The New Spectrogram Evaluated by Enhanced Continuous Wavelet and Short Time Fourier Transforms via Windowing Spectrums

Yih-Nen Jeng (鄭育能)
Professor
Dept of Aero. & Astro. National Cheng-Kung University
Email: Z6208016@email.ncku.edu.tw
and
You-Chi Cheng (鄭又齊)
Tainan, Taiwan, 70101

ABSTRACT

The continuous wavelet transform and short time Fourier transform, both using the Gaussian function as kernels, are enhanced by introducing a Gaussian window to the spectrum of the discrete Fourier spectrum with negligible low frequency error. By truncating two ends to obtain zero value via interpolation, the original data is redistributed via the monotonic cubic interpolation. Two necessary restrictions are: the number of points is of 2^m and more than one point should be placed in every data interval of the original data string. The spectrum attains a negligible low frequency error after using a odd mapping to extending the data string. The Gaussian function factor of the Gabor transform is also related to the wavelength to improve the resolution in time domain. The enhanced Morlet wavelet and the enhanced Gabor transforms with fixed and variable σ 's are applied to evaluate the spectrogram of a voice of the word "hello" . The visibility of all the resulting spectrograms is better than those employing the original Morlet and Gabor transforms. It seems that a coarse frequency resolution or a wide window in frequency direction will modified the true feature of a spectrogram. Moreover, there are many detailed information involved in the resulting the spectrograms which may be related to the emotion and character of speaker.

Keywords: Spectrogram, Gabor transform, Morlet transform, fine frequency resolution

1. INTRODUCTION

The spectrogram is an time-frequency representation which allow a precise description of non-stationary speech signal [1-6]. Two-dimensional spectrogram images are computed by concatenating spectra obtained by short time Fourier transforms, which is also named as the Gabor transform [2]. This transformation assumes that the signal is quasi-stationary for the length of the window. As noted in Ref.[3], time and frequency resolutions are inversely proportional because good resolution in both the time and frequency domains in the same image is not possible. Consequently, a small analysis window leads to poor localization of the frequency components and vice versa. In practice, narrowband spectrograms (with long window lengths) are used for good frequency resolution while wideband spectrograms (with short window lengths) allow good temporal resolution of speech signals. To improve this deficiency, people either employed a better localization or used method of reassignment and had successfully enhanced the accuracy of creating speech features [4-6].

To the authors' knowledge, the resolution of a spectrogram is much coarser than that of a finger-print. In fact, human's speech signals are generated by organic voice system and involve many characteristics of the speaker. Consequently, from a speech signal string, one can understand not only the information of the speech itself but also the emotion, physical function and health state of the speaker. Unfortunately, authors can not find a report about how to extract a speaker's health conditions and emotions from a spectrograms generated by a currently short time Fourier transform.

In general, a data string may or may not involve smooth non-periodic and rapid varied parts. Typical examples of rapidly varied parts are discontinuous jump and spikes of brain neural signals. The corresponding spectrum components of these non-sinusoidal parts also run over the whole spectrum domain. To generate a Fourier spectrum of a finite data string, most people just access an available Fast Fourier Transform (FFT) algorithm. The application of an FFT algorithm to an untreated data string may be equal to enforce the periodic condition at two ends of the data string which leads to certain spectrum error. Therefore, a spectrum generated by an FFT algorithm may contain a significantly large fraction of information of the non-periodic error and non-sinusoidal parts. Although a speech signal might not involve significant non-sinusoidal information, the non-periodic error is still a trouble source. In order to remove the deficiency caused by this error, people successfully employed different windows to extract local information without removing the error [1]. However, a windowing spectrum embedded with unknown degree of error cannot be employed as a precise parameter representation of the original data such as to enhance the visibility of

spectrograms. In a recent study, the authors had proposed a simple strategy to obtain an accurate spectrum by eliminating the non-periodic condition and removing the non-sinusoidal part before applying an FFT algorithm [7]. This algorithm has a penalty that the effective data length is shrunk because the data string is truncated to zero crossing location at two ends.

