INVERSE GABOR TRANSFORM FOR SPEECH ENHANCEMENT

Mohammed A. Al-Manie[1] and William J. Wang[2]
[1]Computer Research Institute, King Abdulaziz City for Science and Technology, Riyadh 11442, Saudi Arabia
[2]School of Engineering and Information Technology, University of Sussex, Brighton, BN1 9QT, England
malmanie@gmail.com

In this paper, the Inverse Discrete Gabor Transform (IDGT) is proposed for signal recovery buried in board-band non-stationary noise. Time-frequency masking filtering technique is implemented to reject the noise from corrupted speech while at the same time maintaining the desired waveform. A synthetic multicomponent non-stationary test signal made up of two chirps was first used to simulate noise; the signals were then separated using this technique. Four English speech signals recorded in different environments such as airport, restaurant, and train buried in wide-band noise were reconstructed. The extracted signals were then compared with the original ones in terms of cleanness and noise removal. The implemented procedure is suitable for this type of wide-band non-stationary interference, which cannot be canceled in the Fourier (frequency) or time domain.

INTRODUCTION

In the year of 1946 a physicist by the name Dennis Gabor proposed the decomposition of a continues-time signal into a set of shifted and modulated elementary discrete signals [1]. The spectrum in this case is obtained by multiplying the analysis signal by a Gaussian window of a chosen length, the Fourier transform is then, calculated for this particular function. The Gabor transform is a time-frequency distribution with a number of useful applications such as speech, seismic, and image-processing signals that have characteristics of a time-varying spectrum not appropriate to analyze in the Fourier transform method [2-4]. The main advantage of this technique is in the ability to detect frequency contents changes with respect to time, which may be very important in cases, such as medical imaging where diagnoses of abnormalities are related to time. Another example is detection of localized faults in mechanical and electrical systems that also vary with time [5,6]. This is in contrast to the classical Fourier analysis, which gives information about the overall spectral content of a particular signal without providing information about how these frequencies evolve in time.

In general, the discrete Gabor transform (DGT) can be thought of as a windowed Fourier transform, which provides a representation of a signal in time and frequency simultaneously. The Gabor coefficients are calculated as the inner product of the test signal and a single analysis window used to calculate the transform [7,8]. A major advantage that DGT offers is the ability to reconstruct the original signal. This is accomplished by summing the translated and modulated parts of the test signal obtained by the fixed synthesis window after being weighted by the Gabor coefficients. The Gabor transforms also offers the ability to change the time-frequency distribution (TFD) of a signal by adjusting the magnitude of the Gabor coefficients and recovering the original signal. This procedure is used as a time-varying filter for non-stationary signals that cannot be recovered in the time or frequency domain. The result of reconstruction can also be compared with the original value for an accuracy test of the Gabor transform representation.

In addition, the inverse discrete Gabor transform (IDGT) has been implemented by a number of researchers for various applications such as seismic de-convolution proposed by Margrave [9]. In this approach, a non-stationary filter was accomplished by modifying the Gabor time-frequency decomposition in order to recover the desired seismic signal. In another procedure of time-frequency synthesis, Xiang et al. [10] proposed an iterative time-varying filtering algorithm in the discrete Gabor domain. A non-stationary chirp test signal was extracted from a wide-band noise by applying an iterative algorithm, then comparing the result with the noisy waveform. In order to decrease computational complexity, Tao et al. [11] suggested a block time-recursive algorithm to find the discrete Gabor coefficients then, extract the original signal for both the critical sampling and the over-sampling cases.

This paper is organized as follows. In Section 2 a brief overview of the discrete Gabor transform (DGT) is provided. The inverse discrete Gabor transform (IDGT) and its applications to signal reconstruction is presented in Section 3. In the next Section 4, experimental results and findings including implementation procedure for this experiment are outlined. In the final part of the paper, Section 5, analysis of results and conclusions are presented.

THE DISCRETE GABOR TRANSFORM (DGT)

In this section, a review of the discrete Gabor expansion is presented, before implementation procedure is conducted in the later part of the experiment.

The discrete Gabor expansion

For a discrete-time finite, real and periodic function $x(n)$ with a period L, the Gabor expansion can be written in the following form [7,12]

$$x(n)=\sum_{m=0}^{M-1}\sum_{k=0}^{N-1}a(m,k)h(n,m\overline{N})e^{\frac{j2\pi kn}{N}} \quad (1)$$

where M is the total number of time sampling points and N is the number of points sampled in frequency. The values of $\overline{N}$ represents the time sampling interval while $\overline{M}$ is the frequency-sampling interval.

On the other hand, the coefficients $a(m,k)$ are calculated according to the following formula

$$a(m,k)=\sum_{n=0}^{L-1}x(n)\gamma(n,m\overline{N})e^{\frac{j2\pi kn}{N}} \quad (2)$$

Note that in the above equation the equality $L=\overline{N}M=N\overline{M}$ holds. The critical sampling point occurs when $\overline{N}\,\overline{M}=NM=L$ where the number of samples from the original signal is equal to the number of Gabor coefficients. Under sampling occurs when $NM>L$ which can leads to information loss. Therefore, for perfect reconstruction of the original signal, the following condition must be satisfied $NM\geq L$.

