum_{m'=-\infty}^{\infty}x(n-m')w(m')e^{-j\omega(n-m')}=x(m')e^{-j\omega m'}*w(m')$$

Hence, for a particular value of ω, the STFT can thought as the output of the system depicted in Figure 6.

We shall now derive, using the graphical representation of the STFT system in Figure 6, what is the interpretation of the STFT.

As the input to the system we have $X(\theta)$. After multiplying by $e^{-j\omega n}$ we yield, due to the modulation property, $X(\theta+\omega)$. And finally after transferring the signal through $W(\theta)$ we have as the systems' output $X(\theta+\omega)W(\theta)$. Hence we interpret the output as a filtered output whose transfer function is given by $W(\theta)$.

Another system displaying the STFT as an LTI system is given in Figure 7. This interpretation utilizes the relation between $X(n,\omega)$ and $\tilde{X}(n,\omega)$ as given in section 0. Utilizing this graphical representation we derive that $X(n,\omega)$ can also be thought of as the result of modulating $e^{-j\omega n}$ with output of a BPF whose impulse response is $w(n)e^{j\omega n}$. Hence, if the Fourier Transform of $w(n)$, $W(\omega)$, is low-pass like, then the output of the system will be a BPF whose central frequency is ω.

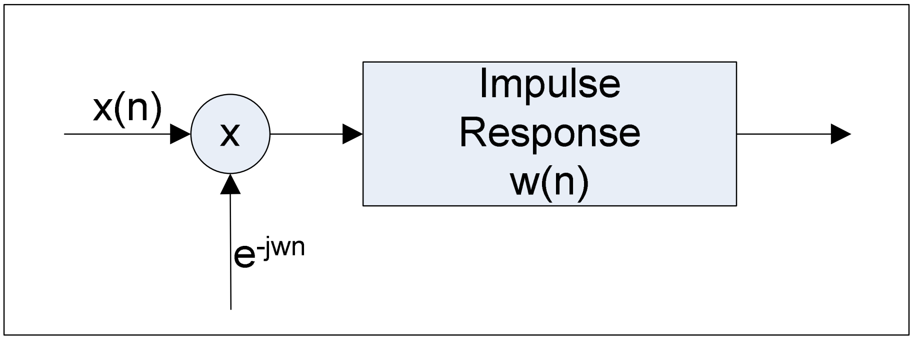


Figure 6 – First Linear Filtering Interpretation of the STFT

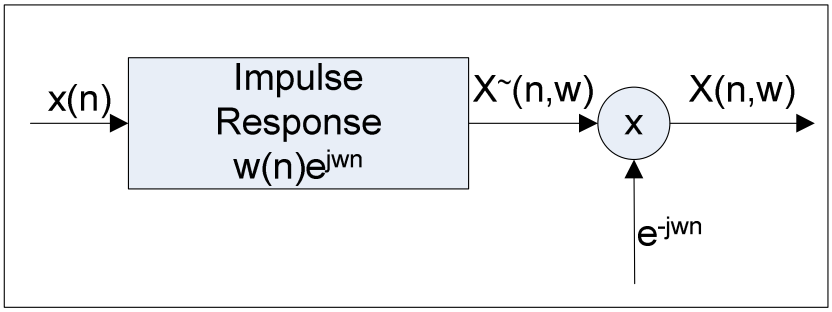


Figure 7 – Second Linear Filtering Interpretation of the STFT

Investigating the properties of the most popularly used windows we indeed reveal that they all possess LPF properties.

The interpretation of the STFT as filtering can be used for the signal's reconstruction as can be seen in section 0.

The Linear Filtering interpretation is often referred to as the Filter Bank Interpretation and so will frequently be done by this paper.

Sampling Requirements in Time and Frequency Domains

It should be lucid by now that the STFT is a function of both time and frequency. We shall now derive what are the sampling rates necessary in order to provide a credible STFT.

In order to derive the sampling requirements in the time domain we utilize the linear filtering interpretation of section 0, where it was shown that the output, for a fixed value of ω, could be shown to be the output of a filter with impulse response $w(n)$ and that most windows have LPF properties. By denoting the BW of $W(\omega)$ as B we immediately conclude that $X(n,\omega)$ has an identical BW. Now, according to the sampling theorem, $X(n,\omega)$must be sample at a rate of at least $2B$ samples/second in order to avoid aliasing.

Due to the fact that $X(n,\omega)$is periodic inω with a 2π period, a finite set of frequencies to be sampled can be determined and used in order to determine the Sampling Requirements in Frequency. We now utilize the Fourier Transform interpretation given in section 0. The Inverse Fourier Transform of $W(\omega)$ is clearly time limited. Hence, the Inverse Fourier Transform of $X(n,\omega)$ is also time limited and the sampling theorem requires that we sample it at a rate of at least its time interval. If the signal contains L samples then $X(n,\omega)$ must be sampled at the set of frequencies given by –

$$\omega_k = \frac{2\pi k}{L}, \quad k = [0, L-1]$$

In total, the overall sampling requirements state that total amount of samples needed in order to credibly reconstruct a discrete time, discrete frequency signal $X(n,\omega_k)$ is–

$$SR = 2B \bullet L \ \text{ samples/ sec}$$

The Fourier Uncertainty Principle and its implications on the STFT resolution

Choosing a window length has implications on the outcome as will be clearly depicted in the sections to come. Choosing a long window will enable high frequency resolution but will yield inferior time resolution. This is offcourse due to the fact that a DFT 's resolution is inversely proportional to the number of samples transformed while the time resolution is linearly proprtional to the number of windows which fit in the entire signals duration. e.g. – a signal with 1024 samples analysed with a windows length of 512 will give good frequency resolution but only two time samples. Meaning the time resolution will be the signal's entire time duration divided by two (which is, of course, a very poor time resolution). A $DSTFT$ using a long window is usually called a narrowband transform whereas a $DSTFT$ using a short windows is usually called a wideband transform.

This resolution trade-off can be depicted as a lower limit to the rectangle size in the signals spectrogram. This is depcited in Figure 8.

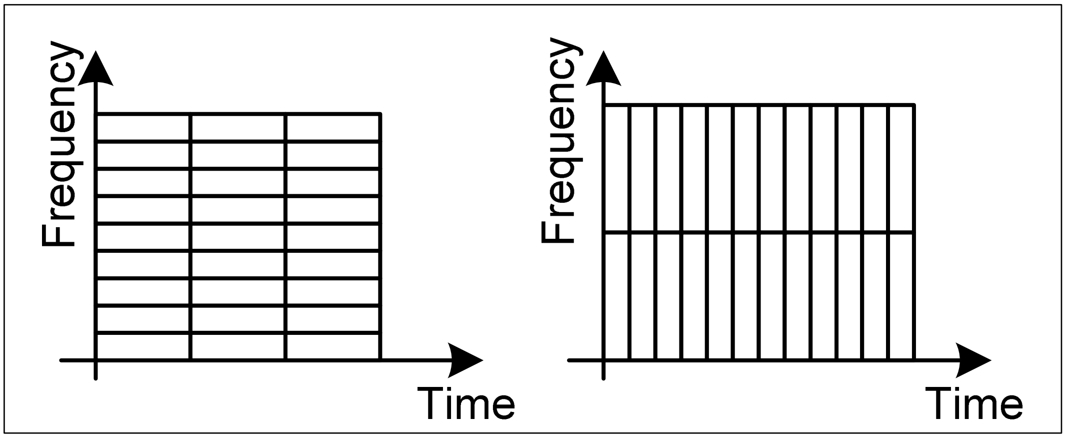


Figure 8 –Short Time Fourier Transform Resolution Trade-Off

Implications of the Window of choice

The usage of an analysis window for purposes of Spectral Analysis requires multiplying the signal in the time domain by the chosen window. Such an operation leaves us with the section in time we wish to analyze. In the frequency domain the consequences of applying an analysis window involves convolving the original signal's Fourier Transform with the Analysis Window's Fourier Transform.

A basic and easy way to understand the choice of window is of course the rectangular window. In the time domain the consequences are easily comprehensible. In the time domain we are left only with the section we wished to analyze. In the frequency domain we convolve our signal with the Fourier Transform of a rectangle. Let's assume we are working in DT, CF regimes (a pragmatic assumption). Hence, the Fourier Transform of the rectangle is the Dirichlet Kernel which may be written as $W_{rect} = \frac{1-e^{-jN\theta}}{1-e^{-j\theta}} = \frac{\sin(0.5N\theta)}{\sin(0.5\theta)} e^{-j0.5\theta(N-1)} = \underbrace{D(N,\theta)}_{Dirichlet\ Kernel} e^{-j0.5\theta(N-1)}$ and which is depicted in Figure 9.

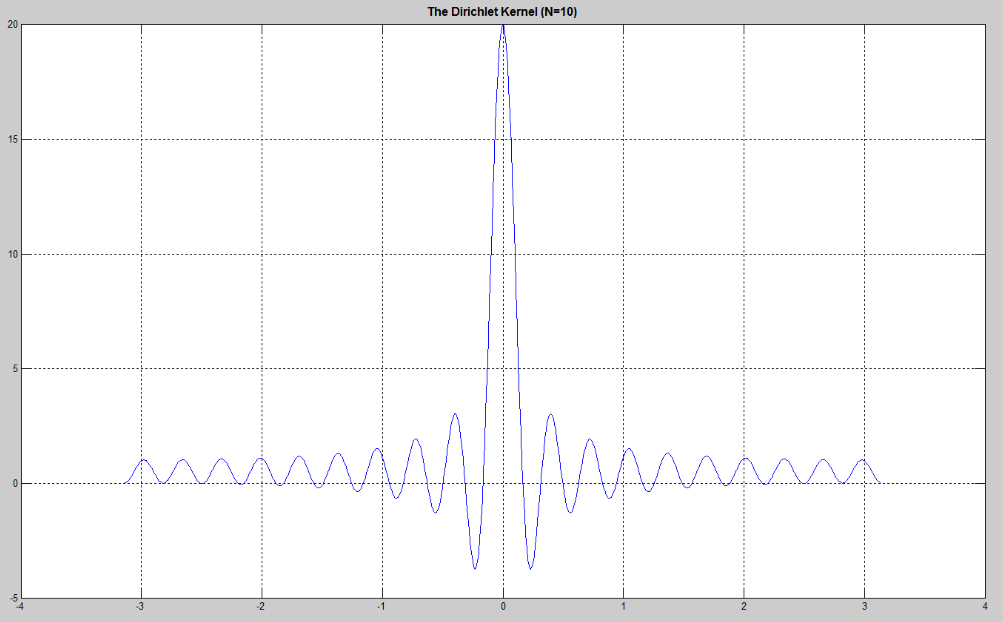


Figure 9 – Dirichlet Kernel with N=10

We immediately notice the appearance of a main lobe and multiple side lobes. Let us recall that the convolution of a signal with a delta function leaves us with the original signal. We also know that when $N \to \infty$ the Dirichlet Kernel converges to the delta function, thus we can deduce that convolving our signal with this window will cause distortions of some type.

The distortions are of two types –

- The loss of frequency resolution due to the finite width of the main lobe. It is obvious that any signal with impulse components, after being convolved with the main lobe, will have its energy spread over the main lobes entire range. Moreover, if two signals' distance one from another is smaller than the main lobes' width, they will turn indistinguishable.
- Due to the side lobes, partial energy from adjacent spectral components leaks into the currently analyzed spectral component.

Hence we want the main lobe to be of minimal width and the side lobes to be of maximal attenuation relative to the main lobe. In general these two demands are tradeoffs and the width of main lobe is inversely proportional to the length of the window. In order to minimize the side lobes we may make a use of "smarter" windows. The consequences of the usage of such windows are the reshaping, in the time domain, of the original signal section.

Table 3 summarizes some common windows properties and Figure 10 depicts their Fourier Transforms.

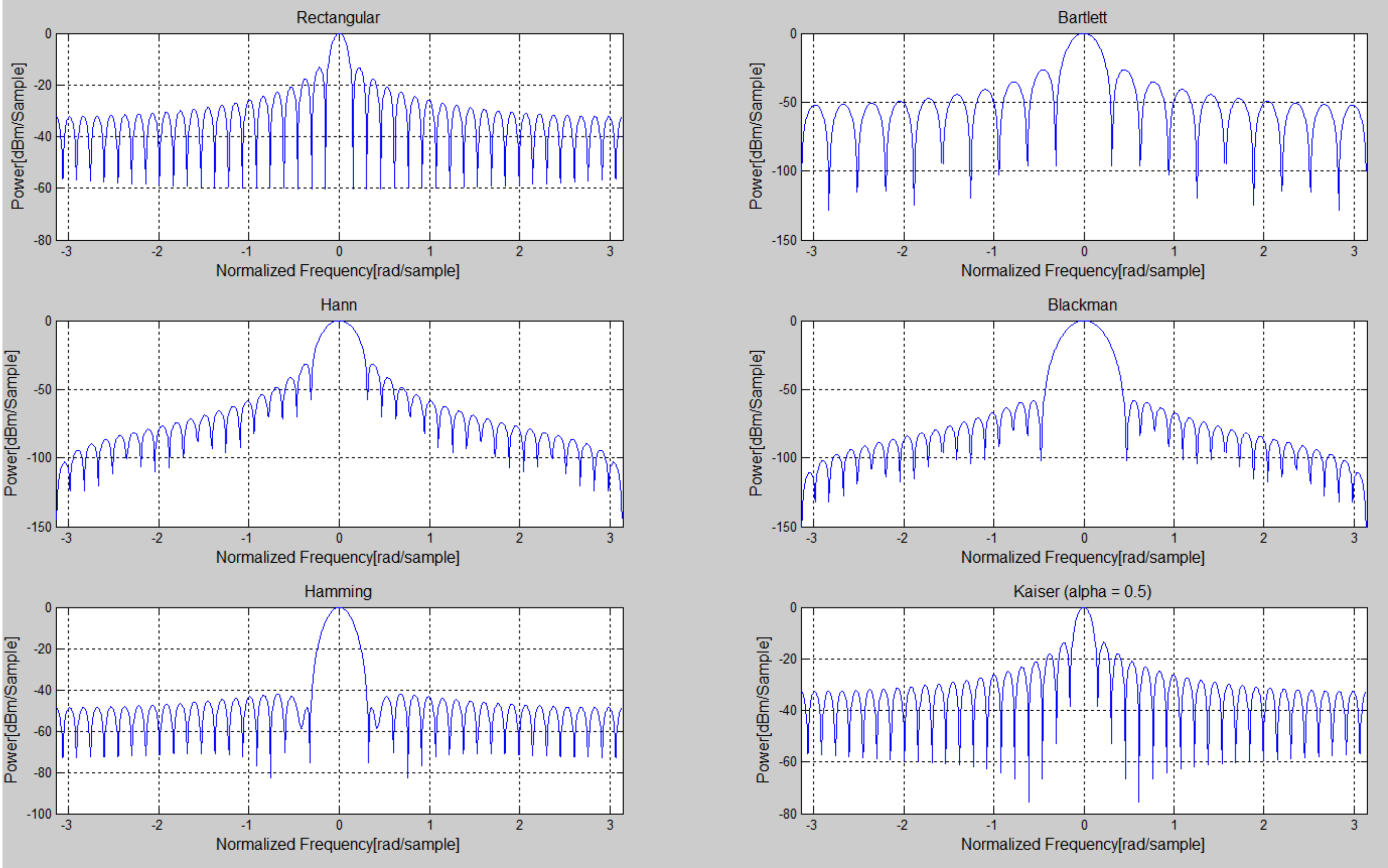


Figure 10 – The Various Analysis Windows

Window Type	Main-lobe width	Side-lobe level
Rectangular	$4\pi / N$	$-13.5dB$
Bartlett	$8\pi / N$	$-27dB$
Hann	$8\pi / N$	$-32dB$
Hamming	$8\pi / N$	$-43dB$
Blackman	$12\pi / N$	$-57dB$
Kaiser (parametric window)	Depends on α	

Table 3 – Various Analysis Windows Properties

Signal Reconstruction Methods

The Overlap Addition Method (OLA)

One method for reconstructing $x(n)$ from the STFT is based on the Fourier Transform Interpretation. As shown earlier $X(n,\omega)$ can be considered as the standard DTFT of the sequence –

$$y_n(m) = x(m)w(n-m)$$

Hence, by computing the Inverse Discrete Fourier Transform of $X(n,\omega_k)$ and eliminating the window multiplication by division. However this method is very susceptible to aliasing errors for reasons that will not be shown in this paper. We will hereby present a more robust synthesis method similar to the OLA method for a periodic convolution using DTFTs.

The Filter Bank Summation Method (FBS)

This reconstruction method is based on the Linear Filtering Interpretation. As stated before, we can consider $X(n,\omega_k)$ to be a low pass filtered signal centered at ω_k.

$$X(n,\omega_k) = \sum_{m=-\infty}^{\infty} w_k(n-m)x(m)e^{-j\omega_k m} = e^{-j\omega_k n}\sum_{m=-\infty}^{\infty} w_k(m)x(n-m)e^{j\omega_k m}$$

By defining $h_k(n) = w_k(n)e^{j\omega_k n}$ we can then express $X(n,\omega_k)$ as –

$$X(n,\omega_k) = e^{-j\omega_k n}\sum_{m=-\infty}^{\infty} h_k(m)x(n-m)$$

Now by defining $y_k(n) = X(n,\omega_k)e^{j\omega_k n}$ we see that –

$$y_k(n) = \sum_{m=-\infty}^{\infty} h_k(m)x(n-m)$$

This representation displays $y_k(n)$ as a band pass filtered (centered at ω_k) version of x by $h_k(n)$.

Now, supposing $X\left(n,\omega_k\right)$ is available at the frequencies $\left\{\omega_k\right\},\quad k=\left[0,N-1\right]$, we consider N BPFs given by $H_k\left(\omega\right)=W_k\left(\omega-\omega_k\right)$. The composite frequency response is given by –

$$\tilde{H}\left(\omega\right)=\sum_{k=0}^{N-1}W_k\left(e^{j\left(\omega-\omega_k\right)}\right)$$

If $W\left(e^{j\omega_k}\right)$ is properly sampled in frequency then it can be shown that –

$$\frac{1}{N}\sum_{k=0}^{N-1}W_k\left(\omega-\omega_k\right)=w\left(0\right)$$

Applying an Inverse Transform to this expression shows us that

$$\tilde{h}\left(n\right)=\underbrace{\sum_{k=0}^{N-1}w_k\left(n\right)e^{j\omega_k k}}_{\substack{Inverse\ transform\ of\\ \sum_{k=0}^{N-1}W_k\left(\omega-\omega_k\right)}}=\underbrace{Nw\left(0\right)\delta\left(n\right)}_{Inverse\ transform\ of\ Nw\left(0\right)}$$

And finally, the reconstructed signal is given by –

$$y\left(n\right)=\sum_{k=0}^{N-1}y_k\left(n\right)=\sum_{k=0}^{N-1}\sum_{m=-\infty}^{\infty}x\left(n-m\right)h_k\left(m\right)=Nw\left(0\right)x\left(n\right)$$

The Spectrogram

The Spectrogram is our way of presenting signals that have been STFT-transformed. It is basically a two-variable function, with the variables being the time and frequency. The Spectrogram is used to depict the changes in Amplitudes of signals (usually graphically presented by changes of color). Similarly to what was done in section 2.2.3.3.1, we can discuss two different kinds of spectrograms - the narrow band Spectrogram and the wide band Spectrogram. When dealing with speech signals, the first kind exhibits the harmonic structure in x(n) as horizontal striations and the second kind exhibits the periodic structure in x(n) as vertical striations.

Amplitude Modulation

Introduction and Motivational Background

Radio signals can be used to carry information. The information, which may be audio, data or any other information, is used to modify (modulate) a high frequency signal known as the carrier. The information superimposed onto the carrier forms a radio signal which is transmitted.
In order to receive the transmitted information the carrier is removed from the radio signal and reconstituted in its original format in a process known as demodulation.
There are many different varieties of modulations but they all fall into three basic categories, namely amplitude modulation, frequency modulation and phase modulation. In this project we will make a use of Amplitude Modulation and will now review some if its properties.
Possibly the most obvious method of modulating a carrier is to change its amplitude in order to represent the information desired to convey. E.g. in order to convey digital binary information one would possibly alter the amplitude between to values. The modulated data would appear as displayed in Figure 11.
An audio signal, which is clearly of major interest to our work, is essentially analog data. We will treat throughout this paper all audio signals as sums of Amplitude Modulated (AM) signals; for wideband signals this will be done by treating each STFT frequency band as an AM signal.