After Morlet proposing a continuous wavelet transform[8], Farge et. al. had successfully employed it to study the turbulent flow data string [9,10]. In fact, a continuous wavelet coefficient plot is in some sense similar to the spectrogram generated by the short time Fourier transform[1,2]. In Ref.[11], the Morlet wavelet transform had been improved by embedding a Gaussian window to the spectrum so that the resulting two-dimensional wavelet coefficient clearly showing many features not seen before. In this study, the coefficient plots of both the enhanced Morlet and Gabor transforms will be examined both embedded with a Gaussian window to the spectrum generated by the algorithm of Ref.[7].

THEORETICAL ANALYSIS

Assuming that a data string is only composed of general sinusoidal waves whose amplitude and frequency may or may not be functions of time. The simple strategy of Ref.[7] is modified to be the followings.

1. Choose zero crossing points at two ends and use an interpolation method to find 0 points there.
2. Use the monotonic cubic interpolation of Ref.[12] to regenerate the data so that total number of points are of 2^m. Note that more than one point should be located in the range between two successive data points of the original data string to reduce interpolation error.
3. Use an odd function mapping to extend the data string so that the number of points becomes 2^{m+1}.
4. A simple and fast Fourier sine transform algorithm is employed to generate the desired spectrum.

Because the zero values at two ends are used and the odd function mapping are employed, no any error due to non-periodic condition is introduced except the interpolation error. However, since the zero values are chosen at two ends, the penalty of shrinking the available data range can not be avoided.

Assume that a discrete data string can be approximated by

$$y(t) = \sum_{n=0}^{N} b_n \cos\left(\frac{2\pi t}{\lambda_n}\right) + c_n \sin\left(\frac{2\pi t}{\lambda_n}\right) \tag{1}$$

The following Morlet transform transfer the data string $y(t)$ into the wavelet coefficient.

$$W(a,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} y(x)\psi^*\left(\frac{x-\tau}{a}\right)dx \tag{2}$$

where $\psi(x) = e^{i6x}e^{-|x|^2/2}$ and a is called as the scale function. If this transform is applied over a range $a_0 \le a \le a_1$, a two-dimensional wavelet coefficient plot is obtained on the (a,τ) plane. After some manipulations, it can be easily shown that the resulting wavelet coefficient is

$$W(a,\tau) \approx \sum_{n=0}^{N} (b_n - ic_n)\sqrt{\frac{\pi a}{2}} \exp\left(-\frac{a^2}{2}\left(\frac{2\pi}{\lambda_n} - \frac{6}{a}\right)^2\right) \times \exp\left[\frac{i2\pi\tau}{\lambda_n}\right] \tag{3}$$

A careful inspection upon this formula reveals that, if the original data is of the form

$$\begin{aligned} y(t) &= \sin(2\pi t/\lambda_m), && t_1 < t < t_2 \\ &= 0 && \text{otherwise} \end{aligned} \tag{4}$$

where $\lambda_m = a\pi/3$, the response of applying Eq.(2) will give a non-zero value in the region of $t_1 - 6.5a < t < t_1 + 6.5a$ and $0.65\lambda_m < \lambda < 2.4\lambda_m$. Since there is not any adjustable parameter in Eq.(2), one cannot change the factor 6.5 in time domain except if another windowing function is chosen.

The following short time Fourier (or Gabor) transform maps the data $y(t)$ into the continuous Gabor coefficient.

$$\begin{aligned} G(f,\tau,a) &= \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} y(t)\psi^*(f, t-\tau)dt \\ \psi(f, t-\tau) &= \exp[i2\pi f(t-\tau)]\exp\left[-\frac{(t-\tau)^2}{2a^2}\right] \end{aligned} \tag{5}$$

The resulting response for a given set (a, f) is

$$G(f,\tau,a) \approx \sqrt{\frac{\pi a}{2}} \sum_{n=0}^{\infty} \{(b_n - ic_n)\exp(i2\pi f_n \tau) \cdot \exp[-4\pi^2 a^2 (f_n - f)^2] \tag{6}$$

For the test function of Eq.(4), at $f = 1/\lambda_m$, the non-zero response range is somewhat similar to that of Morlet transform mentioned above.