Moreover, both the analysis window $h(n)$ and the synthesis window $\gamma(n)$ must be real, periodic, satisfying the biorthogonality condition given by:

$$h(n)=\sum_{n=0}^{L-1}h(n+mN)e^{\frac{j2\pi kn}{\overline{N}}}\;\gamma(n)=(L/MN)\delta_m\delta_k \quad (3)$$

where $0\leq m\geq\overline{M}$ -1, $0\leq k\geq\overline{N}$ -1 and $\delta(n)$ denotes the Dirac delta function. The above expression in Eq. (1) is defined as the inverse discrete Gabor transform; while Eq. (2) is known as the DGT.

THE INVERSE DISCRETE GABOR TRANSFORM (IDGT)

The de-convolution of the discrete Gabor transform refers to recovering the original signal from the time-frequency distribution. In classical Fourier analysis if the signal is narrow-band and stationary, then linear filtering is implemented to recover the original signal. In this case, a simple procedure made up of a band-pass filter covering the band of the signal is used to recover the desired output from the wide-band-noise. In the frequency domain, the band-pass filter transfer function is multiplied by the Fourier transform of the noise as a form of a mask for the spectrum. Then, the inverse Fourier transform of the result is calculated to recover the noise-free signal.

However, when dealing with non-stationary signals such as speech, biomedical or seismic data buried in noise, traditional linear filtering cannot be used. In this case, a different approach known as time-frequency filtering or masking is utilized to extract the original signal. In the discrete Gabor method, reconstruction of the signal $x(n)$ can be obtained from Eq. (1) after satisfying the conditions mentioned in the above section. That is, the inverse FFT of the coefficients $a(m,k)$ is computed for each index k, then taking the point-by-point product with the synthesis window $h(n)$ to obtain a specified portion of the output. To recover the whole signal, the windowed slices are summed over the index m.

Consequently, the above procedure can be used for time-frequency filtering by modifying the Gabor coefficients used to calculate the time-frequency distribution prior to signal reconstruction. This is known as time-varying filtering where two or more non-stationary signals can be separated and the desired one is synthesized. A number of applications take advantage of this particular technique for the objective of signal recovery. Some of these examples include a non-stationary seismic data de-convolution designed to remove unwanted earth attenuation effects and source signature [10]. Another example is the implementation of a recursive Gabor filter for image processing with the least possible number of operations [13]. An important application is speech enhancement, where this approach is exploited in the time-frequency representation to remove inherent noise and recreate the uncorrupted time wave.

EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the IDGT is implemented as a time-varying filter for noise removal and enhancement of speech. First a synthetic non-stationary signal made up of tow chirps that cannot be separated using traditional linear filtering in the Fourier (frequency) or time domain is used to simulate a noise-corrupted waveform. One chirp is treated as noise while the second simulates the desired time-series wave. This is illustrated in Figure 1 showing the time-frequency representation (TFR) of the test chirp. On the other hand, Figure 2 shows the two chirps' distribution with one of the signals depicting noise presence. The original signal is then recovered or separated from the time-frequency distribution by utilizing a masking procedure to remove the present noise. The result is given in Figure 3 with the noise-free signal recovered using the proposed masking technique. The time series plot of error between the original and recovered signals is also shown in Figure 4 providing a fairly accurate result.

In order to create a suitable mask, the time frequency distribution (TFD) of the distorting noise is set to zero while at the same time the desired part is multiplied by unity. This will keep the desired components inside the mask and reject other parts of the signal. The calculation of an appropriate time-frequency mask can be done according to the following procedure:

$$M(n,m)=\begin{cases}1, & (n,m)\in C\\ 0, & otherwise\end{cases} \quad (4)$$

The value of C represents the pass region of the time-frequency mask. Hence, the final result of the masked time-frequency distribution is in the form:

$$\overline{S}(n,m)=\widehat{S}(n,m)\,M(n,m) \quad (5)$$

where $\overline{S}(n,m)$ stands for the result of masking, $\widehat{S}(n,m)$ is the corrupted value, and $M(n,m)$ is the mask function. The signal can then be reconstructed using the inverse GDT from the masked output. In order to obtain an appropriate mask, a threshold can be calculated as the average value of the peaks in the noise- corrupted value. It is important to note that time frequency distributions are not one to one (redundant) or onto transformations. As a result, not every signal in the joint time-frequency representation corresponds to a signal in the time domain.

Furthermore, as an application to the above time-frequency masking procedure, speech utterances corrupted with different kinds of noise are synthesized for the objective of interference cancellation. This type of noise is very similar to the clean signal, which makes the task of removal very difficult or impossible using traditional Fourier techniques. The first example is the word "long" spoken by a female and recorded inside a talking crowed, which of course will affect the fidelity of speech. The signal length is 27 ms, with a sampling rate of 11 kHz. The original noise-free time-frequency distribution is illustrated in Figure 5. On the other hand, Figure 6 depicts the noisy speech, while Figure 7 is the masked result with most of the present noise removed. The error between the extracted waveform and the original is shown in Figure 8. As can be seen from this output, the recovered signal is a good approximation of the original speech with a very small difference.