Amplitude Modulation

Analytical Definition

Assuming the carrier signal is of the form –

$$f_c(t) = A_c(t)\cos(\omega_c t + \varphi_c)$$

Applying an Amplitude Modulation requires the alteration of A_c. Hence, we can set A_c to be of the form –

$$A_c = A\left[1 + m(t)\right], \quad \left|m(t)\right| \le 1$$

We conclude that the modulated signal is $f_c(t) = A\left[1 + m(t)\right]\cos(\omega_c t + \varphi_c)$

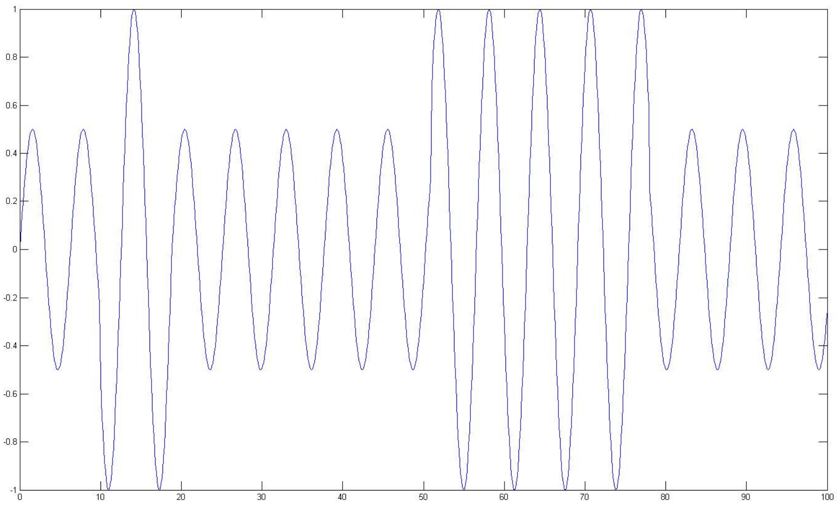


Figure 11 – Waveform of an AM digital binary signal

Spectral Decomposition

Let us assume the modulating signal is $Am(t) = a\cos(\omega_m t + \varphi_m)$. We then deduce that the modulated signal is –

$$f(t) = \left[A + a\cos(\omega_m t + \varphi_m)\right]\cos(\omega_c t + \varphi_c) = A\left[1 + m_a\cos(\omega_m t + \varphi_m)\right]\cos(\omega_c t + \varphi_c)$$

Defining $m_a = a/A$ as the modulation index and imposing $m_a < 1$ in order for the signals amplitude not to change sign (shift phase).

Now by using the trigonometric identity $\cos(\alpha)\cos(\beta) = \frac{1}{2}\left[\cos(\alpha+\beta) + \cos(\alpha-\beta)\right]$ we further conclude that the modulated signal is –

$$f(t) = A\left\{\cos(\omega_c t + \varphi_c) + \frac{m_a}{2}\cos\left((\omega_c + \omega_m)t + (\varphi_c + \varphi_m)\right) + \frac{m_a}{2}\cos\left((\omega_c - \omega_m)t + (\varphi_c - \varphi_m)\right)\right\}$$

It is now noticeable that by modulating the signal we generated three spectral components in the frequencies $(\omega_c),(\omega_c + \omega_m),(\omega_c - \omega_m)$ which are called respectively the carrier, the upper sideband and the lower sideband. Now that the modulating signal isn't a pure cosine we can use Fourier Decomposition in order to deduce its spectral components and display it as –

$$m(t) = \sum_i a_i \cos(\omega_i t + \varphi_i)$$

We can then conclude that the general spectral shape of an AM signal is as displayed in Figure 12.

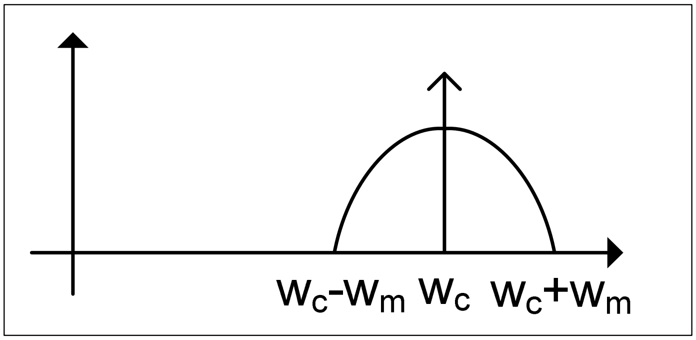


Figure 12 – Spectral Decomposition of an AM Signal

AM Demodulation

The signal may be demodulated using a system whose principle operation method is depicted in Figure 13. Here, the signal is multiplied with a locally-generated signal with the same frequency and phase (Coherent Demodulation) as the carrier. In this way the signal is converted down to the baseband frequency. A by-product of the multiplication is a high frequency component.

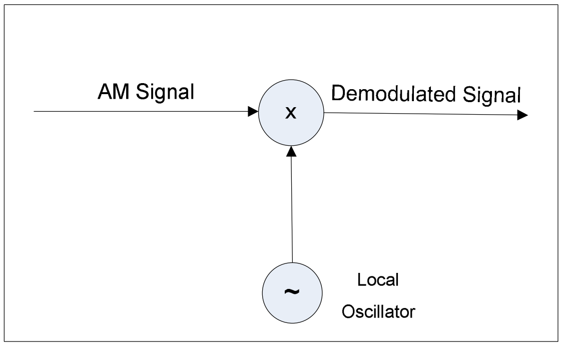


Figure 13 – AM Demodulator

In our work, we will try to use demodulation (envelope detection) in order to learn and observe the "AM properties" of various signals (speech, Cyclo-stationary noise), and we will try to use such properties in the enhancement process.

Modulation Transforms

Introduction and Motivational Background

According to [10] - – *"Speech signals are regarded as the result of a process in which a carrier signal (the speaker's voice), whose properties are given by organic and expressive factors, has been modulated with conventional linguistic speech gestures"* And *"… For the perception of the different types of information in speech, this implies that a demodulation is necessary in order to be able to separate them*".

The study from which we quote here suggests the motivation to withhold some type of modulation analysis on speech signals. The desire to devise an operator whose outcome is a signal displayed in a new domain in which "modulational" properties are visible is self-evident (we will later see that this domain is the “Joint-Frequency” domain).

Modulation transforms, which will be presented in the coming sections, include a two stage transformation. Firstly they make a use of the STFT in order to transform a signal to the Time-Frequency domain. They subsequently use a Modulation Transform in order to transform a signal to the new Modulator domain.

Definitions

The Modulator and the Envelope Detector

In this paper the term Modulator will refer to the Envelope. The term modulator is used in many ways and has many definitions. Hence, this section will implicitly define an Envelope and hereby after, by using this definition, define the Carrier. In section 0 we define various Envelope Detectors implicitly.

Without loss of generality, a modulated signal can expressed as –

$$x(t) = m(t)c(t),$$

where we shall denote the Modulator as $m(t)$ and $c(t)$ as the Carrier.

The Modulator is a low-pass signal that describes the AM of the signal while the carrier is a narrowband signal, typically of a much higher frequency, that describes the FM of the signal.

We denote the CT Envelope Detector $\mathfrak{D}$ by –

$$m(t) \triangleq \mathfrak{D}\{x(t)\}$$

Or in DT regime by -

$$m(n) \triangleq \mathfrak{D}\{x(n)\}$$

Assuming the modulated signal is of the form $x(t)=m(t)c(t)$ we may name the CT Complementary Detector D^c and define it as –

$$c(t)=\mathfrak{D}^c\{x(t)\} \triangleq \frac{x(t)}{\mathfrak{D}\{x(t)\}}$$

Or in DT regime by –

$$c(n)=\mathfrak{D}^c\{x(n)\} \triangleq \frac{x(n)}{\mathfrak{D}\{x(n)\}}$$

It is immediately noticeable that the output of D^c is the signal Carrier. Hence, we name it the Carrier Estimator.

For a signal, which is a function of both time and frequency (a signal which has been transformed by the STFT), the convention is that

$$M(t,\omega) \triangleq \mathfrak{D}\{X(t,\omega)\}, \quad C(t,\omega)=\mathfrak{D}^c\{X(t,\omega)\}$$

Or in DT regime by -

$$M(n,\omega) \triangleq \mathfrak{D}\{X(n,\omega)\}, \quad C(n,\omega)=\mathfrak{D}^c\{X(n,\omega)\}$$

Applying the Envelope Detector to a signal in the time-frequency domain will yield different Carriers for different times. Hence, we adopt the term Carriers (plural) as opposed to a single Carrier when dealing with time domain signals.

Properties of the Envelope Detector

I. An Envelope detector is a non-linear operator. By this Envelope Detectors constitute an extension to Linear System theory.

II. An Envelope Detector or Carrier Estimator must be projections. That is,

$$\mathfrak{D}\{\mathfrak{D}\{x(t)\}\}=\mathfrak{D}\{x(t)\}, \quad \mathfrak{D}^c\{\mathfrak{D}^c\{x(t)\}\}=\mathfrak{D}^c\{x(t)\}$$

III. An Envelope Detector must be frequency-shift invariant. That is,

$$\mathfrak{D}\left\{e^{j\omega_0 t}x(t)\right\} = \mathfrak{D}\left\{x(t)\right\}, \quad \mathfrak{D}^c\left\{e^{j\omega_0 t}x(t)\right\} = \mathfrak{D}^c\left\{x(t)\right\}$$

This property must hold because a signal’s Modulator is independent of the signal’s Carrier or its central frequency location.

IV. The Envelope Detector must preserve BW. As will be shown in further sections, we may want to apply the Envelope Detector on a signal which has been segmented in the frequency domain or divided into sub bands. Hence, we wouldn’t want the Envelope Detector to broaden the BW of each sub band.

The Modulation Transforms

We shall define now new transforms, which will aid us as a main tool in our work. It should be noted that all of the definitions and mathematics in the current and next sections (2.3.2.3 through 2.3.2.7) are based on [8] -.

The Modulation Transform

Using the groundwork developed in section 0 we can now define the Modulation Transform, the DT Modulation Transform and the Discrete Modulation Transform.
The CTMT is defined by

$$X(H,\omega) = \mathrm{M}\left\{x(t)\right\} \triangleq \mathrm{F}\left\{\mathrm{D}\left\{STFT\left\{x(t)\right\}\right\}\right\} = \int \mathrm{D}\left\{\int x(t)\,w(\tau - t)\,e^{-j\omega t}dt\right\} e^{-jH\tau}d\tau$$

Notice the new variable marked by an H , which represents the modulation frequency.
In DT regime we have,

$$\begin{aligned} X(\eta,\omega) &= DTMT\left\{x(t)\right\} \triangleq DTFT\left\{\mathrm{D}\left\{DTSTFT\left\{x(n)\right\}\right\}\right\} \\ &= \sum_m \mathrm{D}\left\{\sum_n x(n)\,w(m-n)\,e^{-j\omega n}\right\} e^{-j\eta m} \end{aligned}$$

And in Discrete time and frequency regimes we have,

$$\begin{aligned} X(i,k) &= DMT\left\{x(n)\right\} \triangleq DFT\left\{\mathfrak{D}\left\{DSTFT\left\{x(n)\right\}\right\}\right\} \\ &= \sum_m \mathfrak{D}\left\{\sum_n x(n)\,w(m-n)\,e^{-j2\pi nk/K}\right\} e^{-j2\pi mi/I} \end{aligned}$$

We shall name the domain in which these transformed functions exist as the "Joint Frequency Domain". It is easy to see that this domain indeed consists, as its name suggests, of 2 frequency axes - a modulator frequency axis and a harmonic frequency axis. The two-stage process which is defined by the Modulation Transforms is depicted in Figure 14 -

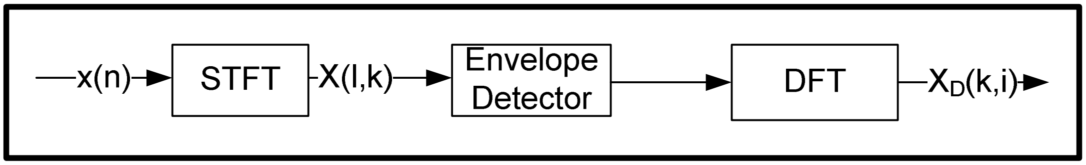


Figure 14 – Depiction of the two-stage process defined by Modulation Transforms

The Short Time Modulation Transform

We define the Short Time Modulation Transform of a signal $x(t)$ by

$$X(T,H,\omega) = STMT\{x(t)\} \triangleq STFT\{\mathrm{D}\{STFT\{x(t)\}\}\}$$
$$= \int \mathrm{D}\left\{\int x(t) w(\tau - t) e^{-j\omega t} dt\right\} v(T-\tau) e^{-jH\tau} d\tau$$

Here, $w(t)$ and $v(\tau)$ are analysis windows for the signal at the various stages of the transformation. $w(t)$ at the first $STFT$ stage and $v(\tau)$ at the second, modulation variable, $STFT$ stage. T represents the new time variable, H the modulation frequency variable and ω the original frequency variable.

Similarly we define the DT Short Time Modulation Transform of a signal $x(n)$ by

$$X(l,\eta,\omega) = DTSTMT\{x(t)\} \triangleq DTSTFT\{\mathfrak{D}\{DTSTFT\{x(n)\}\}\}$$
$$= \sum_m \mathfrak{D}\left\{\sum_n x(n) w(m-n) e^{-j\omega n}\right\} v(l-m) e^{-j\eta m}$$

Again, $w(n)$ and $v(m)$ are the analysis windows at the various transformation stages. Here l represents the new time variable, η the modulation frequency variable and ω the original frequency variable.

Finally we can define the Discrete Short Time Modulation Transform of a signal $x(n)$ by

$$X(l,i,k) = DSTMT\{x(n)\} \triangleq DSTFT\{\mathfrak{D}\{DSTFT\{x(n)\}\}\}$$
$$= \sum_m \mathfrak{D}\left\{\sum_n x(n) w(m-n) e^{-j2\pi nk/K}\right\} v(l-m) e^{-j2\pi mi/I}$$

Following the same naming conventions as in [8] -, we shall name the domain in which these transformed functions exist as the "Short Time Joint Frequency Domain"

Inverse Modulation Transforms

The Inverse Modulation Transforms

Practicing the terms defined above we immediately derive that the Inverse Modulation Transform in CT regime is given by

$$
\begin{aligned}
x(t) &= \mathfrak{M}^{-1}\left\{X(H,\omega)\right\} = ISTFT\left\{\mathfrak{F}^{-1}\left\{X(H,\omega)C(\tau,\omega)\right\}\right\} \\
&= \frac{1}{2\pi}\int\int\left[\frac{1}{2\pi}\int X(H,\omega)e^{jH\tau}dH\right]C(\tau,\omega)e^{j\omega t}d\tau d\omega
\end{aligned}
$$

The Inverse Modulation Transform in DT regime is given by

$$
\begin{aligned}
x(n) &= IDTMT\left\{X(\eta,\omega)\right\} = IDTSTFT\left\{IDTFT\left\{X(\eta,\omega)C(m,\omega)\right\}\right\} \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{m}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} X(\eta,\omega)e^{j\eta m}d\eta\right]C(m,\omega)e^{j\omega n}d\omega
\end{aligned}
$$

The Inverse Short Time Modulation Transforms

Practicing the terms defined above we immediately derive that the Inverse Short Time Modulation Transform in CT regime is given by

$$
\begin{aligned}
x(t) &= ISTMT\left\{X(T,H,\omega)\right\} = ISTFT\left\{ISTFT\left\{X(T,H,\omega)\right\}C(\tau,\omega)\right\} \\
&= \frac{1}{2\pi}\int\int\left[\frac{1}{2\pi}\int X(T,H,\omega)e^{jH\tau}dTdH\right]C(\tau,\omega)e^{j\omega t}d\tau d\omega
\end{aligned}
$$

The Inverse Short Time Modulation Transform in DT regime is given by

$$
\begin{aligned}
x(n) &= IDTSTMT\left\{X(l,\eta,\omega)\right\} = IDTSTFT\left\{IDTSTFT\left\{X(l,\eta,\omega)\right\}C(m,\omega)\right\} \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{m}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{l} X(l,\eta,\omega)e^{j\eta m}d\eta\right]C(m,\omega)e^{j\omega n}d\omega
\end{aligned}
$$

Decimation of Various Transforms

The operation of Discrete Short Time transforms on a signal vector, as exhibited in the past few sections, yield tensors or matrixes. In some cases, the definition of the STFT may cause it to contain data which is redundant and we may want to decimate it (along a certain dimension). We shall mark such an operation by a subscript R , marking the decimation factor.

The decimated $DSTFT$ is defined by

$$X\left(nR,k\right)=DSTFT_R\left\{x\left(n\right)\right\}\triangleq\sum_m x\left(m\right)w\left(nR-m\right)e^{-j2\pi mk/K}$$

The decimated DMT is defined by

$$\begin{aligned}X\left(i,k\right)&=DMT_R\left\{x\left(n\right)\right\}\triangleq DFT\left\{\mathfrak{D}\left\{DSTFT_R\left\{x\left(n\right)\right\}\right\}\right\}\\&=\sum_m\mathfrak{D}\left\{\sum_n x\left(n\right)w\left(mR-n\right)e^{-j2\pi nk/K}\right\}e^{-j2\pi mi/I}\end{aligned}$$

The decimated $DSTMT$ is defined by

$$\begin{aligned}X\left(lRS,i,k\right)&=DSTMT_{R,S}\left\{x\left(n\right)\right\}\triangleq DSTFT_S\left\{\mathfrak{D}\left\{DSTFT_R\left\{x\left(n\right)\right\}\right\}\right\}\\&=\sum_m\mathfrak{D}\left\{\sum_n x\left(n\right)w\left(mR-n\right)e^{-j2\pi nk/K}\right\}v\left(lS-m\right)e^{-j2\pi mi/I}\end{aligned}$$

The Various Domains and Their Relations

In this document we shall denote a Domain $\mathcal{S}$ as the set of complex function of variables λ defined on I , i.e.,

$$\mathcal{S}\left(I,\lambda\right)=\left\{x\left(\lambda\right):I\to\mathbb{C}\right\}$$

Now by carefully examining the Domains defined so far we can summarize

Domain	Definition Region
CT Domain	$\mathcal{S}_{CT}\left(\mathbb{R};t\right)$
DT Domain	$\mathcal{S}_{DT}\left(\mathbb{N};n\right)$
CF Domain	$\mathcal{S}_{CF}\left(\mathbb{R};\omega\right)$
DF Domain	$\mathcal{S}_{DF}\left(\mathbb{N};k\right)$
Continuous Time-Frequency Domain ($STFT$ Domain)	$\mathcal{S}_{CT,CF}\left(\mathbb{R}^2;t,\omega\right)$

Discrete Time-Frequency Domain ($DSTFT$ Domain)	$\mathcal{S}_{DT,DF}\left(\mathbb{N}\times\mathbb{N};n,k\right)$

Table 4 – Definition Ranges of the Prominent Domains

One can easily notice that by applying an Envelope Detector to a signal the outcome is a function belonging the spaces $\mathcal{S}_F$ or $\mathrm{s}_{t,F}$. The same can be deduced regarding the Complementary Detector. Hence, we define the subspaces

Sub Domains	Definition Region
CT Modulator Domain (Envelope Detector)	$\mathcal{S}_{\mathfrak{D}(t,F)}\left(\mathbb{R}^2;t,\omega\right)$
DT Modulator Domain (Envelope Detector)	$\mathcal{S}_{\mathfrak{D}(n,F)}\left(\mathbb{N}\times\mathbb{R};n,\omega\right)$
CT Carrier Domain (Complementary Detector)	$\mathcal{S}_{\mathfrak{D}^c(t,F)}\left(\mathbb{R}^2;t,\omega\right)$
DT Carrier Domain (Complementary Detector)	$\mathcal{S}_{\mathfrak{D}^c(n,F)}\left(\mathbb{N}\times\mathbb{R};n,\omega\right)$

Table 5 – Definition Ranges of the Sub Domains

Assaying the Transformations one can easily notice the need for two more Domains that originate by applying Modulation Transformations.