In order to shrink the response range in frequency domain, both the Morlet and Gabor transform are embedded with a Gaussian window on the spectrum domain. The resulting enhanced Morlet transform is

$$\overline{W}(a,\tau) = \sum_{n=0}^{\infty} (b_n - ic_n)\sqrt{\frac{\pi a}{2}}\left[\exp\left(-\frac{a^2}{2}\left(\frac{2\pi}{\lambda_n} - \frac{6}{a}\right)^2 - \frac{[n - 3T/(a\pi)]^2}{2\sigma^2}\right)\right] \times \exp\left[i\frac{2\pi\tau}{\lambda_n}\right] \tag{7}$$

where $3T/a\pi$ is the mode number associated with the Gaussian window parameter a and n is the mode number. The enhanced Gabor transform is

$$\overline{W}(f,\tau,a) \approx \sqrt{\frac{\pi a}{2}} \sum_{n=0}^{\infty} (b_n - ic_n)\exp(i2\pi f_n \tau)\cdot \exp\left[\frac{-4\pi^2 a^2 (f_n - f)^2}{2}\right]\exp[-\frac{(n-m)^2}{2\sigma^2}] \quad (8)$$

where $f = f_m$, and m is the mode number. The wavelet coefficient of Eq.(7) contiains a user specified parameter σ while the Gabor coefficient of Eq.(8) has two user specified parameters σ and a. The parameter a takes a constant value for all σ and τ. However, this specification froze the window size in time domain. In fact, it can be related to the specific wavelength by defining $a = k\lambda = k/f$, where k is a parameter to adjust the window size on time domain. Consequently, the following new Gabor transform is obtained, say $\overline{W}(f,\tau,k)$

$$\approx \sqrt{\frac{\pi k}{2f}} \sum_{n=0}^{\infty} (b_n - ic_n)\exp(i2\pi f_n \tau)\exp\left[\frac{-2\pi^2 k^2 (f_n - f)^2}{f^2}\right] \cdot \exp[-\frac{(n-m)^2}{2\sigma^2}]$$

$$\approx \sqrt{\frac{\pi k\lambda}{2}} \sum_{n=0}^{\infty} (b_n - ic_n)\exp(i2\pi f_n \tau)\exp\left[-2\pi^2 k^2 \lambda^2 (f_n - f)^2\right] \cdot \exp[-\frac{(n-m)^2}{2\sigma^2}] \quad (9)$$

If one exclude the trem $\sqrt{\pi a/2}\exp[-a^2(2\pi/\lambda_n - 6/a)^2/2 - [n-3T/(a\pi)]^2/(2\sigma^2)$ from Eq.(7), $\sqrt{\pi a/2}\exp[-2\pi^2 \times a^2(f_n - f)^2]\exp[-(n-m)^2/(2\sigma^2)]$ from Eq.(8), and $\sqrt{\pi k\lambda/2}\exp\left[-2\pi^2 k^2 \lambda^2 (f_n - f)^2\right]\exp[-(n-m)^2/(2\sigma^2)]$ from Eq.(9), respectively, the remaining terms are only but some rearranged form of Eq.(1). Consequently, by summing up a or f for a given τ, one can easily gain an inverse transform by taking the whole two-dimensional coefficient domain or excluding specific regions, provided that the calculated a or f can approximately resolve the whole spectrum. For a discrete data string, the upper limit of summations of Eqs.(7-9) becomes a finite value N and similar inverse transforms can be easily obtained too.

RESULTS AND DISCUSSIONS

Now a string of the voice "hello" (shown in Fig.1) is employed to examine the present modifications. The corresponding spectrum is shown in Fig.2. By using 120 uniformly spacing lines to resolve the frequency from 50 to 1350Hz ($\Delta f \approx 11$ Hz), the resulting spectrogram generated by the classical Gabor transform are shown in Fig.3, where Fig.3a is the narrowband spectrogram (with long window size $a = 0.02$ second) and Fig.3b is the wideband spectrogram (with short window size $a = 0.002$ second). These results agree with the well known knowledge that the narrowband spectrogram gives detailed resolution in frequency-direction while the wideband spectrogram shows detailed resolution in time. To obtain a complete information, these two spectrograms should be combined via some procedure. That shown in Fig.4 is the resulting spectrogram of the enhanced Gabor transform employing the variable window size $a(=k\lambda)$ and windowed spectrum with parameter $k = 1$ in Eq.(9). This result is similar to results of the enhanced Gabor transform with fixed windowing size ($a = 0.002$ second) and $k = 1$ and the enhanced Morlet transform with the spectrum windowing factor $\sigma = 1$ and $k = 1$. These results are now shown here because of length limitation of the paper.