As a second example, the word "down" spoken by a male of length 355 ms has a sampling rate of 8 kHz with added noise from an airport environment. The time-frequency distribution is depicted in Figure 9 for the clear signal; while Figure 10 clearly shows noise presence in the time-frequency plane. The following Figure 11 depicts the masking result of the time-frequency distribution after interference cancelation. The extracted waveform represents a very good approximation of the original as Figure 12 illustrates.

The above technique is also implemented for two other English speech examples. The next one is the word "great" spoken by a female and corrupted by a moving train sound. This has a length of 215 ms and a sampling rate of 11 kHz. The outputs are displayed in Figures 13, 14, 15, and 16 showing the error between the recovered speech and the original signal. The fourth and final example is the word "pleasant" spoken by a male inside a noisy restaurant atmosphere. In this case, the signal has a length of 420 ms with a sampling rate of 8 kHz. The answers are displayed in Figures 17, 18, 19, and 20. The extracted wave in this example also provides a promising outcome when compared with the clean signal. The overall results are also displayed in Table 1 representing signal to noise ratio (SNR) of individual speech signals before and after the proposed TFR filtering procedure. A significant improvement in the speech quality and noise reduction is achieved.

Table 1. Signal to noise ratio (SNR) in dB of the four speech utterances in different noisy environments

Test Word	SNR Before Filtering	SNR After Filtering
Long (crowed noise)	1.4804 dB	6.3620 dB
Down (airport noise)	-1.6488 dB	6.2684 dB
Great (train noise)	5.5486 dB	9.2970 dB
Pleasant (restaurant noise)	1.1858 dB	8.7993 dB

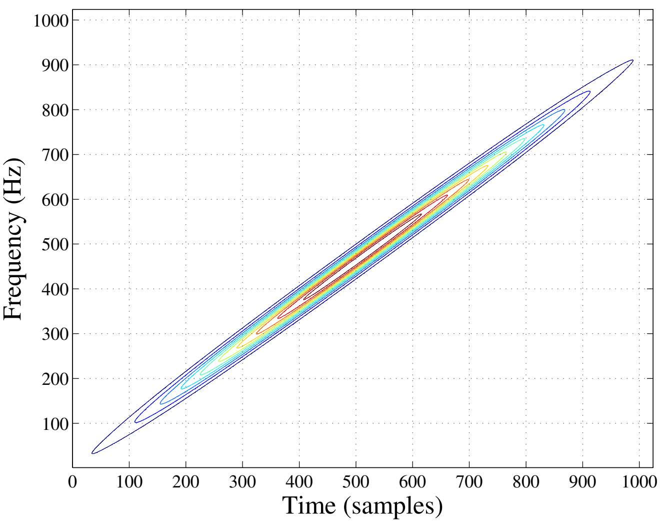

Figure 1. Time-frequency representation of the noise-free test chirp signal

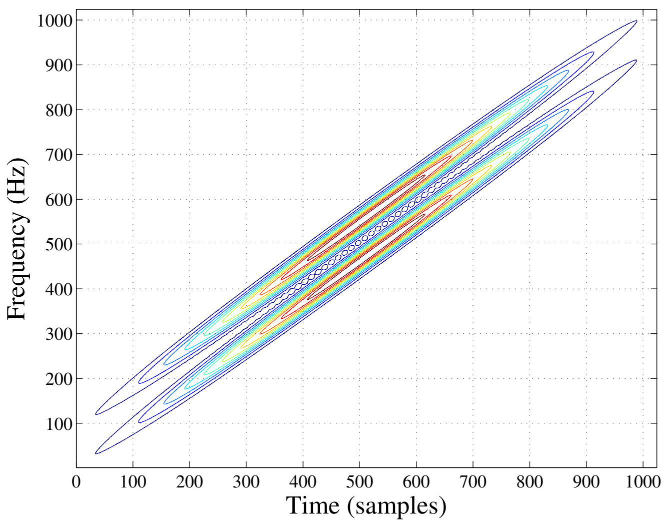

Figure 2. Time-frequency representation of the test signal with noise present

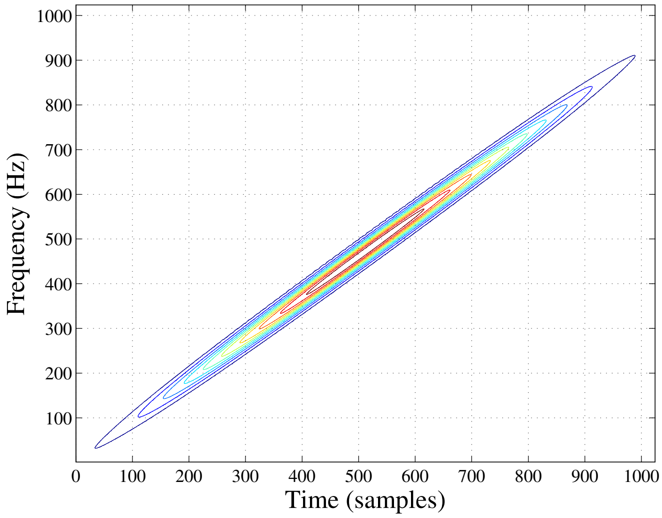

Figure 3. Time-frequency representation of the recovered test signal with noise removed