Domains	Definition Region
Joint Frequency Domain (Modulation Transform)	$\mathcal{S}_{\mathfrak{J}}\left(\mathbb{R}^2;\omega,H\right)$
CT Short Time Joint Frequency Domain (Short Time Modulation Transform)	$\mathcal{S}_{\mathfrak{J}\mathfrak{s}(t)}\left(\mathbb{R}^3;t,\omega,H\right)$
DT Short Time Joint Frequency Domain (Short Time Modulation Transform)	$\mathcal{S}_{\mathfrak{J}\mathfrak{s}(n)}\left(\mathbb{N}\times\mathbb{R}^2;n,\omega,H\right)$

Table 6 – Definition Ranges of the Unaccustomed Domains

Finally By examining the outcome of the transformations defined so far we can derive the domain transitions brought about by them

Operator	Domain Transition
The CT Fourier Transform, $\mathfrak{F}$	$\mathcal{S}_{CT}\left(\mathbb{R};t\right)\rightarrow\mathcal{S}_{CF}\left(\mathbb{R};\omega\right)$
The DT Fourier Transform, $DTFT$	$\mathcal{S}_{DT}\left(\mathbb{N};n\right)\rightarrow\mathcal{S}_{CF}\left(\mathbb{R};\omega\right)$
The CT Short Time Fourier Transform, $STFT$	$\mathcal{S}_{CT}\left(\mathbb{R};t\right)\rightarrow\mathcal{S}_{CT,CF}\left(\mathbb{R}^2;t,\omega\right)$
The DT Short Time Fourier Transform, $DTSTFT$	$\mathcal{S}_{DT}\left(\mathbb{N};n\right)\rightarrow\mathcal{S}_{DT,CF}\left(\mathbb{N}\times\mathbb{R};n,\omega\right)$
The Discrete Short Time Fourier Transform, $DSTFT$	$\mathcal{S}_{DT}\left(\mathbb{N};n\right)\rightarrow\mathcal{S}_{DT,DF}\left(\mathbb{N}\times\mathbb{N};n,k\right)$

The CT Envelope Detector, $\mathfrak{D}$	$\mathcal{S}_{CT}\left(\mathbb{R};t\right)\to\mathcal{S}_{CT}\left(\mathbb{R};t\right)$ or $\mathcal{S}_{CT,CF}\left(\mathbb{R}^2;t,\omega\right)\to\mathcal{S}_{CT,CF}\left(\mathbb{R}^2;t,\omega\right)$
The DT Envelope Detector, $\mathfrak{D}$	$\mathcal{S}_{DT}\left(\mathbb{N};n\right)\to\mathcal{S}_{DT}\left(\mathbb{N};n\right)$ or $\mathcal{S}_{DT,CF}\left(\mathbb{N}\times\mathbb{R};n,\omega\right)\to\mathcal{S}_{DT,CF}\left(\mathbb{N}\times\mathbb{R};n,\omega\right)$
The CT Complementary Detector, $\mathfrak{D}^c$	$\mathcal{S}_{CT}\left(\mathbb{R};t\right)\to\mathcal{S}_{CT}\left(\mathbb{R};t\right)$ or $\mathcal{S}_{CT,CF}\left(\mathbb{R}^2;t,\omega\right)\to\mathcal{S}_{CT,CF}\left(\mathbb{R}^2;t,\omega\right)$
The DT Complementary Detector, $\mathfrak{D}^c$	$\mathcal{S}_{DT}\left(\mathbb{N};n\right)\to\mathcal{S}_{DT}\left(\mathbb{N};n\right)$ or $\mathcal{S}_{DT,CF}\left(\mathbb{N}\times\mathbb{R};n,\omega\right)\to\mathcal{S}_{DT,CF}\left(\mathbb{N}\times\mathbb{R};n,\omega\right)$
The CT Modulation Transform, $\mathfrak{M}$	$\mathcal{S}_{CT}\left(\mathbb{R};t\right)\to\mathcal{S}_{\mathfrak{I}}\left(\mathbb{R}^2;\omega,H\right)$
The DT Modulation Transform, $DTMT$	$\mathcal{S}_{DT}\left(\mathbb{N};n\right)\to\mathcal{S}_{\mathfrak{I}}\left(\mathbb{R}^2;\omega,H\right)$
The CT Short Time Modulation Transform, $STMT$	$\mathcal{S}_{CT}\left(\mathbb{R};t\right)\to\mathcal{S}_{\mathfrak{Is}(t)}\left(\mathbb{R}^3;t,\omega,H\right)$
The DT Short Time Modulation Transform	$\mathcal{S}_{DT}\left(\mathbb{N};n\right)\to\mathcal{S}_{\mathfrak{Is}(n)}\left(\mathbb{N}\times\mathbb{R}^2;n,\omega,H\right)$

Table 7 – Domain Transitions Driven by Transformations

The transitions driven by Inverse Transformations can be inferred simply by inverting the direction of the arrow in the appropriate (forward) Transformation found in Table 7.

Linear Filtering Interpretation for signal modification

Applying discrete modulation transforms requires applying a $DSTFT$. Hence, using the linear filtering interpretation of the $STFT$, we can construct a system used for general signal filtering. Such a system is depicted in Figure 15.

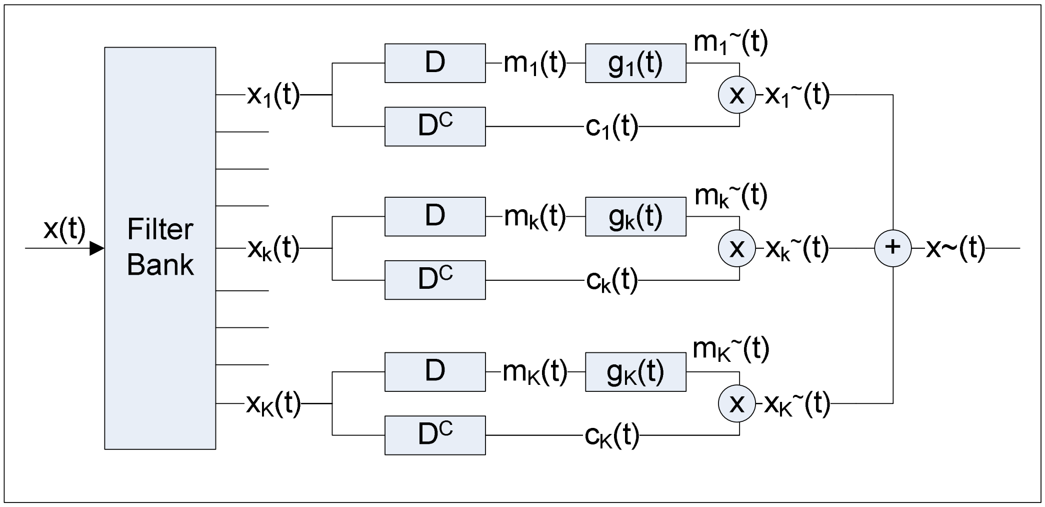


Figure 15 – Linear Filtering Interpretation of Discrete Modulation Transforms

In this system we have, for each sub-branch

$$m_k(t) = \mathfrak{D}\{x_k(t)\}, \quad c_k(t) = \mathfrak{D}^c\{x_k(t)\}, \quad x_k(t) = m_k(t)c_k(t)$$

And due to the linear filtering interpretation of the $DSTFT$, each modulator is filtered by the LTI filter $g_k(t)$,

$$\tilde{m}_k(t) = m_k(t) * g_k(t)$$

Finally the output signal is given by a K -fold summation of the sub branches,

$$\tilde{x}(t) = \sum_{k=1}^{K} \tilde{x}_k(t)$$

Now assuming the filter bank is the exact inverse of what was applied to the input signal by now, we have a system whose output holds,

$$x(t) = \tilde{x}(t) = \sum_{k=1}^{K} \tilde{m}_k(t) c_k(t)$$

We may utilize such a system for filtering a signal, by applying the desired filter on each sub channel.

An interesting property of such a filter bank can be derived by manipulating the mathematical expressions. The filter bank enables perfect signal reconstruction if it satisfies,

$$x(t) - \sum_{k=1}^{K} x_k(t) = x(t) - \sum_{k=1}^{K} x(t) * h_k(t) = x(t) - x(t) * \sum_{k=1}^{K} h_k(t) = x(t) * \left[\delta(t) - \sum_{k=1}^{K} h_k(t)\right] = 0$$

This implies that a perfect reconstruction filter must sum to an impulse:

$$\sum_{k=1}^{K} h_k(t) = \delta(t)$$

Implicit definition of Envelope Detectors

In section 0 we have presented an implicit operator, namely the Envelope Detector. Implicitness allowed us to stay within theoretical boundaries without delving into pragmatic engineering limitations of each individual Envelope Detector. In the sections to come we will explicitly discuss about the various Envelope Detectors, their sweet spots and their weak spots. In general, we will distinguish between *Incoherent Envelope Detection* and *Coherent Envelope Detection*. The basic difference between these two types is that the Incoherent Detection requires the detected envelopes to be real-valued signals, whereas the Coherent Detection allows complex envelopes.

Incoherent Envelope/Carrier Detectors

The Hilbert Envelope Detector (real envelopes)

The Hilbert envelope detector, denoted by $\mathfrak{D}_H\left\{x(t)\right\}$ when operating on $x(t)$, is based on the Hilbert Transform and is used for real-valued signals ([5] -, [8] -). The Hilbert Transform shifts the positive frequency spectral content of a signal's phase by $\frac{\pi}{2}$ radians and negative frequency spectral content of a signal's phase by $-\frac{\pi}{2}$ radians. By defining

$$H\left\{x(t)\right\}=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau$$

It is immediately noticeable that the Hilbert Transform produces a new signal belonging to the same domain.

And also by defining:

$$x^{+}(t)=x(t)+jH\left\{x(t)\right\}$$

We obtain a new analytic signal $x^{+}(t)$. Its Fourier Transform $X^{+}(\omega)$ is related to $x(t)$ Fourier Transform $X(\omega)$ by

$$X^{+}(\omega)=\begin{cases}2X(\omega), & \omega>0\\ X(\omega), & \omega=0\\ 0, & \omega<0\end{cases}$$

We define the Hilbert Envelope Detector, for real-valued signals, as

$$m(t)=\mathfrak{D}_H\left\{x(t)\right\}\triangleq\left|x^{+}(t)\right|=\left|x(t)+jH\left\{x(t)\right\}\right|=\left|x(t)+j\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau\right|$$

The complementary Carrier Detector is then defined by

$$c(t)=\mathfrak{D}_H^c\left\{x(t)\right\}\triangleq\left|\cos\left(\arg\left[x^+(t)\right]\right)\right|=\left|\cos\left(\arg\left[x(t)+jH\left\{x(t)\right\}\right]\right)\right|$$

Utilizing the $STFT$ filter bank interpretation, one may apply an Envelope Detector or Carrier Detector on a signal composed of K sub-bands as depicted and described in section 0.

We may now want to analyze the Hilbert Envelope Detector's product when operated on an AM signal. We will assume the signal we have is of AM form and that $u_m(t)$ has no spectral content above $\frac{\omega}{2\pi}$ (assuming the modulator is of much lower spectral content is equivalent to the requirement of the signal being a narrow band signal. Such requirements are often common for audio signals). Such an AM-signal is:

$$u(t)=u_m(t)\cos(\omega t+\phi)$$

By recognizing the phase shifts that the Hilbert Transform produces we have

$$\hat{u}(t)=H\left\{u_m(t)\cos(\omega t+\phi)\right\}=u_m(t)\sin(\omega t+\phi)$$

We may now easily reconstruct the signal's waveform by,

$$(\omega t+\phi)_{\mathrm{mod}\ 2\pi}=\tan^{-1}\left(\hat{u}(t),u(t)\right)=\arg\left(u(t)+i\hat{u}(t)\right)$$

The Magnitude Envelope Detector (complex envelopes)

The Magnitude Envelope Detector is practically an extension of the Hilbert Envelope Detector for complex values signals. We may define and denote the Magnitude Envelope Detector by

$$m(t)=\mathfrak{D}_{\|}\left\{x(t)\right\}\triangleq\left|x^+(t)\right|=\left|x(t)+jH\left\{x(t)\right\}\right|$$

And the complementary Carrier Detector by

$$c(t)=\mathfrak{D}_{\|}^c\left\{x^+(t)\right\}=\exp\left\{j\arg\left[x^+(t)\right]\right\}$$

Weak Spots of Incoherent Envelope Detectors

We hereby emphasize the Incoherent Envelope Detectors limitations and by this derive the need for Coherent Envelope Detectors.

I. Employing an Envelope Detector on sub-bands usually yields signals which exceed the BW of the original signal. Hence, the Incoherent Envelope Detector usually does not preserve the BW.
Let us consider the signal the band-limited signal

$$x(t)=e^{j\omega_c t}\cos(\omega_m t), \quad \omega_c>\omega_m$$

We consider the spectrum of the envelope noticing the complex-valued envelope and thus utilizing the Magnitude Envelope Detector

$$m(t)=|x(t)|=\left|e^{j\omega_c t}\cos(\omega_m t)\right|=|\cos(\omega_m t)|$$

Now, due to the discontinuities at $t=\frac{\pi}{2\omega_m}+\frac{\pi k}{\omega_m}, \quad k\in\mathbb{N}$ the spectrum of the envelope is infinite and obviously not band-limited.
A sufficient condition for signals to be of a band-limited envelope is in the case of a signal composed of a monochromatic carrier multiplied by a band-limited non-negative real envelope:

$$x(t)=a(t)c(t), \quad c(t)=e^{j\omega_0 t}$$

In this case we yield

$$\begin{aligned}&m(t)=|x(t)|=a(t)\to\\&M(\omega)=A(\omega)\end{aligned}$$

Where $A(\omega)$, as stated above, is band-limited.

It is suggested that due to emprical observations ([1] -) such a requirement is also a necessary condition.

II. The Incoherent Envelope Detector, by usage of the Hilbert Transform, forces a conjugate symmetric spectrum on the Modulator. This does not normally apply to most natural signals ([1] -).

III. Applying a convolution or filtering on a signal which fulfills the requirements stated in the first statement of this section yields a signal which doesn't fulfill these requirements. e.g. – a signal of a non-negative real envelope, after being convolved, isn't necessarily band-limited. Hence the modulator, as extracted by an Envelope Transform, doesn't necessarily belong to $\mathcal{S}_{\mathfrak{D}}$. The Modulator Domain is then said to be "not closed" under a convolution.

Coherent Envelope/Carrier Detectors

It can be shown [1] that by acknowledging that ability of modulators to be complex valued, the problematic issues discussed in section 0 are almost plenarily mitigated. By doing so we can define new Carrier Detectors. It should be noted that our work, that will be presented in the next chapters, did not make usage of Coherent Envelope Detectors.

The Smoothed Hilbert Carrier Detector

Considering a narrow band signal $x(t)=a(t)e^{j\phi(t)}$, the signal's phase is given by $\phi(t)$. Adopting the form suggested in [1], we assume the phase of the signal can be written as

$$\phi(t)=\omega_m t+\phi_0+\theta(t)$$

To say, the phase is constituted of a linear phase component, an initial phase ϕ_0 and a phase deviation term. We may choose ω_m and ϕ_0 in such way so $\phi(t)$ has zero mean and such a choice is unique. The frequency ω_m is comprehendible as an "average" frequency of the narrow band signal. We refer to this frequency as the "Mid-Band" frequency.
The smoothed (band limited) signal's phase is given by

$$\tilde{\phi}(t)=\omega_m t+\phi_0+\left[\theta(t)*h_{lp}(t)\right]$$

The Carrier Estimator is then defined by

$$c(t)=\mathfrak{D}_{SH}^{c}\left\{x(t)\right\}\triangleq e^{j\tilde{\phi}(t)}$$

and the Coherent Envelope is given by

$$m(t)=\mathfrak{D}_{SH}\left\{x(t)\right\}\triangleq a(t)e^{j\phi(t)}e^{-j\tilde{\phi}(t)}$$

It noticable that by choosing $h_{lp}(t)=\delta(t)$, the signal isn't smoothed and we yield the Hilbert Envelope defined in section 0. By choosing $h_{lp}(t)=1$ we yield a fully-smoothed envelope whose carrier is located at the Mid Band frequency.

The Instantaneous Frequency Carrier Detector

The Second type of Coherent Carrier Detector is based on [1] - and comprises of an alterred FM detector. Let us define a number of new signals and use the signals' representation in polar form. We shall also assume the ability to deompose the signal into its modulator and carrier when presented in the suggested polar form.

$$\begin{aligned}
&x(t)=m(t)c(t)=a_x(t)e^{j\phi_x(t)}=\left[a_m(t)e^{j\phi_m(t)}\right]\left[a_c(t)e^{j\phi_c(t)}\right]\\
&a_x(t)=a_m(t)a_c(t),\quad \phi_x(t)=\phi_m(t)+\phi_c(t)
\end{aligned}$$

The objective is to indentify $a_m(t),\phi_m(t)$ given $a_x(t),\phi_x(t)$. For this we define a number of new signals –

$$\begin{aligned}
&I=\mathrm{Re}\left(x(n)\right),Q=\mathrm{Im}\left(x(n)\right)\\
&Z_I=I(n-1)I(n+1)+Q(n-1)Q(n+1)\\
&Z_Q=I(n-1)Q(n+1)-Q(n-1)I(n+1)\\
&Z=Z_I+jZ_Q
\end{aligned}$$

Now by calculating -

$$\begin{aligned}
&\alpha(n)=\begin{cases}\sqrt{\left(Z(n)/\left|Z(n)\right|\right)}, & \left|Z(n)\right|>\varepsilon\\ \alpha(n-1), & \left|Z(n)\right|\le\varepsilon\end{cases}\\
&W(n)=W(n-1)\alpha(n)
\end{aligned}$$

We yield $\alpha(n)$ which is the instantaneous phase, where as $W(n)$ is an integrated term containing the decomposed carrier signal. By multiplying the input signal by the conjugate of its detected carrier we yield the envelope

$$X(n) = A(n)e^{j\phi(n)}$$
$$I = \mathrm{Re}\left(x(n)\right) = A(n)\cos\left(\phi(n)\right), Q = \mathrm{Im}\left(x(n)\right) = A(n)\sin\left(\phi(n)\right)$$

Placing the suggested terms for I, Q in Z yields –

$$Z = A(n-1)A(n+1)\exp\left\{j\left[\phi(n+1) - \phi(n-1)\right]\right\}$$
$$\alpha(n) = \exp\left\{j\left[\phi(n+1) - \phi(n-1)\right]\right\}$$
$$W(n) = \exp\left\{j\phi(n)\right\}$$

Cyclo-Stationary Noise

Introduction and Motivational Background

Due to the DMT's ability to extract AM content in audio signals, we are highly motivated to try to use this ability in order to apply some kind of Noise-Enhancement technique and or signal separation. The question regarding what noise model should we work on then arouses. Hence, we begin our work with the analysis of the DMT's ability to enhance an audio signal with the interference of Cyclo-Stationary noise.

A Cyclo-Stationary noise is basically an AM modulated noise. Common and easily comprehensible examples are engine noises.

Mathematical Definition

In this work we chose to use a specific form of WGN modulated noise as our model for Cyclo-Stationary noise (from here on referred to as CSN). Its mathematical representation is as follows:

We shall first define a Random Gaussian Process:

$$\gamma(n) \sim N\left(0,\sigma^2\right)$$

And we will define the noise, $b(n)$, as the product of the multiplication of this process by a cosine envelope:

$$b(n) = \gamma(n)\left(1+\cos\left(2\pi f_{\text{mod}} n + \phi\right)\right)$$

As suggested above, this is a vector of white noise multiplied by a cosine with a DC level of 1. The DC level is shifted from zero in order to be able to use the Magnitude Envelope Detector without the creation of DC artifacts. Figure 16 depicts this noise as created using Matlab –

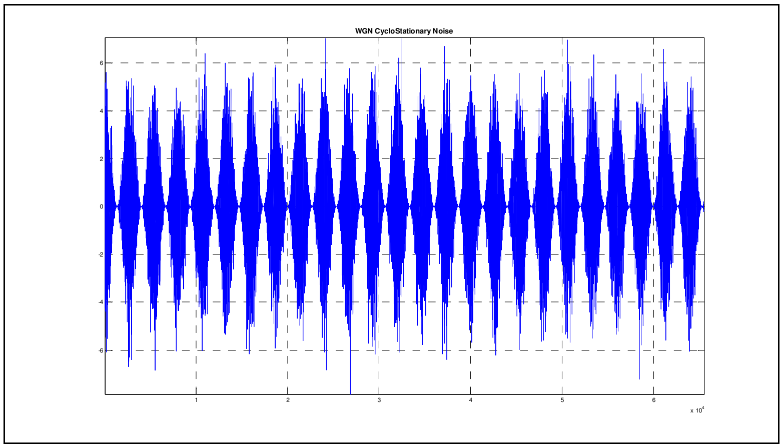


Figure 16 – WGN Cyclo-Stationary Noise

Objective Measures

Introduction and Motivational Background

Any scientific work requires methods of comparison to past attempts/experiments. When dealing with separation or filtering attempts, the human ear provides us a very subjective comparison tool. Such a tool is, in many cases, not sufficient to determine which, out of several audio files, "sounds better", and in what "sense". In addition, it cannot provide any numerical appreciation to an experiment's success/failure.

All of the above set the need to establish, or mathematically define, agreed criteria to appreciate separation/enhancement results. The next chapter details these criteria, which are also referred to as "objective measures", and the way they were defined. As we shall see, some objective measures use "original signals", i.e. "pure" speech or "pure" noise (before mixed with each other). This is, of course, a valid theoretic approach though it is obvious that in real-life separation/enhancement attempts such signals do not exist.

It should perhaps be noted that some of these figures of merit were used not only to compare between the results of various enhancement attempts we performed, but also to compare our attempts and algorithms with the results achieved by other public separation algorithms, such as OM-LSA.

Mathematical Definitions

NMSE

We shall define the Normalized Mean Square Error (NMSE) as following:

$NMSE = 10\log_{10}\left(\frac{\sum_i \left|S[i]-\tilde{S}[i]\right|^2}{\sum_i \left|S[i]\right|^2}\right)$, where S[i] and $\tilde{S}[i]$ are the i-th samples of a desired and an estimated signal, respectively.

It is easy to see, from the way it is defined, that NMSE compares signals in the time domain. The motivation behind its definition is to measure the "distance" between a signal achieved by enhancement and the original signal (before mixing it with noise). Thus, decreases in this objective measure indicate better separation, with ideal separation achieved when $NMSE = -\infty$.
The NMSE is relevant and more indicative when applied on frames and not on specific samples. This implies that its main strength is the ability to give a numeric judgement of how well an entire frame was separated. On the other hand, it also implies that the NMSE cannot indicate good (or bad) separation in certain times (in the same frame), and that the NMSE judgement might become less valueable for longer frames.
We experimentally saw that the NMSE was not good enough as an objective measure, in the sense that it did not always numerically reflect improvements in the way enhanced signals sounded.