A careful comparison between Figs.3 and 4 reveals that the three proposed methods: the enhanced Morlet transform and enhanced Gabor transform with fixed and variable a are slightly better than the narrowband spectrogram generated by the original Gabor transform. In some sense, for a coarse frequency resolution, it means that all the present enhanced transformations generate a result which assemblage information of narrow and wideband spectrograms. However, the window size a of both narrow and wideband of the original Gabor need a try and error procedure for a given frequency resolution. On the other hand, all the three enhanced transform are relatively insensitive with respect to their corresponding parameters.

Now the calculation is restricted to be the range of 400 to 500Hz with 100 uniformly spacing lines so that a fine frequency resolution is achieved. The best resolved narrow and wideband spectrograms are shown in Fig.5. Results of employing the enhanced Morlet transform with $\sigma = 1$, the enhanced Gabor transform with fixed $a = 0.002$ second and $\sigma = 1$, and the enhanced Gabor transform with variable a ($k = 1$) and $\sigma = 1$ are shown in Fig.6 through Fig.8, respectively. In order to get a more complete impression about these results, the 3-D plot of the spectrograms are also included. It seems that the narrowband spectrogram generated by the original Gabor transform cannot suggest a fine enough resolution in the frequency-direction. Note that, whenever $a << 1$ second, the region with a non-zero value of $\exp\left[-4\pi^2 a^2 (f_n - f)^2\right]$ covers a wide range in the frequency-direction. That means too many information are folded into the coefficient $G(f,\tau,a)$ of a line of $f = c$ on the wideband spectrogram diagram. Consequently, the accuracy of this wideband data is not sufficiently high such that it is rather difficult to assemblage both the narrow and wideband spectrograms into a detailed spectrograms.

On the other hands, all the figures of Fig.6-8 show more detailed information than that of Figs.5. By examining the three-dimensional amplitude (energy) plots of Fig.5, it is seen that the original Gabor transform cannot resolve the horizontal strips very well

because it fold too many information from many wave components with different frequency at the same time. On the other hand, the horizontal strips structure of Figs.6-8 are obvious. Figure 9a shows the real part plot corresponding to Fig.8, while Fig.9b shows the detailed plot from $t = 0.2$ to 0.3 second. These real plots clearly show the phase information. Figure 9b clearly shows phase difference between wave components whose frequencies close to each other. Consequently, either for a coarse frequency resolution plot with $\sigma = 1$ or a fine frequency resolution plot with $\sigma \geq 2$, the wavelet coefficient or the Gabor coefficient will larger around the in phase region and smaller around the out-of phase region than those fine frequency resolution plot such as Fig.9a. Although human ear can not resolve such a fine frequency resolution feature, it relates to the detail states of speaker. In other words, if one tries to correlate the spectrogram with speaker's mood, organic structure, and health state etc., the fine frequency resolved spectrogram generated by the present methods are necessary.

A careful inspection upon Figs.6-8 reveals that, for the capability to resolve the detailed structure of the spectrogram, the best one is the enhanced Gabor transform with variable parameter a (Fig.8), the worst one is the enhanced Morlet transform. It should be noted that the enhanced Gabor transform with a fixed a has the problem to find the best parameter a. On the other hand, if σ =1 is selected, the enhanced Morlet transform and the enhanced Gabor transform with a variable a characterized by the parameter k =1 need not to worry about how to determine the parameter value. Therefore, it is recommend to employ either the enhance Morlet transform or the enhance Gabor transform with a variable parameter a related to the wavelength or frequency of the Gabor coefficient.

CONCLUSIONS

The present study purposes the enhance Morlet transform, the enhanced Gabor transforms with fixed and variable window size factors on time domain. Their capabilities to resolve a voice "hello" onto the spectrogram are all better than that generated by the original Gabor transform. These new techniques have the potential to examine the conditions of speaker.

REFERENCE

[1] T. F. Quatieri, Discrete-Time Speech Signal Processing Principles and Practice, 2002 Prentice Hall, PTR.