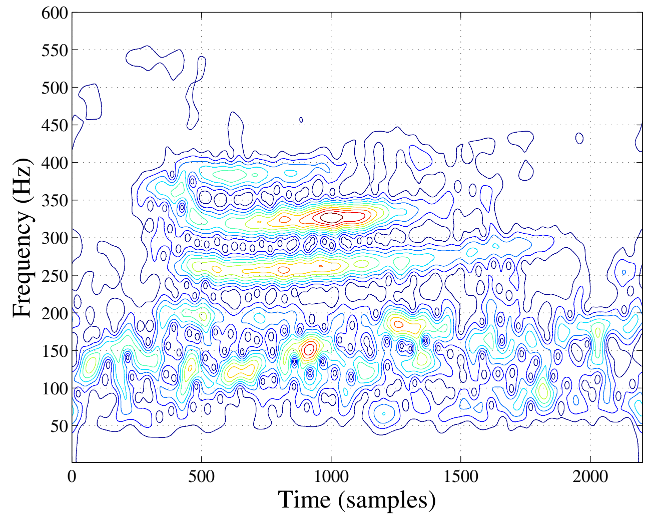

Figure 6. Time-frequency representation of the noisy speech signal "long"

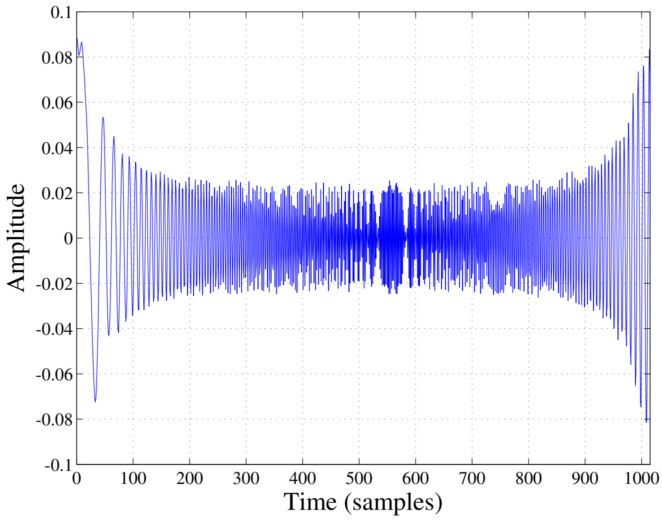

Figure 4. Plot of error between the original and the recovered signal

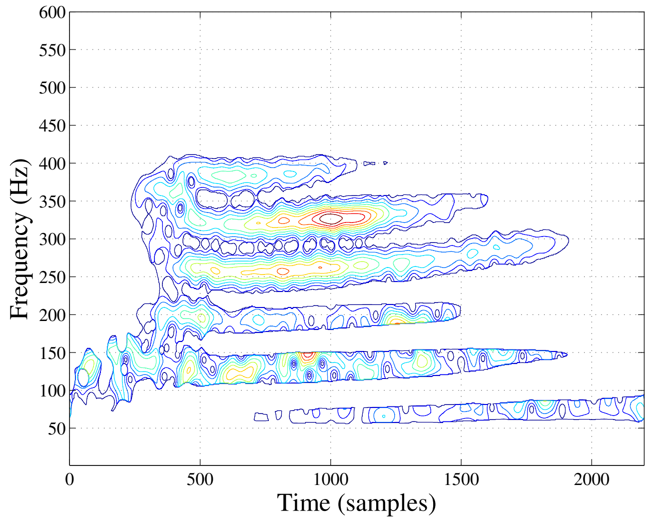

Figure 7. Time-frequency representation of the recovered speech signal "long" with most of the noise removed

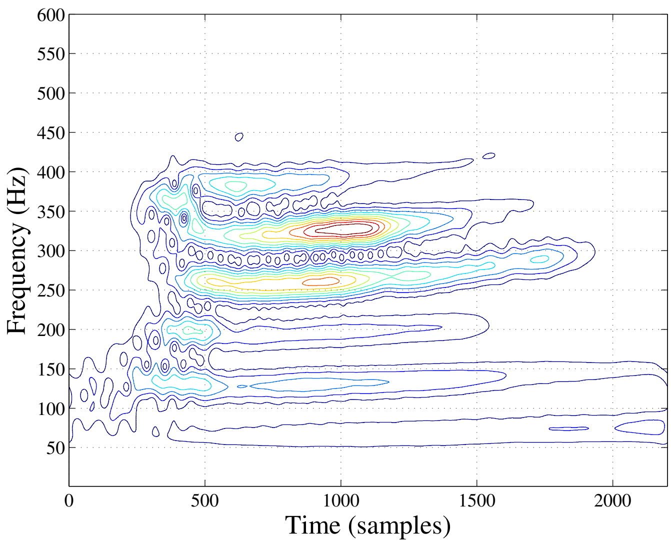

Figure 5. Time-frequency representation of the noise-free speech signal "long"