LSD

We shall define the LSD the following way:

$$LSD = \sqrt{\frac{1}{TM}\left(\sum_{k=0}^{T-1}\sum_{l=0}^{M-1} 10\log_{10}\left|S_n(l,k)\right|^2 - 10\log_{10}\left|\hat{S}(l,k)\right|^2\right)}.$$

This objective measure compares the original and estimated speech in the STFT domain, and measures (on a logarithmic scale) how "far" they are from one another. Here T and M denote the number of points in each axis of an STFT matrix, i.e. number of frequencies multiplied by the number of time samples (thus $\frac{1}{TM}$ is just a normalization factor). LSD decreases, again, indicate an improved estimation, ideal recovery/enhancement is indicated by $LSD = -\infty$.

It should be noted that the LSD, in this definition, gives the same significance to all frequencies in the spectrum, including those that might be of less significance. This might be considered a disadvantage of this appreciation method. However, it does give a better insight on an estimation's quality than the NMSE, as it takes into consideration more factors (i.e. distortions in certain frequencies etc.).

SNR

The output SNR of a filtering system will be defined as:

$$SNR = 10\log_{10}\left(\sum_{n=0}^{N-1}\left(s(n)\right)^2 / \sum_{n=0}^{N-1}\left(s(n)-\hat{s}(n)\right)^2\right)$$

The SNR is measured over a time frame, and it measures the cumulative ratio, over time, between an estimated speech signal and the original noise. The input SNR of a system can be defined similarly, by replacing the estimated speech with the original one.

Measuring only the input SNR, or only the output SNR, gives very little information. It is much more useful to measure both, and to compare their sizes. This objective measure indicates how good a filtering attempt was based on how much it improved the SNR, i.e. how big is the difference between the output and the input SNR values.

This objective measure pays more interest to signal intensities than other figures of merit do. It can be considered an advantage because it easily distinguishes between filtering attempts that were made in different conditions. It is, intuitively, much harder to separate a speech signal from a louder noise, but this objective measure shows, numerically, how harder it really is. On the other hand, this objective measure is time-based, thus it pays less importance to changes in certain frequencies.

Thoughout our work, we found this objective measure very useful, and it was used by us quite commonly.

DMT Norm Distance

This objective measure is unique to our work, and was a first attempt (that we were aware of) to define comparison criteria based on the Joint Frequency domain. We shall denote the norm of a matrix M:

$$Norm(M) = \sum_i \sum_j \left|M(i,j)\right|$$

And the Norm Distance between 2 matrices (M_1 and M_2) as:

$$\sum_{k,i} \left\| M_1(k,i) \right| - \left| M_2(k,i) \right\|$$

We shall refer to signals that were DMT-tramsformed as matrices (where the rows and coloumns represent the carrier and modulation frequency behaviours) in order to use this objective measure.

This objective measure helps in the analysis of changes in the Joint Frequency domain (due to filtering or other actions). As such, it also shows how "close" 2 signals are in the Joint Frequency domain. The differences between an original signal's DMT and a filtered signal's DMT are reflected through their DMT matrices' norms, and ideal recovery of a signal would result in a difference of zero between the norms. In case of non-ideal Joint Frequency filtering, it shows how much (if at all) the noise's intesity was weakened.

In our work this objective measure played a role in the estimation of the modulation frequency of Cyclo-Stationary noises, as we will explain in the following chapter.

3 Proposed Algorithms

3.1 Envelope Detectors

As stated before, we chose to work with the Magnitude Detector due to its advantages over both other detectors (precision and accuracy) in spite of the artifacts it causes, which is his most notable disadvantage.

Magnitude Envelope Detector

In order to evaluate the performance of this Envelope Detectors' ability to carry out enhancement we first devise two signals. One is a synthesised AM signal and the second is a speech recording of a female pronouncing a sentence from TIMIT.
The AM signal is constructed by superposition of 4 AM signals :

$$x(t)=\sum_{i=1}^{4}\cos\left(2\pi\Omega_{c,i}t\right)\left[0.5+0.4\cos\left(2\pi\Omega_{m,i}t\right)\right]$$

$$\left\{\Omega_{c,i}\right\}=\left[1031,1344,2250,2844\right]$$

$$\left\{\Omega_{m,i}\right\}=\left[16.6,9.8,19.5,12.7\right]$$

Each signal was then analysed in the $\mathcal{S}_{DT}$, $\mathcal{S}_{DF}$, $\mathcal{S}_{DT,DF}$ and $\mathcal{S}_{\Im}$ domains. Table 8 summarizes the chosen parameters in each transformation.

$\mathcal{S}_{DF}$	**Time Duration**	**Sample Frequency**				
	0.4999[Sec]	10[Khz]				
$\mathcal{S}_{DF}$	**FFT Points**					
	1024					
$\mathcal{S}_{DT,DF}$	**FFT Points**	**Window Type**	**Window Length**	**Everlap Samples**		
	512	Hamming	128	127		
$\mathcal{S}_{\mathfrak{I}}$	**Harmonic Frequency FFT Points**	**Window Type**	**Window Length**	**Everlap Samples**	**Modulation Frequency FFT Points**	**Envelope Detector Type**
	128	Hamming	128	0	128	Magnitude

Table 8 – Various Transformations Chosen Properties (Magnitude Envelope Detector)

3.1.1.1 Analysis of the various transformations performance

Figure 17 presents the signals' waveform, DFT , $DSTFT$ and DMT .

AM signal properties aren't easily stated in the $\mathcal{S}_{DT}$ domain (the waveform). The superposition of the AM signals causes the beats of each individual signal to be unnoticeable and the total beat rate to appear as very low (the total pattern repeating itself has a very low frequency). Hence, attempting to perform a separation of some type whose outcome is the signals' sub-components seems rather unpractical.

In the $\mathcal{S}_{DF}$ domain (the DFT) one can easily notice the harmonic frequencies, however the AM properties are left unveiled and latent. Hence, assuming a speech signal is much denser in spectrum, causing the harmonic frequencies to be indistinguishable, leaves us left with the inability to decompose the signal into its sub components.

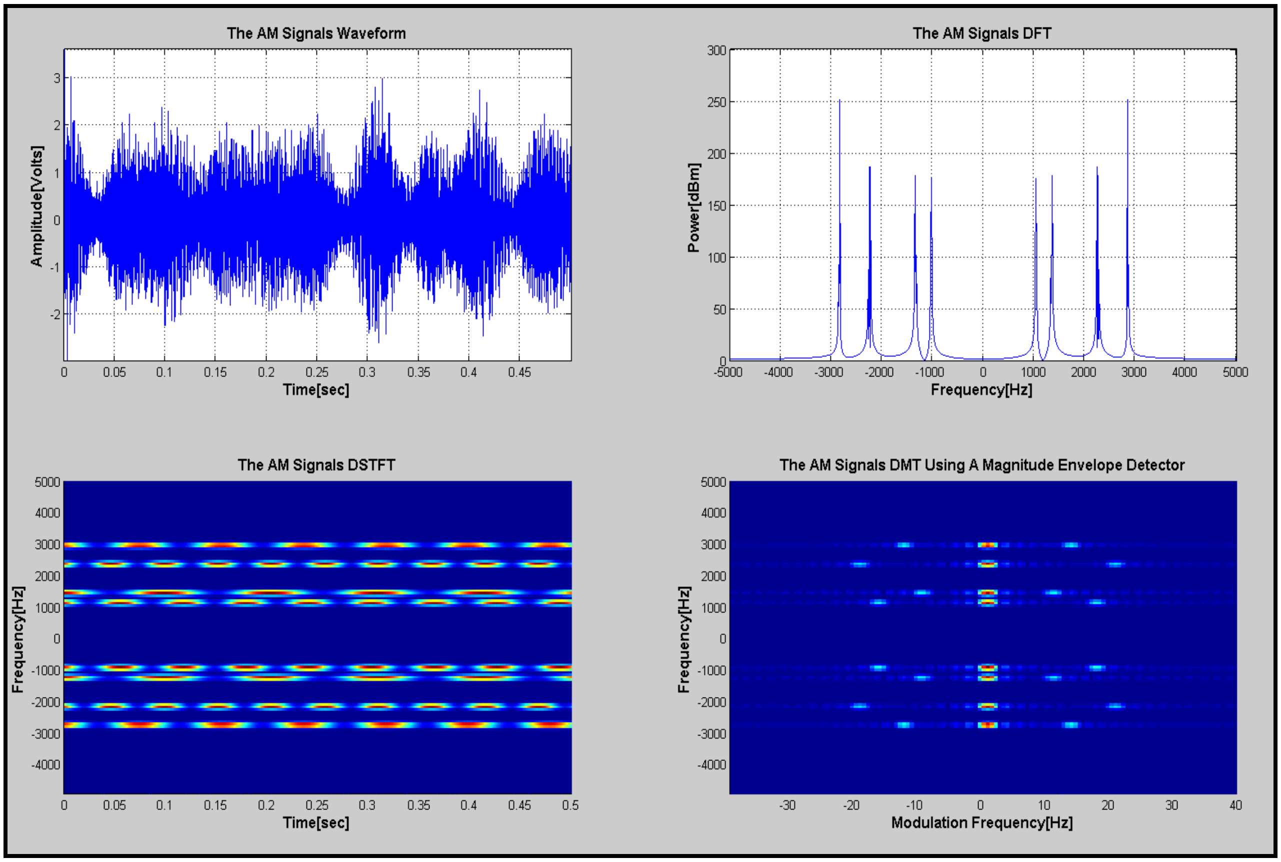


Figure 17 – DMT using Magnitude Envelope Detection Performance Evaluation

In the $S_{DT,DF}$ domain the AM nature of signal is easily noticeable but the exact modulation frequencies are left indeterminable. The analysis of the beat rate of the signal in this domain may perhaps facilitate the exact identification of the signal's modulators' frequencies but such a task seams rather unnecessary given the DMT s' outcome.

In the $S_{\Im}$ domain both the exact modulation frequencies and the carriers' frequencies are easily distinguishable. The careful observer may notice the DC components which arose. In the signal we transformed a DC component does exist, however due to an artifact of the Magnitude Envelope Detector we should have expected a DC component even if the original signal did not include one. This artifact is that if we take a signal with no DC component (time average of zero) and operate the Magnitude Envelope Detector on it, we gain a DC component. For signals with low frequencies (i.e. close to DC) this artifact may be intrusive.

In order to perform an enhancement of some type one may devise a signal mask in the $S_{\Im}$ domain working as a modulator frequency filter.

One may also take into account that the Magnitude Envelope Detector isn't reversible or at least isn't $1:1$ due to its un-linearity. This in turn means that in order to use it for separation we must save the Carriers in advance, in order to compose them together with the modified envelopes.

Smooth Hilbert Envelope Detector

Figure 18 displays the exact operation of the Smooth Hilbert Envelope Detector. By using a filter to smooth the signal's phase we get a close-to-linear slope which can then be used to derive the signal's slope in order to derive the signal's carrier frequency. Choosing a strong low-pass filter will yield a slope which converges to zero and is insensitive to the time phase variations. Choosing a weak low-pass filter (or wide LPF) will cause the Detector to follow strongly after the phase variations, hence being unable to extract the signals' carrier. Thus a smart choice of the filter kernel must be taken into consideration.

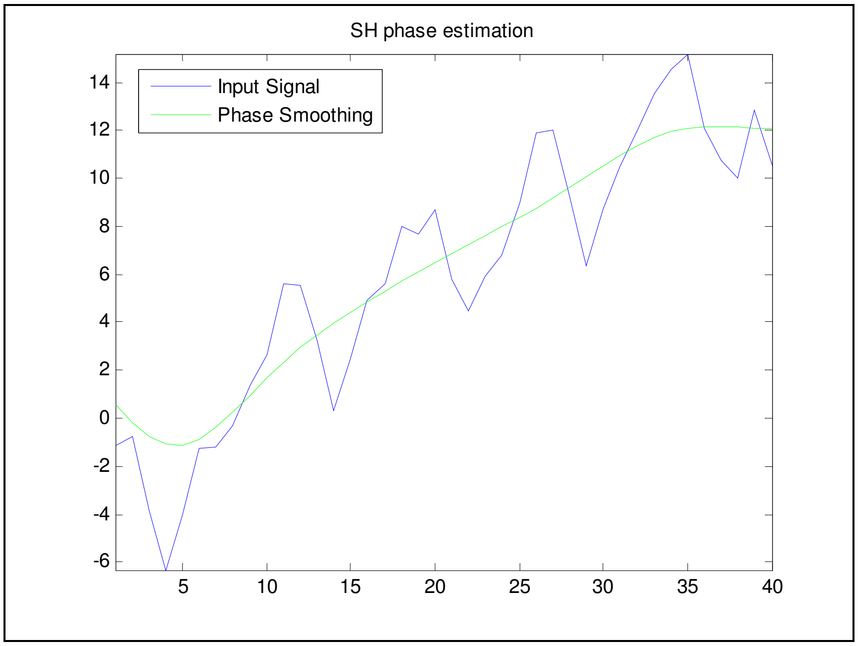


Figure 18 – Carrier estimation using Smooth Hilbert Envelope Detection

Figure 19 displays one's ability to identify the carriers' frequencies using the Smooth Hilbert Envelope Detector. An imperfect choice of the filter, as was made here, shows the fact that in comparison to the Magnitude Envelope Detector the width of the lobes representing the Detected Carrier is rather wide.

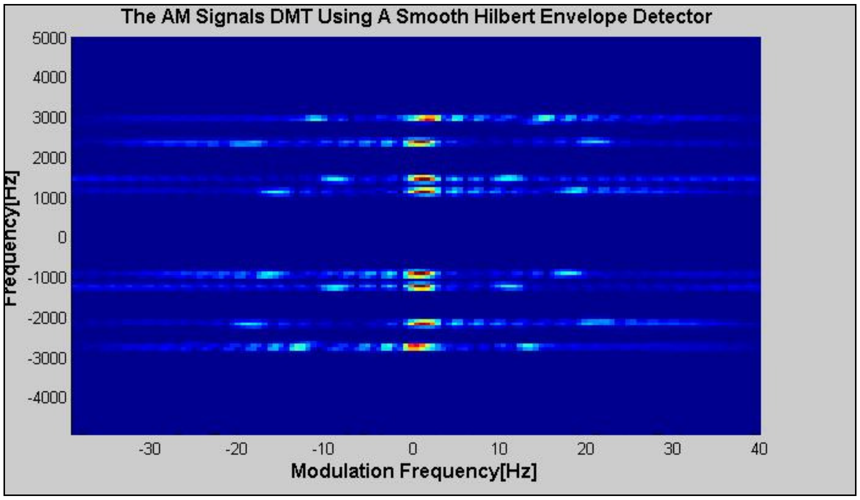


Figure 19 – DMT using the Smooth Hilbert Envelope Detector

Instantaneous Frequency Detector

Figure 20 displays the Instantaneous Frequency Envelope Detectors' ability to identify the carriers' frequencies. A currently not-understood phenomenon creates a "flip" of the frequency axis in some of the bins.

Comparing the various Envelope Detectors' outcomes as displayed in Figure 21 displays the phenomenon clearly.

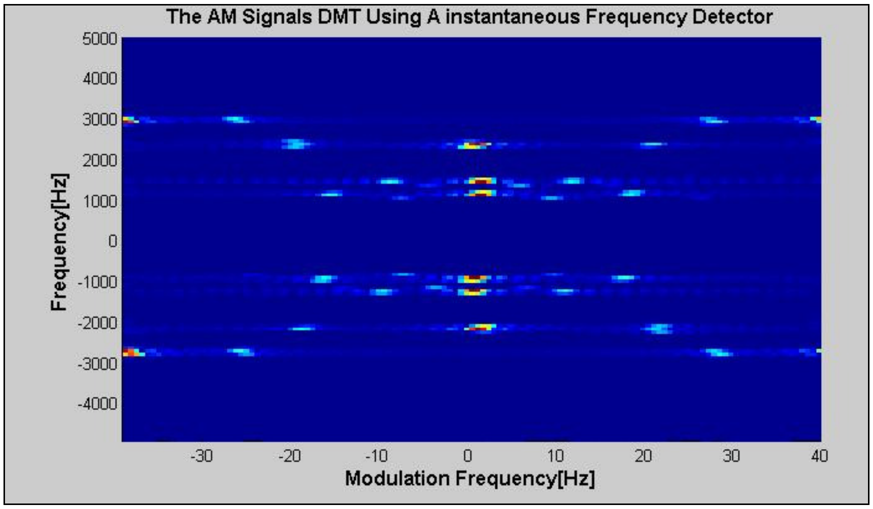


Figure 20 – DMT using the Instantaneous Frequency Detector

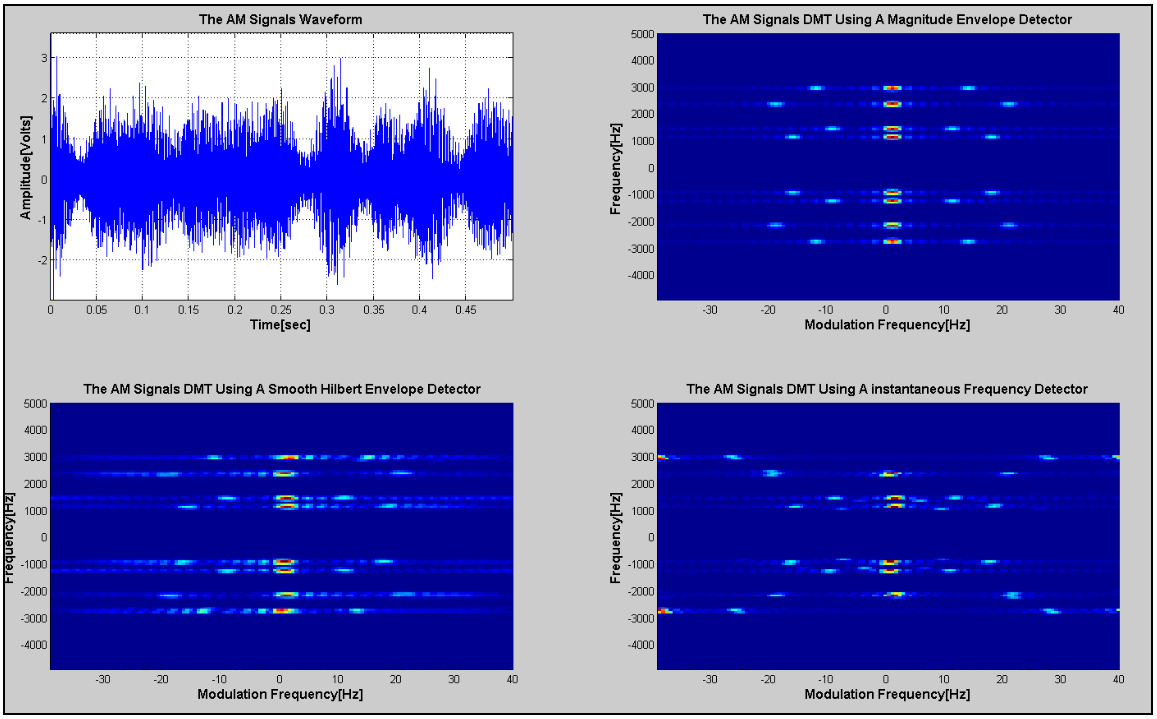


Figure 21 – DMT using the various Envelope Detectors

3.2 DMT Usage

The next section shows various cases of signal transformations to the JF Domain. The purpose of these examples is to demonstrate the ability and necessity of the newly proposed domain and methods to filter and/or enhance a signal.

A good example of a naturally modulated signal is a metronome. Spectral decomposition of the modulation content of "pulses" of the metronome yields a multitude of Fourier coefficients. And indeed transforming a metronome's signal using the DMT to the JF Domain can be seen in Figure 22. Is it easily identifiable that the multitude of constantly decaying spectral content in the modulation frequency axis (shown by the X axis in this figure). The Y axis suggests the Wide Bandwidth of the signal harmonic spectral content. Notice that Y axis ranges up to 220 KHz whereas the Modulated spectral content axis X ranges up to a mere 22Hz. Now using this signal representation one can attempt to employ various types of masks either in the JF domain or use the new data about the signal in order to employ different masks in the Time-Frequency domain. One may also make a use of the new data acquired in the JF domain in order to attempt to estimate various signal properties such as the modulation frequency of the metronome. This ability suggests the ability to easily filter noise models such as CSN. Such a type of noise is common when recording has been done in proximity to various types of mechanical machines.