[2] R. Carmona, W. L. Hwang, and B. Torresani, Practical Time-Frequency Analysis, Chapter 3 and 4, Academic Press, 1998.

[3] M. J. Palakal, and M. J. Zoran, "Feature Extraction from Speech Spectrum Using Multi-Layered Network Models," IEEE international Workshop on Tools for Artificial Intelligence, Architectures, Languages and Algorithms, 1989, pp.224-230.

[4] F. Plante, G. Meyer, and W. A. Ainsworth, "Speech Signal Analysis with Reallocated Spectrogram," Proc. IEEE-SP int. Symposium On Time-Frequency and Time-Scale Analysis, 1994, pp.640-643.

[5] F. Plante, G. Meyer, and W. A. Ainsworth, "Imporvement of Speech Spectrogram Accuracy by the Method of Reassignment," IEEE Trans. Speech and Audio Processing, vol.6, no.3, pp. 282-286, May 1998.

[6] J. G. Vargas-Rubio and B. Santhanam, "An Improved Spectrogram Using the Multiangle Centered Discrete Fractional Fourier Transform," IEEE-ICASSP conference, pp.IV-505-5.8, 2005.

[7] Y. N. Jeng and Y. C. Cheng, "A simple Strategy to Evaluate the Frequency Spectrum of a Time Series Data with Non-Uniform Intervals," Trans. Aero. Astro. Soc. R. O. C., vol.36, no.3, pp.207-214, 2004.

[8] A. Grossmann and J. Morlet, "Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape," SIAM J. Math. Anal. Vol.15 no.4, July 1984.

[9] M. Farge, "Wavelet Transforms and Their Applications to Turbulence," Annu. Rev. Fluid Mech., vol.24, pp.395-457, 1992.

[10] M. Farge, N. Kevlahan, V. Perrier, and E. Goirand, "Wavelts and Turbulence," Proc. IEEE, vol.84, no.4, pp.639-669, April 1996.

[11] Y. N. Jeng, J. C. Chen, and Y. C. Cheng, "A New and Effective Tool to Look into Details of a Turbulent Data String," AIAA paper, Reno, Jan 2005.

[12] H. T. Huynh, "Accurate Monotone Cubic Interpolation," *SIAM J. Number. Anal.* vol.30, no.1, pp57-100, Feb.1993.

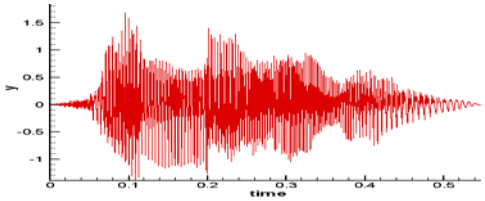


Fig.1 The raw data of the voice "hello".

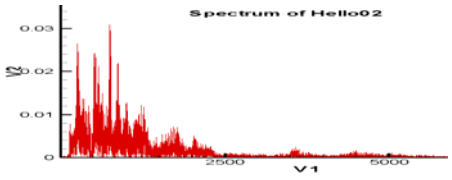
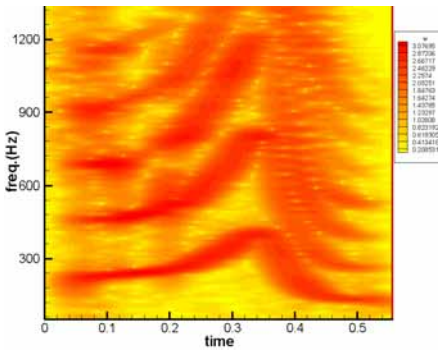


Fig.2 The spectrum of three voice "hello" of Fig.1, respectively with the horizontal axis be the mode number and the vertical axis be the amplitude of each mode.

(3a)



(3b)

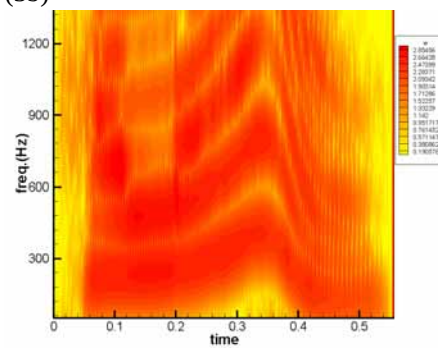


Fig.3 The spectrogram generated by the original Gabor transform: (3a) is the narrowband spectrogram with Gaussian window factor $a = 0.02$ second, and (3b) is the wideband spectrogram with $a = 0.002$ second.