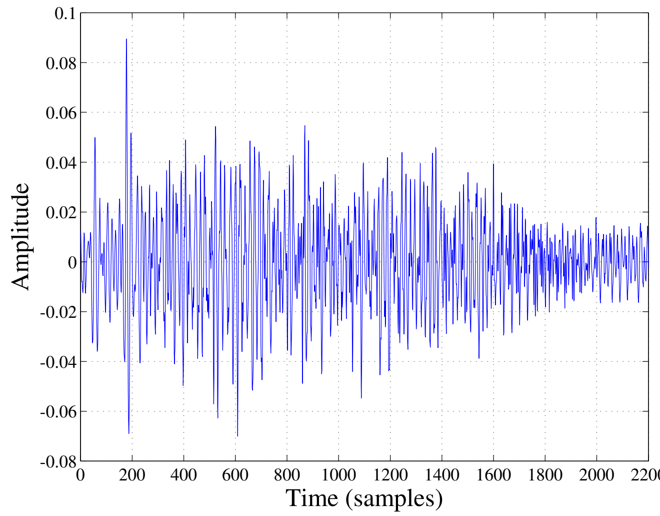

Figure 8. Plot of error between the original speech signal "long" and the recovered signal

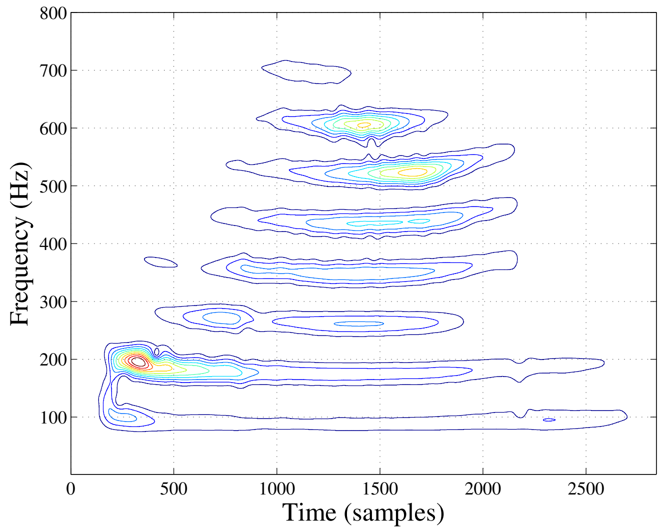

Figure 9. Time-frequency representation of the noise-free speech signal "down"

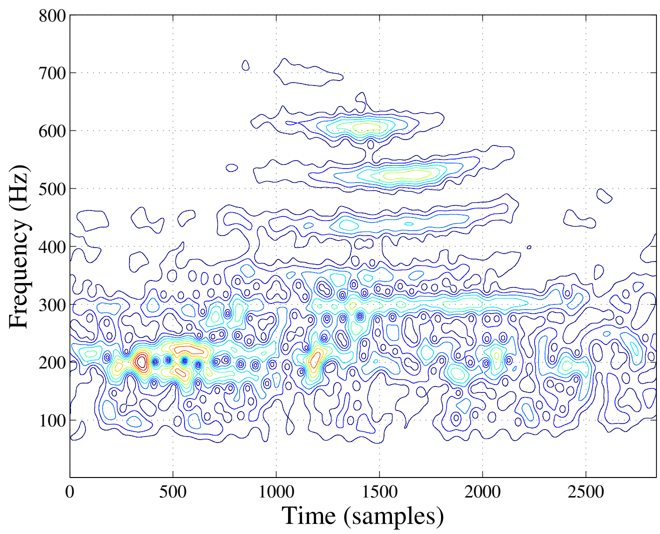

Figure 10. Time-frequency representation of the noisy speech signal "down"

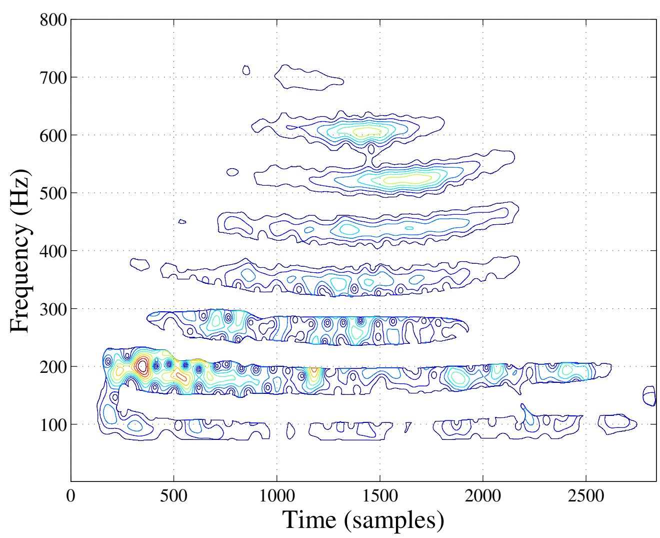

Figure 11. Time-frequency representation of the recovered speech signal "down" with most of the noise removed