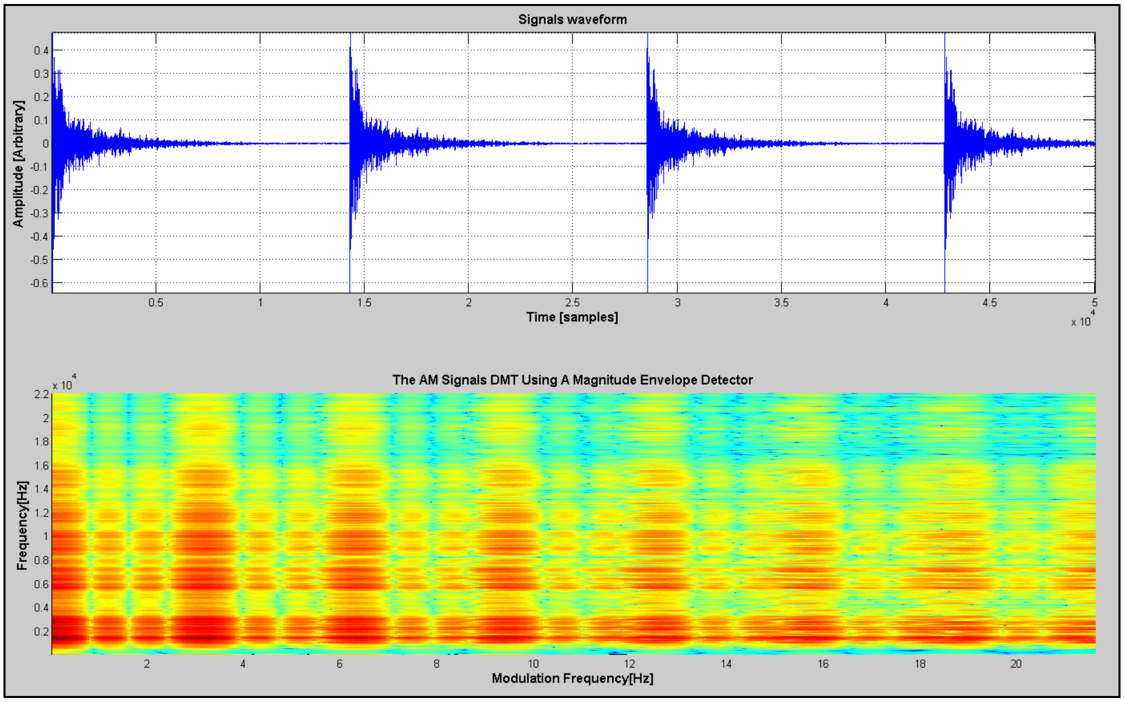


Figure 22 – DMT of a metronome

As stated in previous sections, in the framework of this project we model the human speech system as an AM-modulated one. Hence, the need to test the JF domain's ability to filter/separate simple AM signals is obvious. And indeed employing a simple binary mask in the JF domain as is depicted in Figure 23 proves the potential of the newly proposed domain to separate between AM-modulated signals. The left upper plot depicts the original signal composed of three AM-modulated sines. The careful eye can notice 12 lobes in the bottom left corner displaying the DMT of the original signal (not counting the DC products). We find 12 lobes due the symmetric properties of the Fourier Transform. Now in the lower right corner we employ a simple binary mask by setting the unwanted signal to zero, after which we transform the signal back to the time domain. Finally the desired single AM-modulated signal can be found in the upper left corner of the figure.

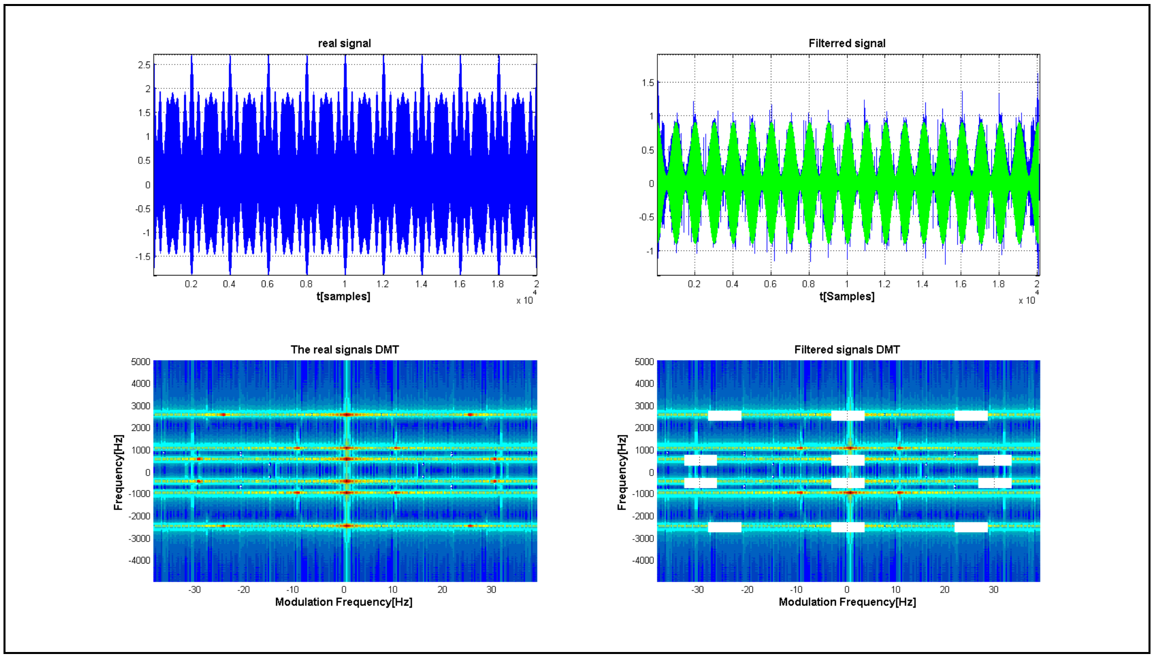


Figure 23 – DMT of a superposition of three AM signals

Another assumption stated in previous sections is our usage of CSN a reference model to AM-modulated noise. Hence, we would like to test the DMT's ability to easily depict such noise. Figure 24 shows that the CSN behavior in the different domains. The figure (especially the DMT representation) displays the fact that this noise is composed of white noise modulated at a frequency of about $5Hz$. We can now begin suggesting various automatic techniques to extract the exact modulation frequency of the noise. Such a task would be much more difficult using past-defined tools such as the STFT or simple Fourier Transform.

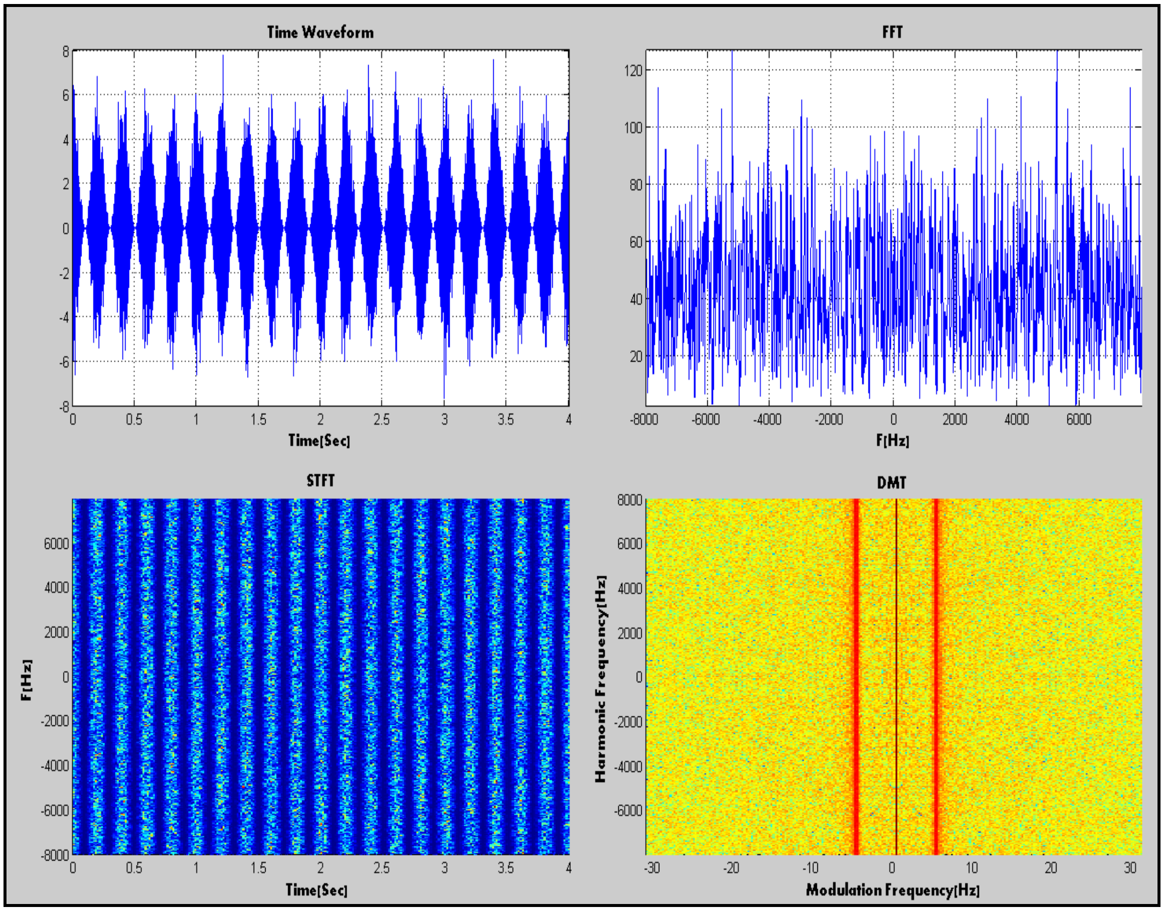


Figure 24 – WGN Cyclo-Stationary Noise in various domains

3.3 Enhancement Attempts

3.3.1 Filtering/Masking in the JF domain

A naïve and simple approach would be to try to filtering/mask the unwanted spectral content by using the DMT in order to employ some kind of mask in the Joint Frequency Domain. By masking or changing the numerical values that appear in the modulation frequency bins where we expect to find the WGN CSN we can attenuate the unwanted spectral content.

Figure 26 shows a signal in which an audio sample of woman is interferred by CSN as suggested in 0. Figure 27 shows the signal after employing a simple binary mask in the JF domain. Examining the signal's time representation after the filtering attempt clearly shows the ability to filter some of the unwanted spectral content, the CSN. Examining the signal's STFT also shows that the relative amplitudes of the speech are much stronger. This fact is depicted by the new color scale which suggests that the speech spectral content is much "warmer".

Calculating the filtered signal's SNR shows only a small improvement as shown in Table 9. We assume this is caused due to newly created artifacts. A subjective sound test confirms this assumption.

In some cases the desired signal has a very narrow-band spectral content in the JF domain. In such cases we may prefer filtering out the desired signal and finaly subtracting between the produced output and original output.

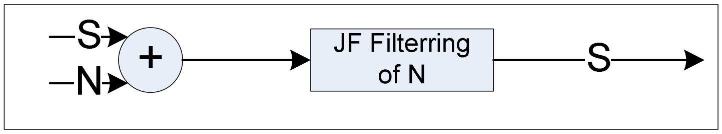


Figure 25 – JF Masking Conceptual Block Diagram

Figure 28 shows another example of the JF masking method to separate between two mixed signals. The mixed signal in this case was composed of two Metronomes which tap at different modulation frequencies. We then detect one metronome and its harmonies by analysing the column widths and their frequencies. We expect harmonies of the same metronome to appear in constant multiples of a basic frequency. After detecting them we mask the undesired columns. The time waveform of the filtered signal shows the ability to separate between the two signals.

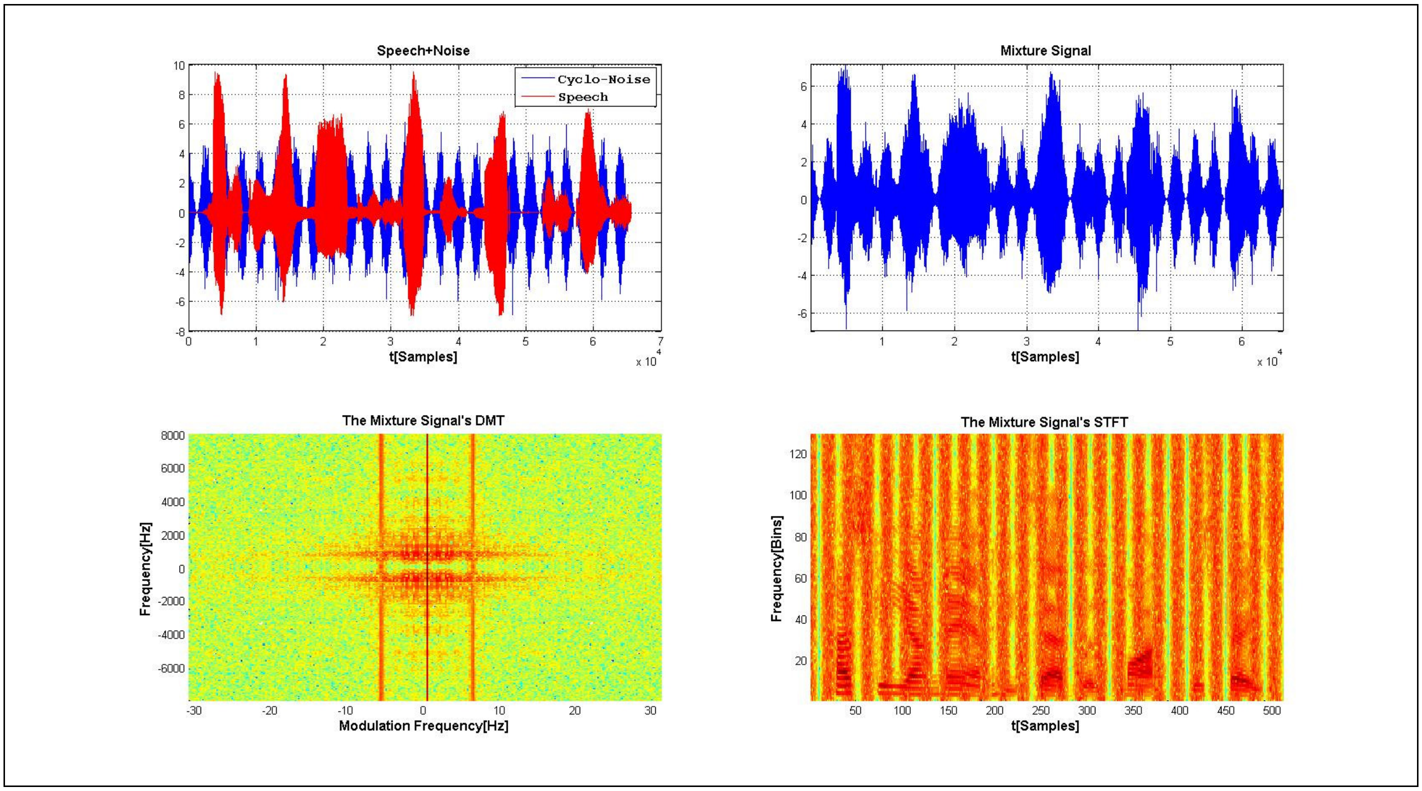


Figure 26 – Masking in the JF Domain – CSN+Speech signal

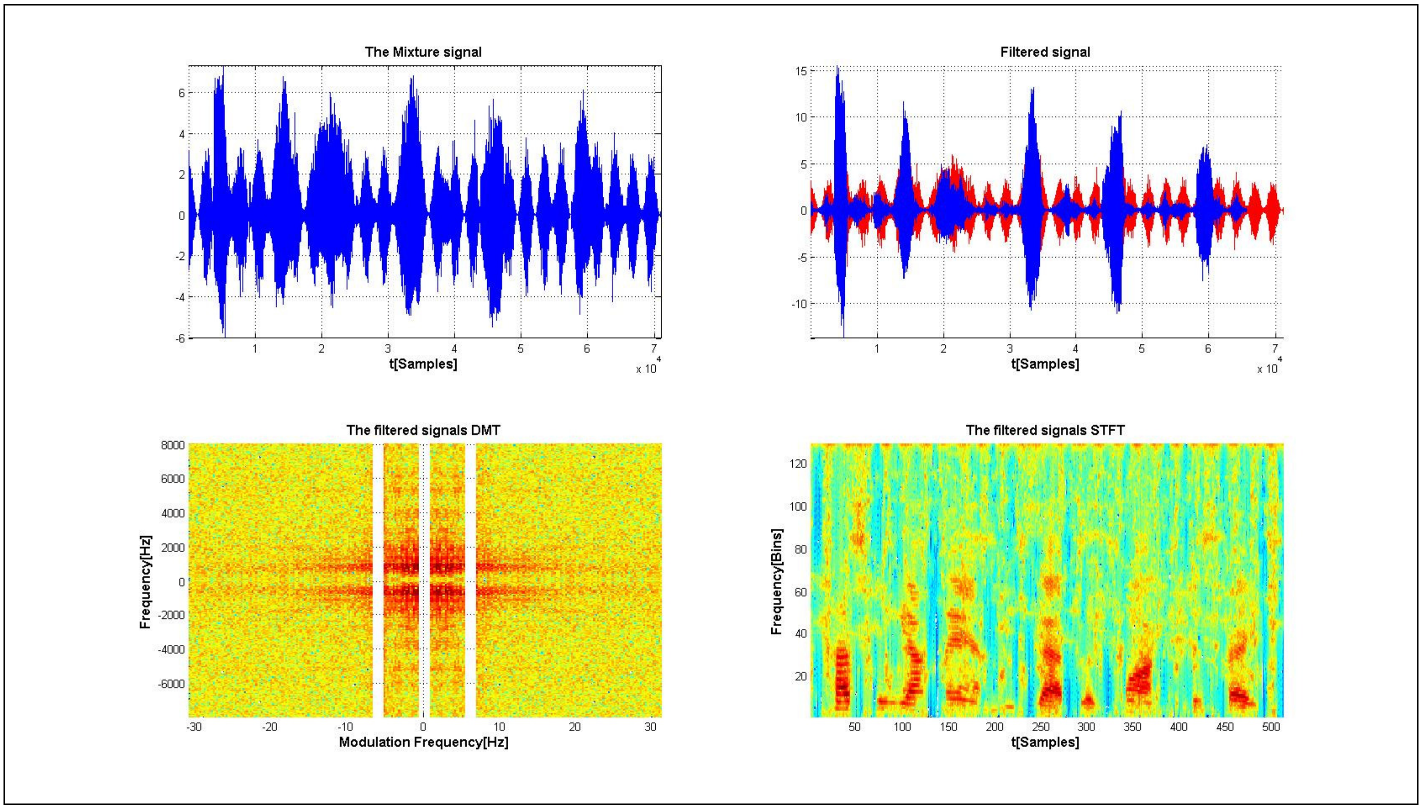


Figure 27 – Masking in the JF Domain - Filtered – CSN+Speech filtered signal

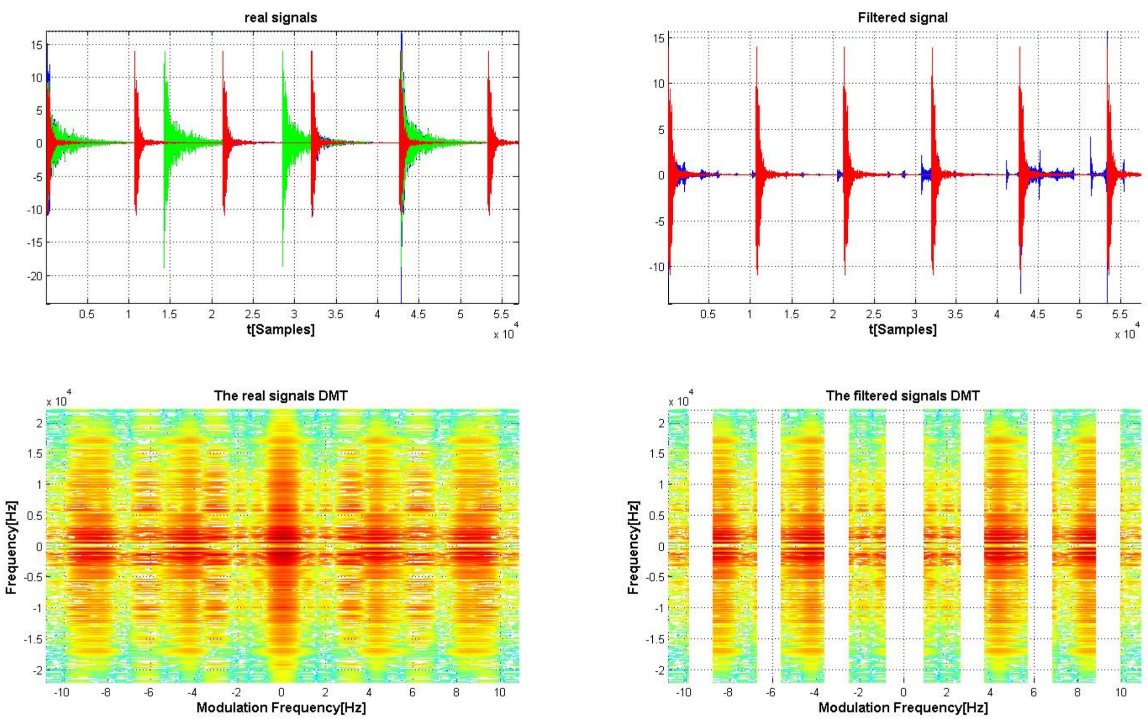


Figure 28 – Separation of two Metronomes by JF Masking

We can attempt to filter the unwated spectral content by masking the wanted spectral content and finaly by subtracting the outcome from the original mixture. This seems to be an effective method for minimizing artifacts when the wanted spectral content is of a narrow-nand nature in the JF domain (i.e. a signal which consists of a relatively small number of different AM signals).

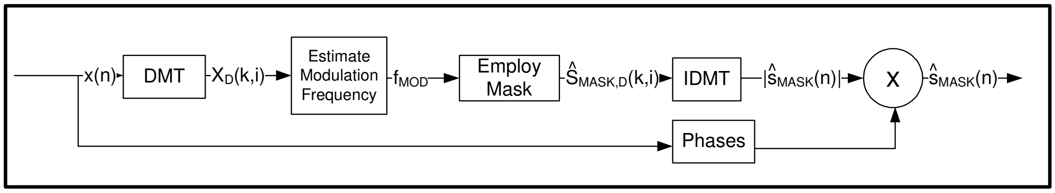


Figure 29 – JF Masking Conceptual Block Diagram

Figure 30 shows an attempt to separate between an AM signal (which we consider as the noise in this case) and speech. Due to the fact that the AM signal is of a very narrow-band nature in the JF domain we prefer attempting to separate between the signals by masking the AM signal and not the speech. And indeed the time waveform of the separated signal shows the ability to separate between the mixed signals.

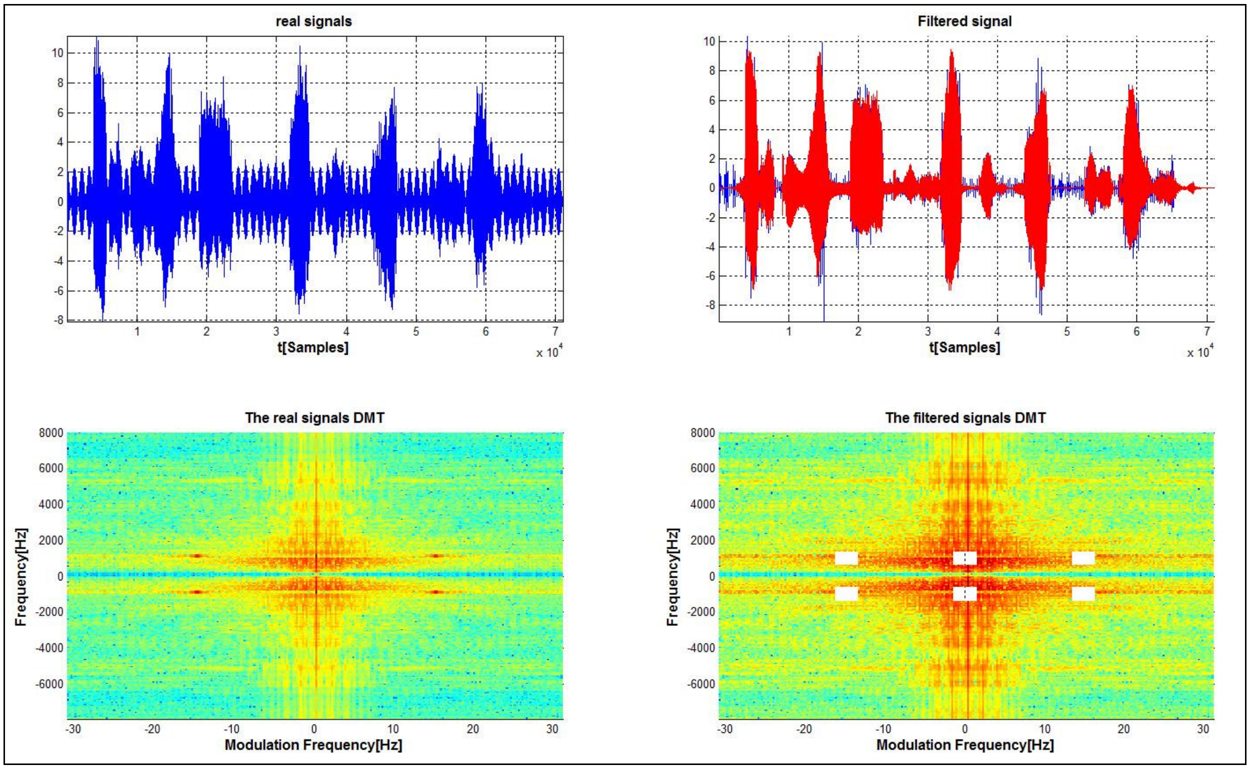


Figure 30 – Masking in the JF Domain – Speech+AM signals

3.3.2 Subtraction of AM signals

Another method we have tried in order to filter out undesired spectral content is to try to immitate a signals spectral content in the JF domain. Then generate an AM signal with a simmilar amplitude and phase and subtract between the two in time or in the STFT domain. This approach was tested on simple AM signals and was found very sensitive to the exact phase and frequency estimation. Estimation of the wrong phase and/or modulation frequency simply yielded a new spectral content in a different frequency.

3.3.3 Maximum Likelihood Amplitude Estimator

Due to the fact that all of the methods described above proved very partial success due to artifacts caused by the actions performed in the Joint-Frequency domain, we tried a new approach, which was to estimate desired signals in the STFT domain. This method also uses the DMT and the Joint Frequency domain, however in an indirect manner and as a tool in the estimation process.

The basic idea behind this method is to take an observation of speech interfered by CSN, to estimate the different parameters of the noise, and to use this knowledge in order to build a Maximum Likelihood Estimator (MLE) for the speech amplitude in the STFT domain. We will first present a block diagram of the entire process, and in the following sections we will describe in detail what is done in each step in the diagram. The diagram is presented in Figure 31.

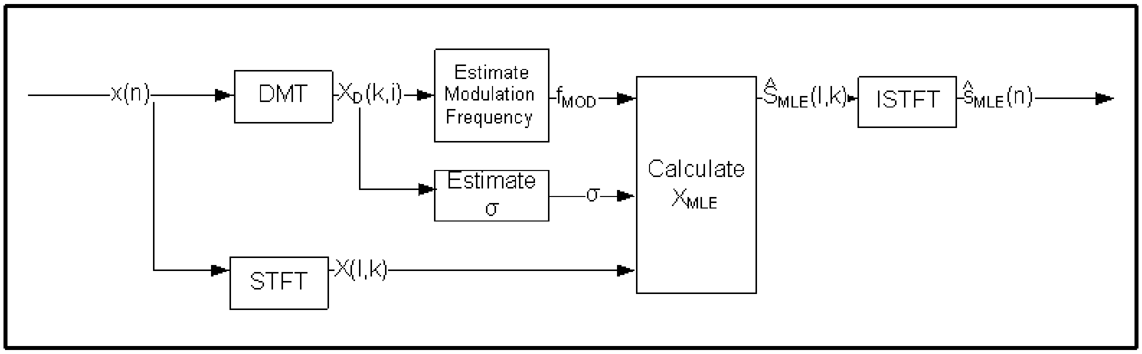


Figure 31 – A block diagram presenting the MLE algorithm

It should be noted that in all of our work in this method, we assumed that the CSN's initial phase is zero and should not be estimated. In addition, we should note that in this method we will estimate only the speech signal's amplitude (see further discussion in the next section).

3.3.3.1 Mathematical foundations

We will begin by presenting the explicit calculation of the speech MLE. For this section, we assume that the CSN parameters have been estimated and are now known to us, and we will show how they should be used in order to estimate the speech signal.

We can describe the observation signal we have in the time domain ($x(n)$) as a sum of the speech ($s(n)$) and the CSN ($b(n)$):

$$x(n) = s(n) + b(n)$$

The mathematical model of our noise implies directly that it will be transformed into an additive Gaussian noise in the STFT domain (this is due to the fact that a real-valued Gaussian in the time domain is Fourier-transformed into a complex-valued Gaussian signal in the frequency domain). We will assume that the CSN and speech signals are statistically independent (which is a very reasonable assumption to make), thus we may assume that for each STFT bin, regardless of the times and frequencies dealt with, the speech and CSN's variances may be summed to receive the observation's variance:

$$\sigma_X^2(l,k) = \sigma_S^2(l,k) + \sigma_B^2(l,k)\ ,$$

and also,

$$\left|X\left(l,k\right)\right|^2=\left|S\left(l,k\right)\right|^2+\left|B\left(l,k\right)\right|^2(1+\cos(\omega_{\text{mod}}l))^2$$

We would now like to estimate the speech out of the amplitudes of our observations in the STFT. We basically want to maximize $P(S\mid X)$, which is equal, by the Bayes theorem, to:

$$P(S\mid X)=\frac{P(X\mid S,\omega)P(S\mid\omega)}{P(X\mid\omega)}$$

Taking the speech and observation probabilites (given ω) as constants, we understand that:

$$\hat{S}_{ML}=\arg\max_S(P(X\mid S))=\arg\max_S(P(B+S\mid S))=\arg\max_S(P(B\mid S))$$

And in order to maximize this probability, we shall first describe it in terms of the speech and observation signals:

$$P_B=P\left(\sqrt{\frac{\left|X\left(l,k\right)\right|^2-\left|S\left(l,k\right)\right|^2}{\left(1+\cos(\omega l)\right)^2}}\right)$$

And now differentiate it, using the chain rule, to receive:

$$0=\frac{dP_B}{dB}\frac{\partial(X-S)}{\partial S}=\left(\frac{-\left|S\right|\left(u\left(S\right)-0.5\right)}{\left(1+\cos(\omega l)\right)\sqrt{\left|X\right|^2-\left|S\right|^2}}\right)\frac{d}{dB}\left(\frac{Be^{-\frac{B^2}{2\sigma^2}}}{\sigma^2}\right)=-\frac{1}{\sigma^2}\left(\frac{\left|S\right|\left(u\left(S\right)-0.5\right)}{\left(1+\cos(\omega l)\right)\sqrt{\left|X\right|^2-\left|S\right|^2}}\right)\left(e^{-\frac{B^2}{2\sigma^2}}+Be^{-\frac{B^2}{2\sigma^2}}*\frac{-2B}{2\sigma^2}\right)=$$

$$=-\frac{1}{\sigma^2}\left(\frac{\left|S\right|\left(2\bullet u\left(S\right)-1\right)}{\left(1+\cos(\omega l)\right)\sqrt{\left|X\right|^2-\left|S\right|^2}}\right)e^{-\frac{B^2}{2\sigma^2}}\left(1-\frac{B^2}{\sigma^2}\right)$$

where u is the Heaviside function. The only relevant solution for this equality is:

$$1-\frac{B^2}{\sigma^2}=0$$

And this implies that:

$$B(l,k)=\pm\sigma$$

$$\sqrt{\frac{\left|X\right|^2-\left|S\right|^2}{\left(1+\cos(\omega l)\right)^2}}=\pm\sigma$$

$$\hat{S}(l,k)=\sqrt{\left|\hat{X}(l,k)\right|^2-\sigma^2(1+\cos(\omega_{\text{mod}}l))^2}$$

And this is the estimation for the speech amplitude in each STFT bin.

As for the phases estimation, two things should be noted – first of all, throughout the mathematical process we have done here, we assumed that the initial phase of the CSN is zero. An MLE for non-zero initial phase can be developed as well, and a very similar experssion will be obtained:

$$\hat{S}(l,k)=\sqrt{\left|\hat{X}(l,k)\right|^{2}-\sigma^{2}(1+\cos(\omega_{\mathrm{mod}}l+\phi))^{2}}$$

Secondly, we did not try to estimate the phases of the speech signal, but chose to use the phases of the observation. The phases of the observation were proved to be the MSE estimation for the phase of the enhanced speech, by Ephraim and Malah [4] -.

3.3.3.2 Modulation Frequency Estimation

In order to estimate the modulation frequency of the CSN we use one of the objective measures defined earlier, i.e. the DMT Norm Distance.

This norm here is marked as ΔCSN and is the integration of the spectral distance between a manually generated CSN and a given signal –

$$\begin{aligned}&X \triangleq input\ signal\\&B \triangleq generated\ \ CSN\\&\Delta CSN=\sum_{\substack{entire\\matrix\\values}}\left|X_D-B_D\right|\end{aligned}$$

Now, by applying common minima search algorithms such as "Newton-Raphson" we can generate various $B's$ and search for the minimal attainable norm with the modulation frequency of B as the independent paramter. This search method is basically equivalent to differentiation, and is based on the idea of "best matching" of the places where the vertical lines, which characterize the CSN in the JF domain, fall.

3.3.3.3 CSN Variance Estimation

The calculation of the noise variance is also done in the JF domain, and uses the estimated value of f_{mod}, the CSN's modulation frequency. After the modulation frequency has been estimated, and we have found where the are the CSN-characteristic lines in the JF domain, we can perform the following calculation:

$$\sigma=\frac{1}{LK}\sum_{k}\left|X_D(l_{f_{MOD}},k)\right|$$

This way of calculating the CSN's standard deviation is equivalent to an alternative calculation which we could perform, which would be to average on the intensity of the noise in the STFT domain. The two calculations are equivalent due to the fact that averaging the peaks of the noise in the STFT domain is "translated", via the Fourier Transform, into averaging over a specific vertical line in the DMT matrix (i.e. the line containing the CSN spectrum). This is simply because the cosine function is transformed into an impulse modulation frequency function in the JF domain, whereas the harmonic frequencies remain unchanged.

3.3.3.4 Recorded Noise and the Modified MLE Enhancement Approach

A look at the synthesized noise's spectrogram as shown in Figure 40 gives us a good idea as to what should the subtracted envelope resemble. One would suggest a subtraction envelope which resembles the one given in Figure 33. This signal has the exact same modulation frequency, and an amplitude which bestly fits the White Gaussian CSN Variance which was present in the synthesized noise.

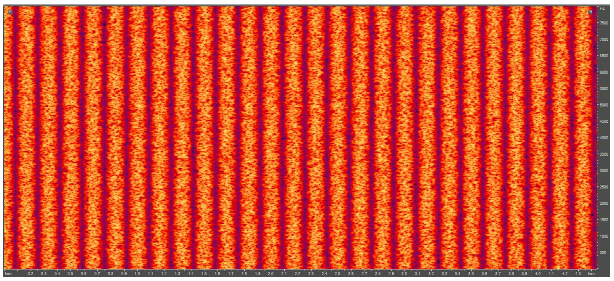


Figure 32 – Synthesized Noise Spectrogram

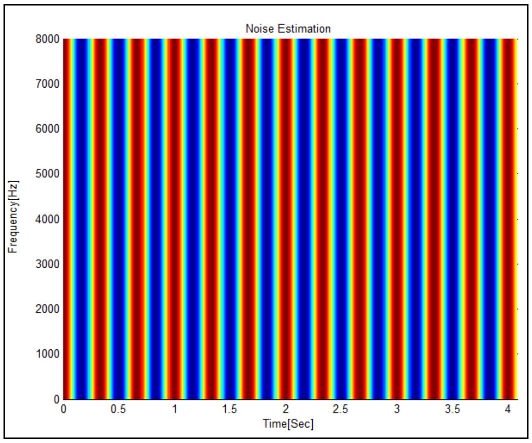


Figure 33 – Synthesized Noise – desired envelope signal

Now let us look at the recorded noise spectrogram. What "shape" of subtraction envelope would we want to use in order to enhance such an observation. A reasonable guess would be the one depicted in Figure 35. A general approach to an extimation of a "good" envelope noise signal for subtraction would be to look at the noise's STFT and try to estimate its envelope as a function of time. We may also use the intensity of the noise as a function of time in order to estimate the noise's variance as a function of harmonic frequency (the noise is now not truly white).

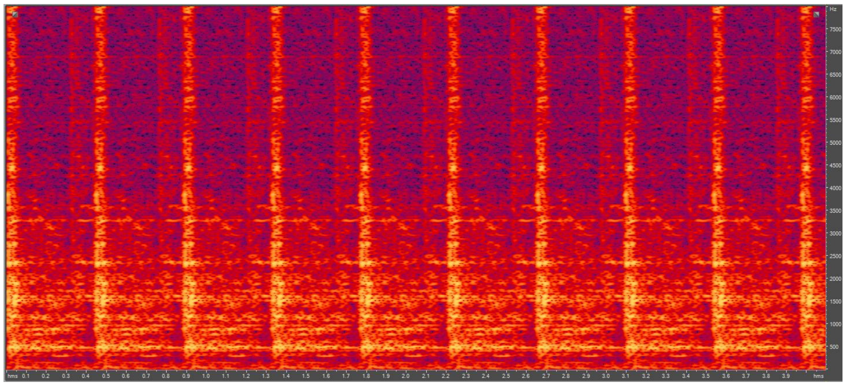


Figure 34 – Recorded Noise Spectrogram

The envelope of the recorded noise we have used in our work resembles a rectangular pulse. Hence we have "modified" the subtraction signal to have an envelope of a rectangular pulse with a 20% duty-cycle. We expanded the rectangular pulse to the fifth order using a fourier series –

$$Modified\ MLE\ Envelope = 1 + \sum_{i=1}^{\infty} \frac{1}{\pi^i} \sin\left(\pi i D\right) \cos\left(2\pi i f_{\mathrm{mod}} l\right)$$

We have then plugged in this term instead of the former term of -

$$MLE\ Envelope = 1 + \cos\left(2\pi f_{\mathrm{mod}} l\right)$$

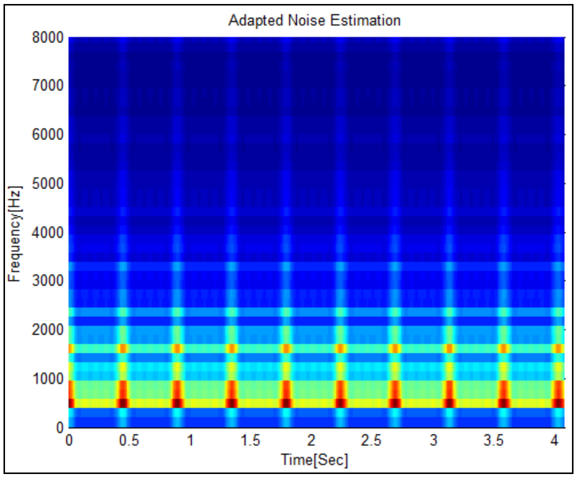


Figure 35 – Recorded Noise – suggested envelope signal

4 Experimental Results

The results given in this section display the algorithms ability to enhance a specific choice of an audio file taken from the TIMIT resources. These tests were done on a 4 more audio files and gave similar results.

Synthesized Noise Enhancement Results

In order to test the enhancement abilities on sythesized CSN, we used the setup described in Figure 36. Our input speech signal was an audio recording of a woman from TIMIT (the woman was saying the sentence "ask a young man, she said with mock distaste"). We synthesized random noises in Matlab, in different modulation frequencies, and we calculated SNR and LSD values according to the definitions given in previous sections.

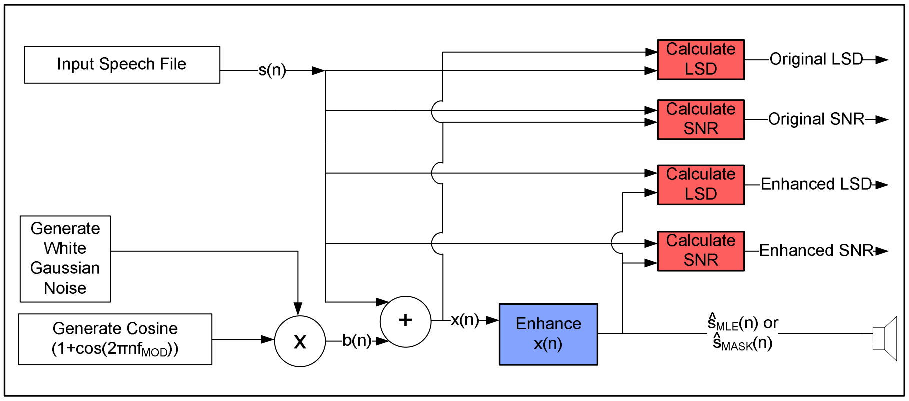


Figure 36 – Test Setup for Sythesized Noise Signals

Table 9 summarizes our results. The results present the improvement (Δ) in the SNR and LSD after enhancement compared to the input SNR and LSD as given in the first row.

	ΔSNR	ΔLSD
Observation Wave File	$0dB$	$13.91dB$
Masking Enhanced Wave File	$\Delta = 1.77dB$	$\Delta = 6.75dB$
MLE Enhanced Wave File	$\Delta = 6.92dB$	$\Delta = 7.35dB$
OM-LSA Wave File	$\Delta = 5.18dB$	$\Delta = 3.8dB$

Table 9 – Synthesized Noise Objective Measures Performance Analysis

It can be seen from the table that the MLE beats both the OM-LSA enhancement algorithm and the masking attempts in the JF domain.