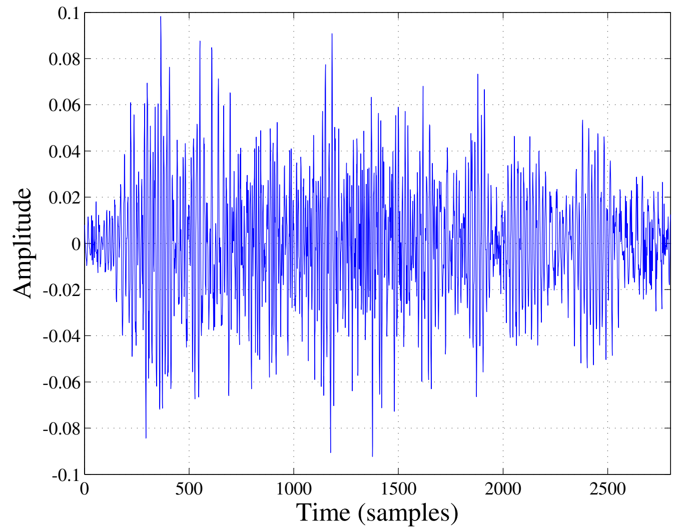

Figure 12. Plot of error between the original speech signal "down" and the recovered signal

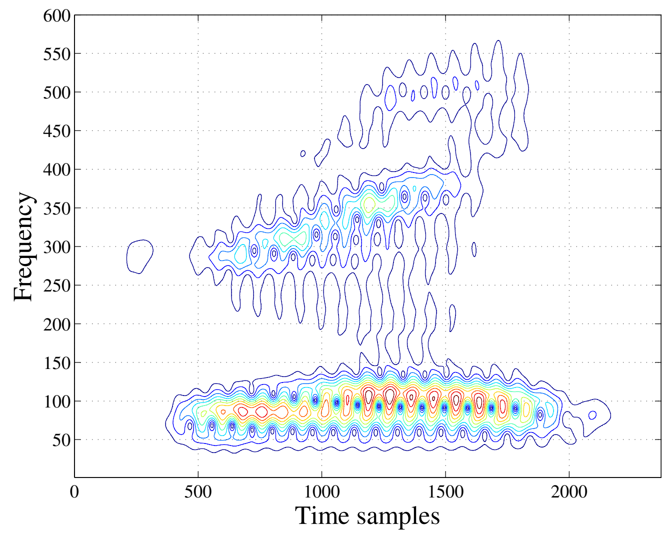

Figure 13. Time-frequency representation of the noise-free speech signal "great"

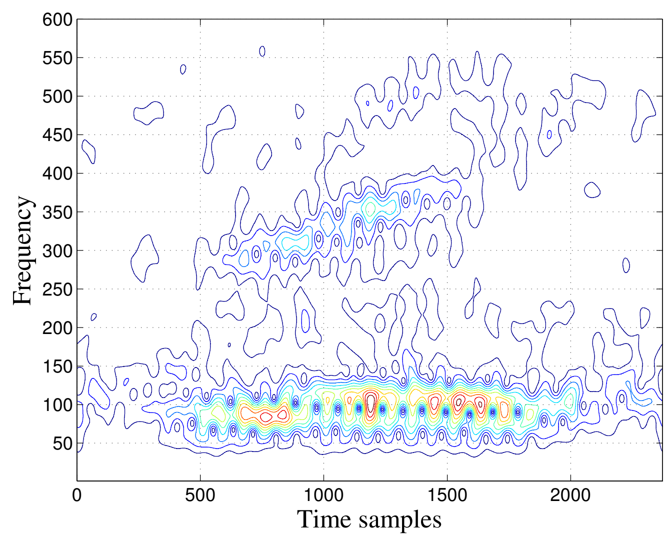

Figure 14. Time-frequency representation of the noisy speech signal "great"

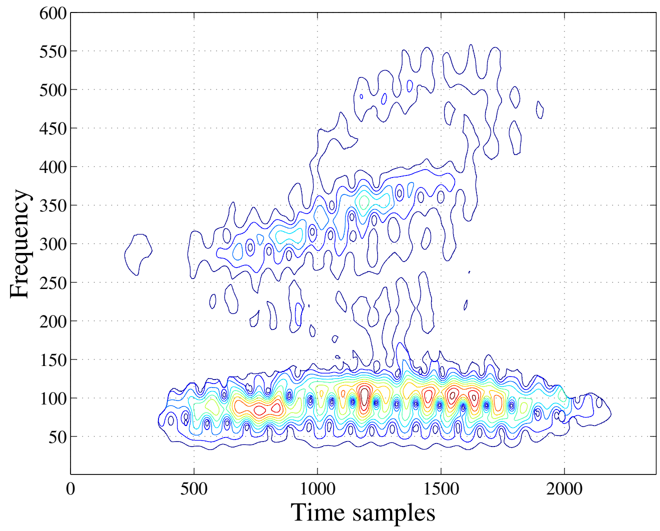

Figure 15. Time-frequency representation of the recovered speech signal “great” with most of the noise removed