Another interesting thing to check is the dependence of the objective measure improvements on the intesity of the speech in the observation (which may be measured and noted as SNRin). Following are the graphs depicting this behaviour:

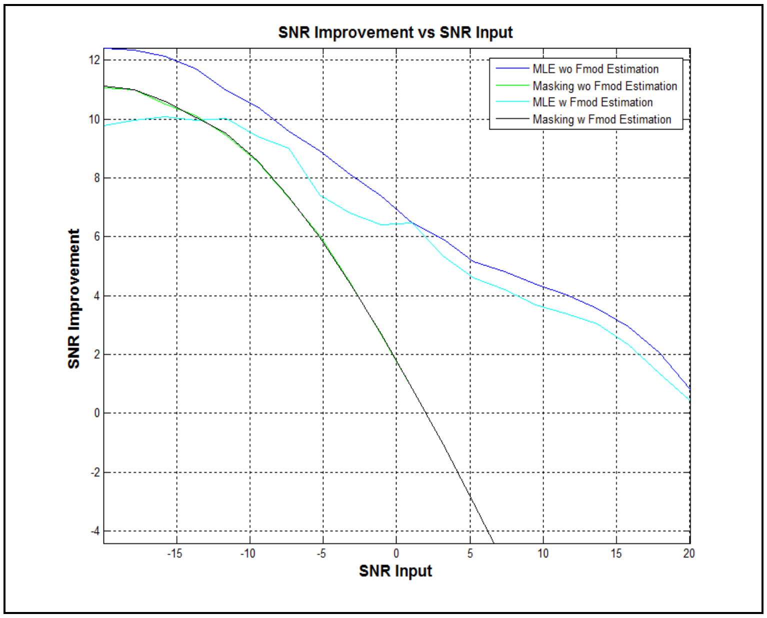


Figure 37 – Synthesized Noise – SNR Improvement vs. SNR Input

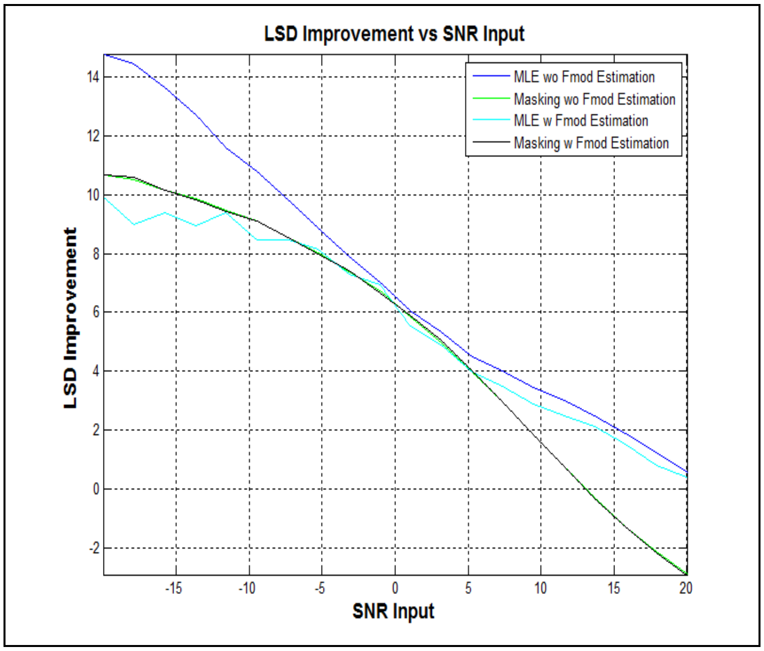


Figure 38 – Synthesized Noise – LSD Improvement vs. SNR Input

These graphs were plotted by calculating the average enhancement result out of 50 noise draws. Further insights on the enhancement process may be obtained from looking at the various signals involved in the STFT and JF domains:

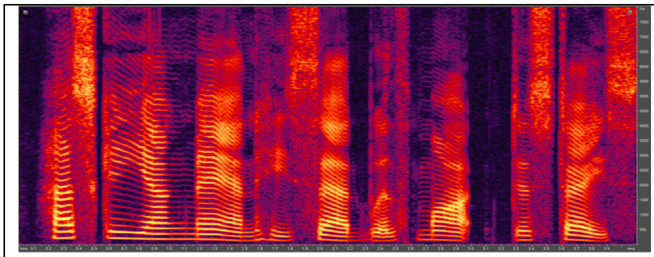
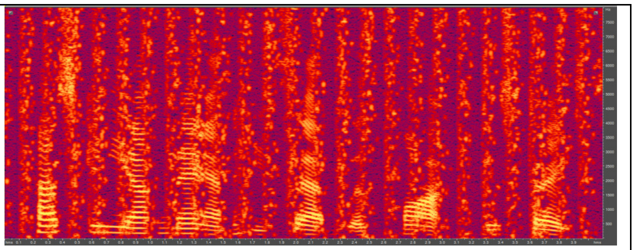
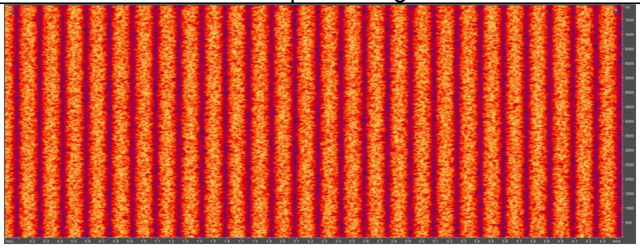
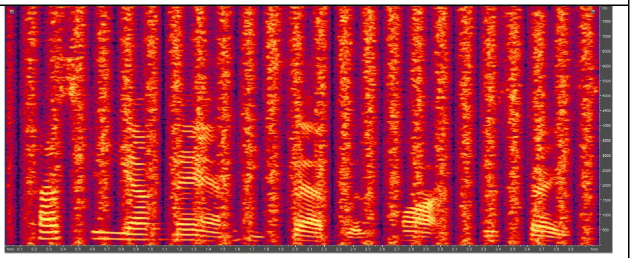
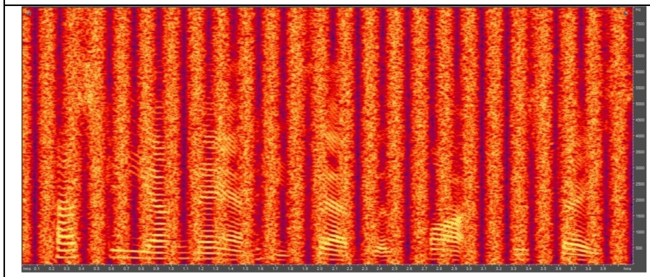
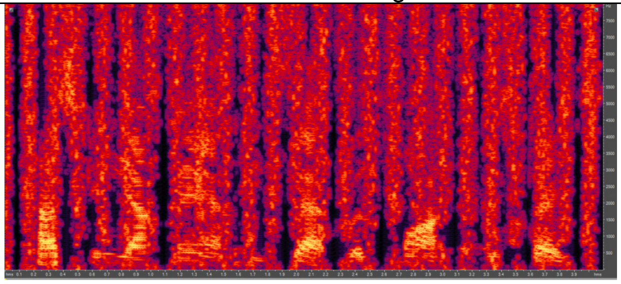
 Clean Speech Signal	 OM-LSA Enhanced Signal
 Noise Signal	 MLE Enhanced Signal
 Noisy Signal	 Masking Enhanced Signal

Table 10 – Synthesized Noise Enhancement - Spectogram Performance Analysis

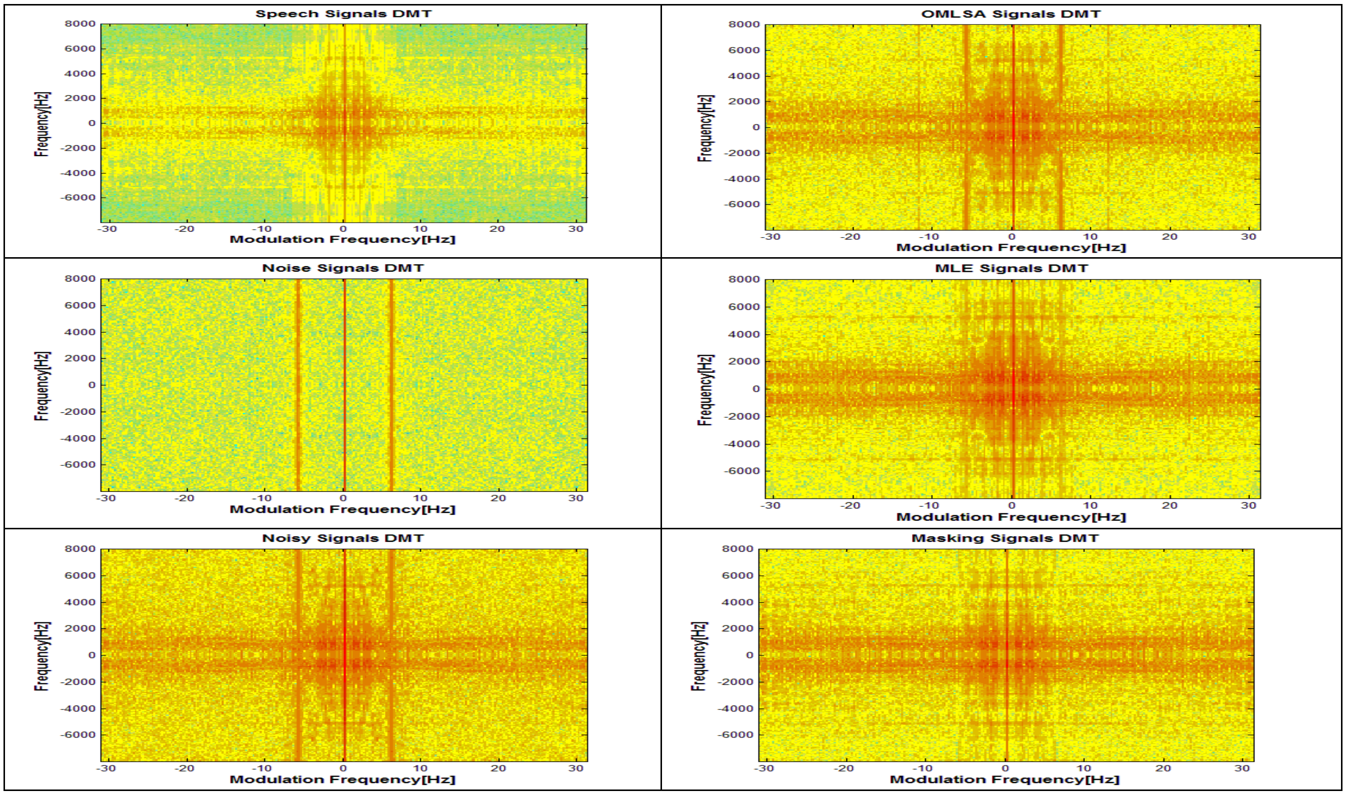


Table 11 – Synthesized Noise Enhancement – DMT Performance Analysis

One might notice that Masking is insensitive to the estimation of fmod. This is due to the size of the masking rectangles.

Masking is misleading. It seems like the best method in the JF domain but returning to the time domain proves otherwise. This is why we preferred working in the STFT domain.

Recorded Noise Enhancement Results

Testing our algorithms on real-life noises is, of course, much more interesting than testing them on mathematically-synthesized noises. Such tests require small modifications in our setup. The new setup is depicted in the following figure:

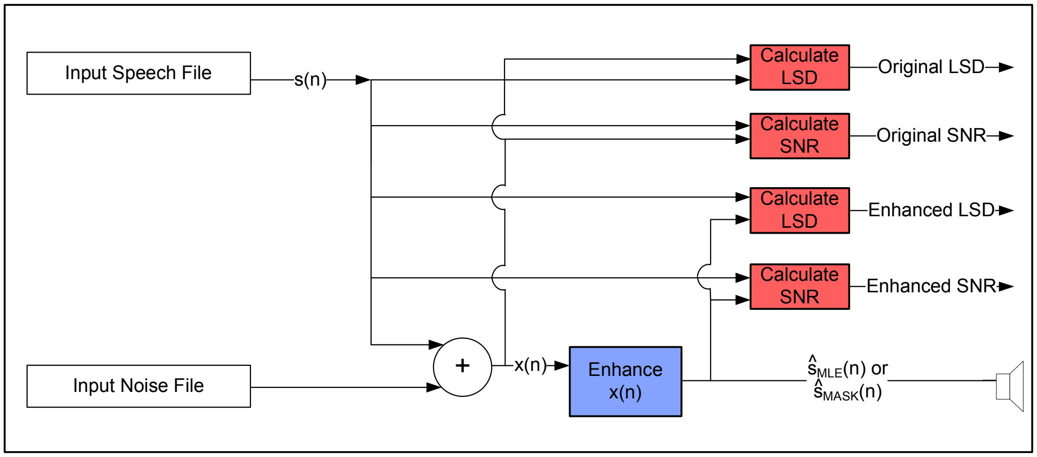


Figure 39 – Test Setup for Recorded Noise Signals

As a real-life noise we chose a recording of a Cesna engine during a start up process. Following is the recorded noise's intensity in the time domain:

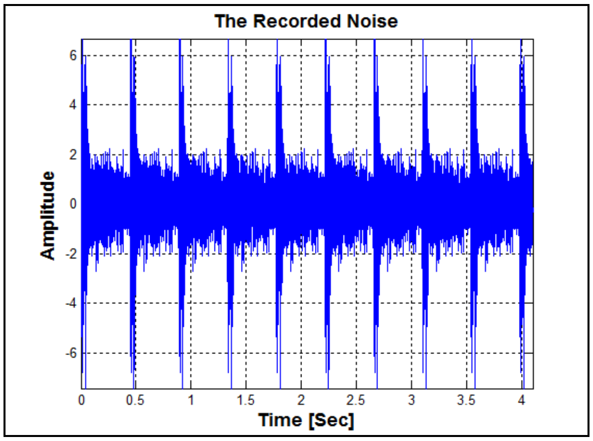


Figure 40 – Recorded Noise – Waveform

Again, enhancement results and the signal's JF and STFT waveforms are presented. It can be seen that for this kind of signal, our MLE performance is not as good as for the synthesized CSN but still gives SNR improvement results which are close the OM-LSA. The LSD results of Masking are still the best but listening to the Wave files proves otherwise. The Wave file after Masking are very distorted and contain a vast amount of artifacts.

The following table summarizes our results. The results present the improvement (Δ) in the SNR and LSD after enhancement compared to the input SNR and LSD as given in the first row.

	ΔSNR	ΔLSD
Observation Wave File	$0dB$	$11.06dB$
Masking Enhanced Wave File	$0.97dB$	$3.06dB$
MLE Enhanced Wave File	$1.92dB$	$1.95dB$
OM-LSA Wave File	$2.07dB$	$2.88dB$

Table 12 – Recorded Noise Objective Measures Performance Analysis

Looking at the spectrograms of the various enhancement algorithms as depicted in Table 13, one can again notice the good ability of the MLE algorithm to enhance the signal and supress the noise. In our opinion the spectrograms of the MLE are the best looking ones, in terms of least noise. Again the JF Domain may seem misleading. Table 14 suggests that the MLE is the bestly enhanced signal whereas the DMT figures suggest otherwise.

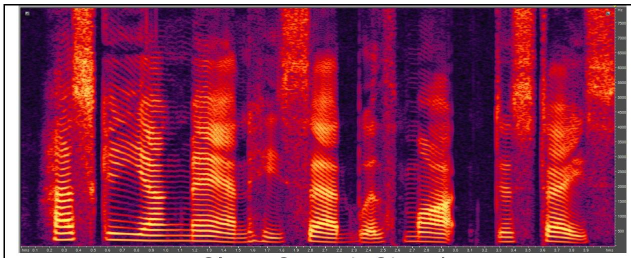
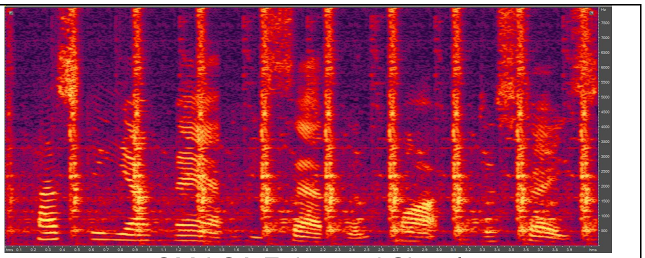
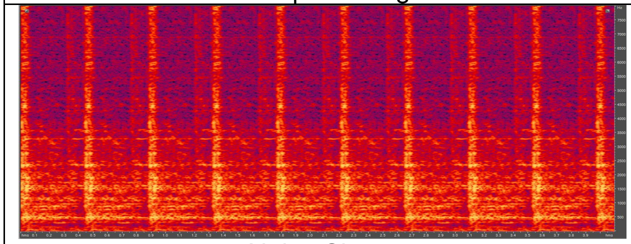
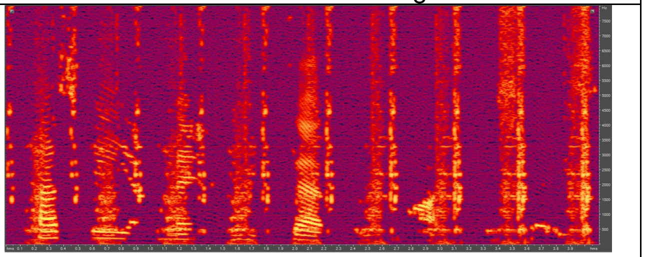
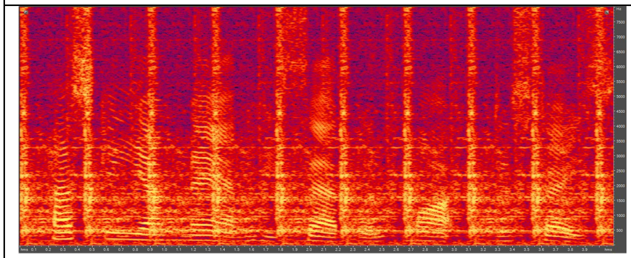
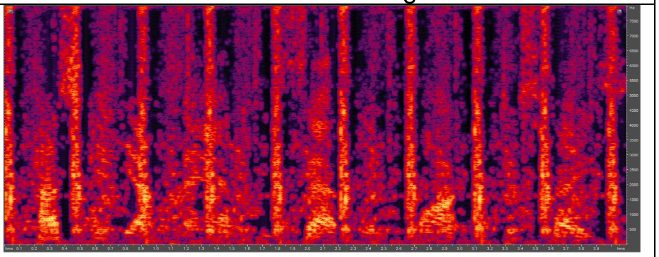
 Clean Speech Signal	 OM-LSA Enhanced Signal
 Noise Signal	 MLE Enhanced Signal
 Noisy Signal	 Masking Enhanced Signal

Table 13 – Recorded Noise Enhancement - Spectogram Performance Analysis

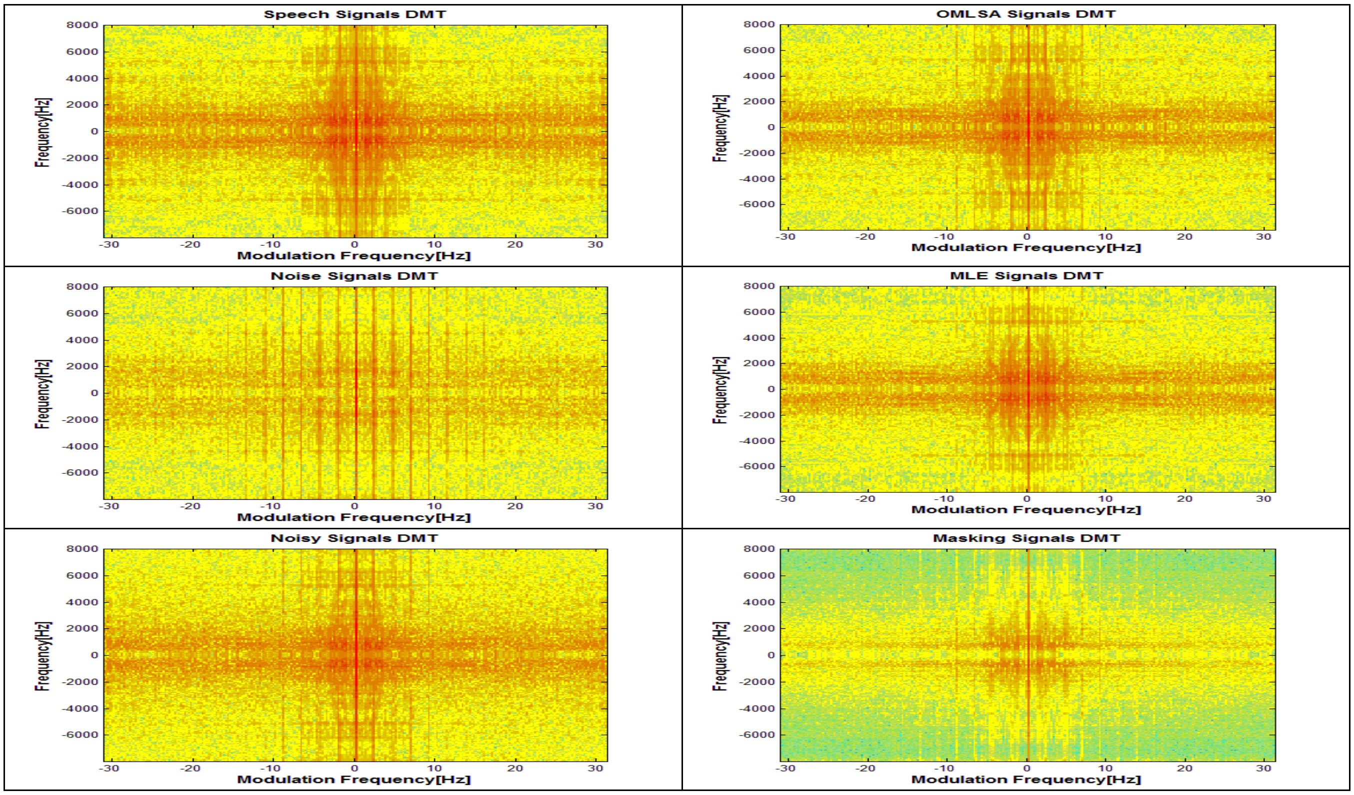


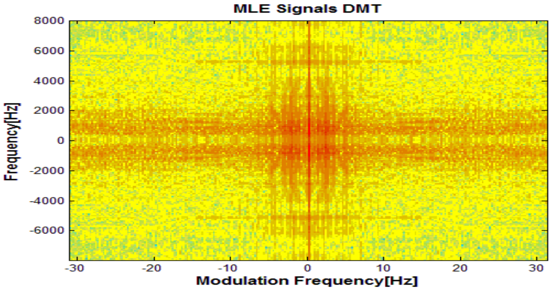
Table 14 – Recorded Noise Enhancement – DMT Performance Analysis

As stated in section 3.3.3.4 we have attempted to modify the MLE approach to better suite "real-world" noises. Table 15 summarize the objective measures results of the mMLE vs. MLE and OM-LSA. We can see that there is a clear gain in using a modified envelope to serve as the estimation of the noise's envelope. We now beat OM-LSA in SNR and have gained an extra $0.8dB$ in LSD.

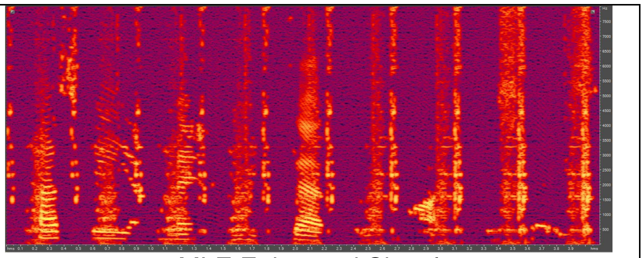
	ΔSNR	ΔLSD
MLE Enhanced Wave File	$1.92dB$	$1.95dB$
mMLE Enhanced Wave File	$2.63dB$	$2.75dB$
OM-LSA Wave File	$2.07dB$	$2.88dB$

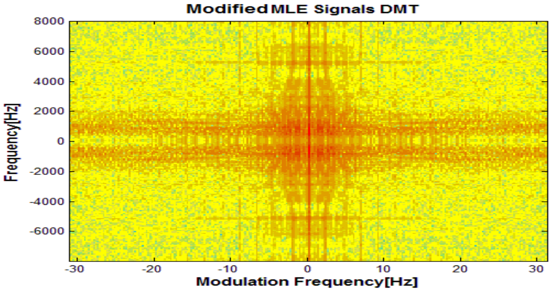
Table 15 – Recorded Noise Objective Measures Performance Analysis

Table 16 depicts the comparison of spectrograms and DMT plots of the mMLE vs. MLE. The spectrograms clearly depict the better ability of the mMLE approach the enhance the signal.

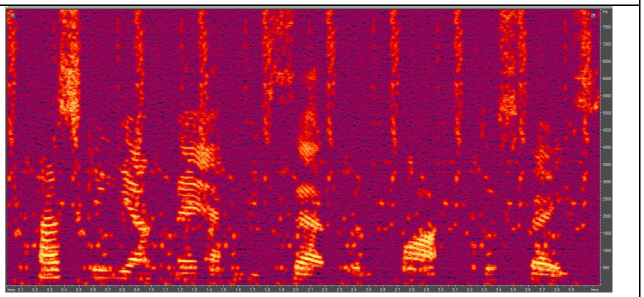
<table>
<tr>
<td>


MLE Signals DMT

</td>
<td>


MLE Enhanced Signal
</td>
</tr>
<tr>
<td>


Modified MLE Signals DMT

</td>
<td>


mMLE Enhanced Signal
</td>
</tr>
</table>

Table 16 – Recorded Noise Enhancement – mMLE vs MLE Spectogram and DMT performance analysis

5 Conclusions and Future Work Directions

Usage of the suggested algorithms for enhancement

Both of the two objective measures and subjective noise tests prove the ability of the MLE algorithm to enhance a signal stamped with CSN. We do not recommend usage of the Masking algorithm. This approach is naïve and the objective measures results are deceiving. Moreover, the subjective sound tests prove unequivocally that the MLE approach is superior to masking. Spectrograms clearly show the enahncement of the MLE approach. The noise signal is suppressed after applying the MLE algorithm which may be seen in both synthesized noise spectrograms and recorder noise spectrograms.

Using solely the DMT as a domain for both estimating the noise parameters and enhancing is a naïve approach and may yield to a vast amount of artifcats. We suggest using the DMT as a tool for signal parameter estimation. Moreover, we suggest using the DMT as a tool for estimating CSN parameters, which can in turn be plugged in other enhancement algorithms such as OMLSA.

As depicted by the figures in this work, one may notice that the measure of enhancing is deceiving when looked at the DMT plots. The noised seems to be well attenuated whereas returning the signal to the time domain and/or STFT proves otherwise. This in turn explains why LSD measures should better be calculated in the STFT domain rather than in the JF domain.

Figure 37 and Figure 38 show that above certain SNR inputs one may prefer not to use these algorithms. These algorithms cause degradation of the SNR rather then any improvement.

The mMLE approach may be a basis for a more flexible approach for enancing speech signals. By using this approach together with STDMT's (which is spoken about in section 0) we can "modify" the subtracted envelope in order to "fit" for the specific noise. Using this method we can perhaps suite our algorithm to work on noises which aren't cyclo-stationary.

A noticable artifact caused by the MLE and mMLE approach is the raising of the noise floor. Observing the figures in Table 10 show that the time regions which were rather clean prior to the enhancing attempt now have a "warmer" shade. This is an inherent product caused by the subtraction which is used in the MLE and mMLE approaches. When we subtract a region which was clean prior to the subtracion we now have in turn inserted unwanted noise. It is well known that MLE approaches aren't best suited for enhancing speech which is corrupted by WGN (and/or Cyclo-stationary WGN). Hence, other approaches such as MMSE, Weiner or Log spectral amplitude are suggested.

Another intersting phenomena one may notice is that the masking algorithm has an intersting ability to attentuate low frequency noise. This specific ability may be of some interest to certain application.

Future work directions

Using the JF as a domain for non stationary noise parameter estimation

Another approach would be to use the JF domain as a "tool" for the sole purpose of estimating the non-stationary signal's parameters. We can then "plug" these parameters into other algorithms such as OM-LSA and by this enjoy the benefits of both worlds. By this we practically enable the ability of algorithms which aren't well suited for enhancement of non-stationary noises to enhance signals which suffer from such noises. As stated in section 0 MLE approaches aren't best suited for enhancing speech which is corrupted by WGN (and/or Cyclo-stationary WGN). Hence, other approaches such as MMSE, Weiner or Log spectral amplitude are suggested. Finally, using the DMT as a tool for noise parameter estimation can be a basis for expanding OMLSA's abilities in order to suite non-stationary noises.

Using the other envelope detectors

In the scope of this work we have made use solely of the Magnitude Envelope Detector. Continuing this research may require exploring the results given by using the various envelope detectors. This may prove a better ability to estimate the modulation frequency and/or the noise variance. This may in turn improve the enhanced signal's SNR and/or LSD and/or subjective noise tests.

Moving to STDMT

Real-life noise signals do not necessarily have a constant modulation frequency and/or a constant variance. Hence, in order to adapt the algorithm to varying noise signals one must move to using the Short time DMT. By this the ability to "update" the modulation frequency and/or variance estimation when there is a need is enabled. This may also be found to improve the SNR results of a constant modulation frequency noise. This is due to the fact the the modulation frequency estimation is never exact and hence the subtraction signal suffers from a cumulating phase differnce between the real noise obsevation and the estimated noise signal. By constantly updated the noise parameters such as phase we can improve the sensitivity to such phenomena.

6 Bibliography

[1] - Arons B. A Review of the Cocktail Party Effect. Retrieved October 6 2008, from: http://xenia.media.mit.edu/~barons/pdf/arons_AVIOSJ92_cocktail_party_effect.pdf

[2] - Atlas L. E. & Janssen C. (2005). Coherent Modulation Spectral Filtering for Single-Channel Music Source Separation, ICASSP 2005, 461-464. Retrieved September 20 2008, from: http://www.ee.washington.edu/research/isdl/papers/atlas-2005-icassp.pdf

[3] - Cohen I. and Berdugo B. (2001). Speech Enhancement for Non-Stationary Noise Environment, Signal Processing, 81, 2403-2418.

[4] - Ephraim Y. and Malah D. (1984). Speech Enhancement Using a Minimum-Mean Square Error Short-Time Spectral Amplitude Estimator, IEEE Transactions on Acoustics, Speech and Signal Processing, 32, 1109-1121.

[5] - Johansson M. The Hilbert Transform. (Masters Dissertation, Växjö University, Växjö) Retrieved September 29 from: http://w3.msi.vxu.se/exarb/mj_ex.pdf

[6] - Porat, B. (1997). A Course in Digital Signal Processing. New York: Wiley.

[7] - Rabiner L.R. & Schafer R.W. (1978). Digital Processing Of Speech Signals. Engelwood Cliffs, N.J.: Prentice Hall.

[8] - Schimmel S. M. (2007). Theory of Modulation Frequency Analysis and Modulation Filtering, with Applications to Hearing Devices. (PhD. Dissertation, University of Washington, Seattle, 2007). Retrieved September 20, 2008 from: http://isdl.ee.washington.edu/people/stevenschimmel/publications/dissertation.pdf

[9] - Schimmel S. M., Atlas L. E. and Nie K. (2007). Feasibility of single channel speaker separation based on modulation frequency analysis, ICASSP 2007. Retrieved September 20 2008,from: http://isdl.ee.washington.edu/people/stevenschimmel/publications/SchimmelAtlasNie_ICASSP2007.pdf

[10] - Traunmüller H. Evidence for Demodulation in Speech Perception. Retrieved October 6 from: http://www.ling.su.se/STAFF/hartmut/demod.pdf

7 Appendix

Implemented Matlab Code

Henceforth, one can find the crucial functions implemented during this work

STFT Function

```
function Y = stftm(x,N,M,swin,iscomplex,no_zero_freq_shift,Nfft)
    if nargin<2, N = 512; end;
    if nargin<3, M = floor(N/2); end;
    if nargin<4, swin = hamming(N); end;
    if nargin<5, iscomplex = false; end;
    if nargin<6, no_zero_freq_shift = false; end;
    if nargin<7, Nfft = N; end;

    if isempty(swin), swin = hamming(N); end;
    if isempty(iscomplex), iscomplex = false; end;
    if isempty(iscomplex), iscomplex = false; end;
    if isempty(no_zero_freq_shift),no_zero_freq_shift=false;end;
    x = x(:);
    swin  = swin(:);
    awin = biorwin(swin,M);

    xf = buffer(x, N, N-M, 'nodelay');
    xf = diag(awin)*xf;

    % Y = sqrt(N)*ifft(xf, [], 1);
    Y= fft(xf, Nfft, 1);

    if ~any(any(imag(x))) && ~iscomplex
        Y = Y(1:Nfft/2+1,:);   %use the symetric character in the freq. domain
        w = (0:2*pi/Nfft:pi)';
    else
        w = (0:2*pi/Nfft:2*pi/Nfft*(Nfft-1))';
    end
    % Y = conj(Y);

    if ~no_zero_freq_shift
        n  = 0:M:M*(size(Y,2)-1);
        phaseFix = exp(-j*w*(n+Nfft));
        Y = Y.*phaseFix;
    end
```

Inverse STFT Function

```
function x = istftm(Y,N,M,swin,iscomplex,no_zero_freq_shift)
    if nargin<2, N = 512; end;
    if nargin<3, M = floor(N/2); end;
    if nargin<4, swin = hamming(N); end;
    if nargin<5, iscomplex = false; end;
    if nargin<6, no_zero_freq_shift = false; end;

    if isempty(swin), swin = hamming(N); end;
    if isempty(iscomplex), iscomplex = false; end;

    N2 = N/2;

    if ~iscomplex
        if ~no_zero_freq_shift
            w = (0:2*pi/N:pi)';
            n  = 0:M:M*(size(Y,2)-1);
            phaseFix = exp(j*w*(n+N));
            Y = Y.*phaseFix;
        end

        Y(N2+2:N,:)=conj(Y(N2:-1:2,:));
        xf=real(ifft(Y));
    else
        xf=ifft(Y);
    end

    clear Y;

    xf = xf .* repmat(swin, [1 size(xf, 2)]);
    x = overlap_and_add(xf, M, false);
```

Overlap and Add Function

```
function y = overlap_and_add(frames, inc, average_weight)
    if nargin<3; average_weight=true; end;
    win = size(frames, 1);
    len = inc*(size(frames,2)-1)+win;
    y = zeros(len, 1);
    weight = zeros(len, 1);
    for k=1:size(frames, 2)
        offset = (k-1)*inc;
        curr_range = offset+1:offset+win;
        y(curr_range) = y(curr_range)+ frames(:,k);
        weight(curr_range) = weight(curr_range)+ ones(win, 1);
    end
    y(weight==0) = [];
    weight(weight==0) = [];

    if average_weight
        y = y ./ weight;
    end
```

DMT Function

```
function  [DataOut,ModFreqVec,HarmonFreqVec,ModFreqFs,Carriers] =
BSS_DMT(DataIn,HarmNFFT,ModNFFT,Win,OverLapLen,DetType,Fs)
%BSS_DMT - Implementation of DMT.
%
% Syntax - [DataOut,ModFreqVec,HarmonFreqVec] =
% BSS_DMT(DataIn,HarmNFFT,ModNFFT,Win,OverLapLen,DetType,Fs)
%
% Inputs:
%    DataIn - Data To Be Transformed
%    NFFT - FFT Block Size (power of two is advised)
%    Win - Temporal Window Vector
%    OverLapLen - Number Of Overlapping Samples Between Windows
%
%
% Outputs:
%    DataOut - The Transformed Data. The Output Matrix's Lines Are The
%                           Frequency Dimension Whereas The Columns Are the Time
Dimension
%    TimeVec - A Normalized Time Vector Ranging Between 0 and the Number of
%                          Whole Windows In DataIn.
%    FreqVec - A Normalized Frequency Vector Ranging Between -pi and pi
%
% Author: Omry Sendik
% Last revision: 1/11/08

DataTemp = stftm(DataIn,HarmNFFT,HarmNFFT-OverLapLen,Win,1,0);
HarmonFreqVec = -Fs/2+Fs/HarmNFFT:Fs/HarmNFFT:(Fs/2);
[CarrierOut EnvelopeOut] = BSS_EnvDet(DataTemp,DetType);
Carriers = CarrierOut;
DataOut = fft(EnvelopeOut,ModNFFT,2);
ModFreqFs = (Fs/(length(DataIn)/size(DataTemp,2)));
ModFreqVec = -ModFreqFs/2+ModFreqFs/ModNFFT:ModFreqFs/ModNFFT:ModFreqFs/2;
```

Envelope Detector Choice Function

```
function  [CarrierOut EnvelopeOut] = BSS_EnvDet(DataIn,DetType)
%BSS_EnvDet - Implementation of Various Envelope Detectors.
% syntax -  [DataOut] = BSS_EnvDet(DataIn,DetType)
%
% Inputs:
%    DataIn - Data Input
%    DetType - Envelope Detection Type
%
% Outputs:
%    DataOut - The Output Data.
%
% Author: Omry Sendik
% Last revision: 15/11/08
EnvelopeOut = DataIn*0;
CarrierOut = DataIn*0;

switch DetType
    case 'Mag' % Incoherent Magnitude
        EnvelopeOut = abs(DataIn);
        CarrierOut = (DataIn./EnvelopeOut).*(EnvelopeOut>=1E-8);
    case 'CSH' % Coherent Smooth Hilbert
         for index = 1:size(DataIn,1)
            [Envelope Carrier] = shce(DataIn(index,:),64);
            EnvelopeOut(index,:) = Envelope;
            CarrierOut(index,:) = Carrier;
         end
    case 'CIF' % Coherent Instantaneous Frequency
         for index = 1:size(DataIn,1)
              [Carrier Envelope] = BSS_IFCohEnvDet(DataIn(index,:));
             EnvelopeOut(index,:) = Envelope;
             CarrierOut(index,:) = Carrier;
         end
    otherwise
        disp('Syntax Error - Unknown Detection Type');
end;
```

Smooth Hilbert Envelope Detector

```matlab
function [Carrier,Envelope] = BSS_IFCohEnvDet(DataIn)
% BSS_IFCohEnvDet - This Function Implements the Coherent Instantaneous
%                                              Frequency Envelope Detector
%
% Syntax:  [Carrier,Envelope] = BSS_IFCohEnvDet(DataIn)
%
% Inputs:
%    DataIn - Data Input
%
% Outputs:
%    Carrier - The Signals' Carrier
%    Envelope - The Signals' Envelope
%
%
%
% Author: Omry Sendik
% Last revision: 14/11/08
DataIn = DataIn(:).';

I = real(DataIn);
Q = imag(DataIn);

ZI = I(1:end-2).*I(3:end) + Q(1:end-2).*Q(3:end);
ZQ = I(1:end-2).*Q(3:end) - Q(1:end-2).*I(3:end);
Z = ZI + i.*ZQ;
Z = [Z(1) Z Z(end)];

Thresh = 1E-5;

alpha = zeros(1,length(Z));
alpha(1) = ((Z(1)/abs(Z(1)))^0.5-1)*(abs(Z(1))>Thresh) + 1;

alpha(2:end) = ((Z(2:end)./abs(Z(2:end))).^0.5).*((abs(Z(2:end))) > (Thresh)) +...
                                alpha(1:end-1).*(abs(Z(2:end)) <= Thresh);

N = 512;
lpwin = hamming(N)'/sum(hamming(N));
alpha = conv2(alpha, lpwin, 'same');
alpha = alpha./abs(alpha);


Carrier = zeros(1,length(alpha));
Carrier(1) = alpha(1);
for index = 2:length(Carrier)
    Carrier(index) = Carrier(index-1).*alpha(index);
end


Envelope = DataIn.*conj(Carrier);

%%% Filter The Envelope
%
h=firls(100,[0 0.6  0.7 1],[1 1 0 0]);
Envelope = conv2(Envelope,h,'same');

end
```

Instantaneous Frequency Envelope Detector

```
function [a c] = shce(X1, N, W)
    if nargin<3, W=17; end;
    Xbuf = buffer(X1, N);
    a = zeros(size(Xbuf));
    c = a;
    t=(0:N-1)';
    f = hamming(W);
    f = f./sum(f);

    phi = unwrap(angle(Xbuf), [], 1);
    omega_m = mean(diff(phi, 1, 1));
    omega_m_t = t*omega_m;
    phi_0 = ones(N, 1)*mean(phi - omega_m_t, 1);
    theta = phi - (omega_m_t + phi_0);
    theta_h = conv2(theta, f, 'same');
    c = exp(j*(omega_m_t+phi_0+theta_h));
    a = Xbuf.*conj(c);


    len = length(X1);
    c = c(1:len);
    a = a(1:len);
```

Inverse DMT

```
function  [DataOut] =
BSS_IDMT(DataIn,HarmNFFT,ModNFFT,Win,OverLapLen,HarmFs,ModFs,Carriers)
%BSS_DMT - Implementation of IDMT.
%
% Syntax - [DataOut,ModFreqVec,HarmonFreqVec] =
% BSS_DMT(DataIn,HarmNFFT,ModNFFT,Win,OverLapLen,DetType,Fs)
%
% Inputs:
%    DataIn - Data To Be Transformed
%    NFFT - FFT Block Size (power of two is advised)
%    Win - Temporal Window Vector
%    OverLapLen - Number Of Overlapping Samples Between Windows
%
%
% Outputs:
%    DataOut - The Transformed Data. The Output Matrix's Lines Are The
%                           Frequency Dimension Whereas The Columns Are the Time
Dimension
%    TimeVec - A Normalized Time Vector Ranging Between 0 and the Number of
%                          Whole Windows In DataIn.
%    FreqVec - A Normalized Frequency Vector Ranging Between -pi and pi
%
%
% Author: Omry Sendik
% Last revision: 1/11/08
DataOut = ifft(DataIn,ModNFFT,2);

DataOut = DataOut(:,1:size(Carriers,2));
DataOut = max(0,real(DataOut)).*Carriers(:,1:min(size(Carriers,2),size(DataOut,2)));
[DataOut] = istftm(DataOut,HarmNFFT,HarmNFFT-OverLapLen,Win,1,0);
```