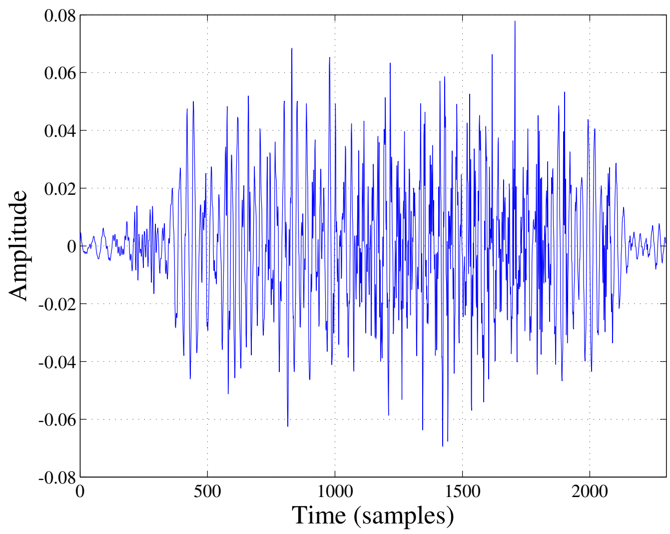

Figure 16. Plot of error between the original speech signal “great” and the recovered signal

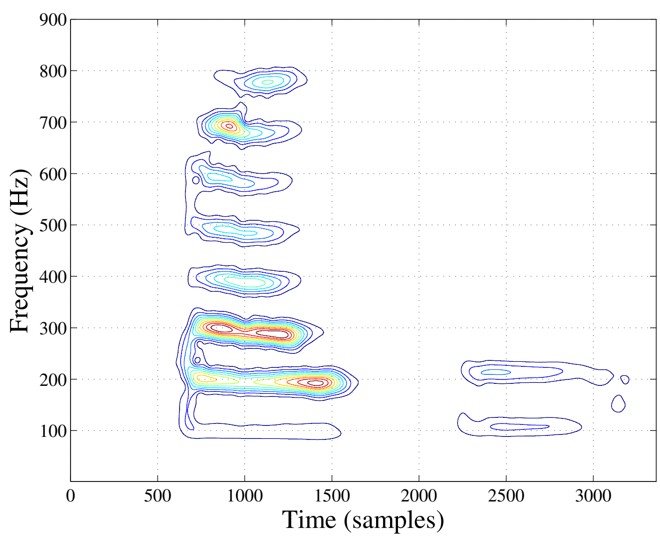

Figure 17. Time-frequency representation of the noise-free speech signal “pleasant”

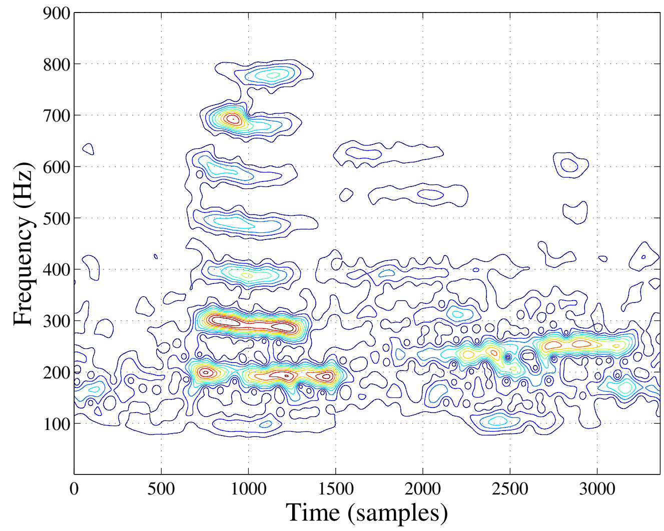

Figure 18. Time-frequency representation of the noisy speech signal “pleasant”

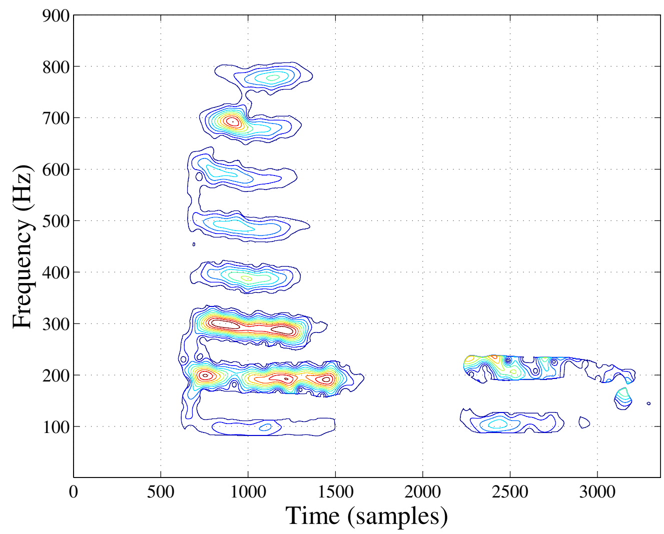

Figure 19. Time-frequency representation of the recovered speech signal “pleasant” with most of the noise removed

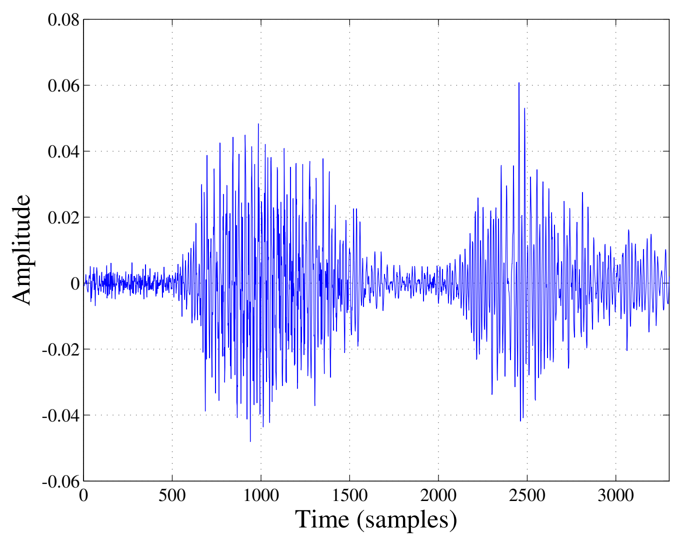

Figure 20. Plot of error between the original speech signal “pleasant” and the recovered signal

CONCLUSION

The Inverse discrete Gabor Transform was implemented to recover corrupted speech signals using a time-frequency masking approach. The main contribution of this work is that the speech signal is buried in a wide-band non-stationary noise very similar in characteristics to the original wave form, which makes it impossible to recover using classical Fourier transform techniques. First, a synthetic non-stationary multicomponent chirp was tested then; four different examples of English words recorded in noisy environments were used to evaluate the effectiveness of the proposed procedure. The original signals were then reconstructed from the time-frequency distribution of the noise-infected speech after processing. The recovered waveforms were compared to the originals in terms of fidelity or noise presence and were found to be a very good approximation of the recorded clean ones. Since this type of noise cannot be separated from the desired signal in time or frequency; this technique provides a practical alternative to the traditional methods that are not capable of resolving this issue.

REFERENCES

[1] D. Gabor, "Theory of communications", *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication* **93**, 429-459 (1946)

[2] M.P. Lamoureux, P.C. Gibson, J.P. Grossman, and G.F. Margrave, "A fast, discrete Gabor transform via a partition of unity", *CREWES Technical Report Volume 15*, University of Calgary – Consortium for Research in Elastic Wave Exploration Seismology, 2003

[3] T.T. Chinen and T.R. Reed, "A performance analysis of fast Gabor transform methods", *Graphical Models and Image Processing* **59**, 117-127 (1997)

[4] A. Akan and L.F. Chaparro, "Multi-window Gabor expansion for evolutionary spectral analysis", *Signals Processing* **63**, 249-262 (1997)

[5] M.A. Al-Manie and W.J. Wang, "Time-frequency analysis by evolutionary periodogram with application in gear fault diagnosis", *International Journal of Wavelets, Multiresolution and Information Processing* **8**, 679 -693 (2010)

[6] W.J. Wang and P.D. McFadden, "Application of wavelets to gearbox vibration signals for fault detection", *Journal of Sound and Vibration* **192**, 927-939 (1996)

[7] S. Qian and D. Chen, "Discrete Gabor transforms", *IEEE Transactions on Signal Processing* **41**, 2429-2438 (1993)

[8] J. Yao, P. Krolak and C. Steele, "The generalized Gabor transform", *IEEE Transactions on Image Processing* **4**, 978-988 (1995)

[9] G.F. Margrave, P.C. Gibson, J.P. Grossman, D.C. Henley and M.P. Lamoureux, "Gabor deconvolution: Theory and practice", *Proceedings of the EAGE 66th Conference and Exhibition*, Paris, France, 7-10 June 2004

[10] X.-G. Xia and S. Qian, "An iterative algorithm for time-variant filtering in the discrete Gabor transform domain", *Proceedings of the 1997 IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP-97)* **3**, pp. 2121-2124, 21-24 April 1997

[11] L. Tao and H.K. Kwan, "Parallel lattice structures of block time-recursive discrete Gabor transform and its inverse transform", *Signal Processing* **88**, 407-414 (2008)

[12] M.J. Bastiaans and C.W. Geiln, "On the discrete Gabor transform and the discrete Zak transform", *Signals Processing* **49**, 151-166 (1996)

[13] I.T. Young, L.J. van Vliet and M. van Ginkel, "Recursive Gabor filtering." *IEEE Transactions on Signal Processing* **50**, 2798-2805 (2002)