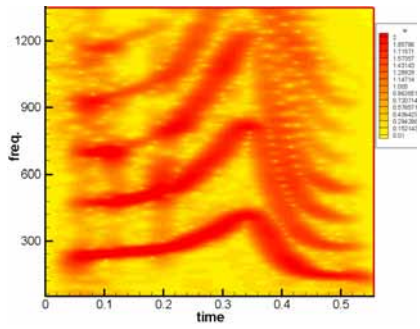
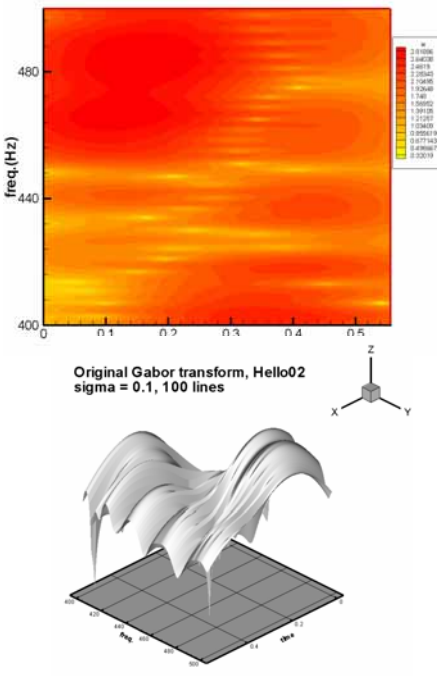


Fig.4 The spectrogram generated by the new Gabor transform with variable windowed size ($k = 1$) of spectrum and variable Gaussain window factor a .

(5a)



(5b)

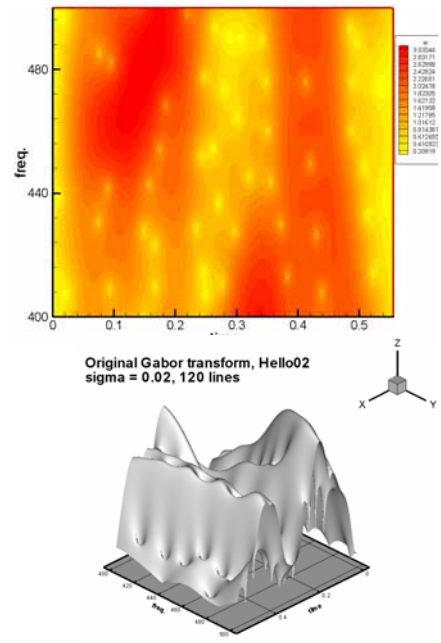


Fig.5 The spectrograms with fine frequency resolution generated by the original Gabor transform: (5a) is the narrowband spectrogram ($a = 0.1$ second) and (5b) is the wideband spectrogram ($a = 0.02$ second).

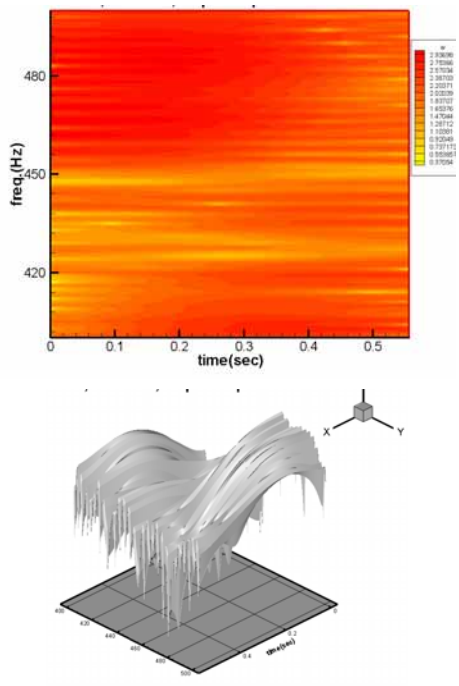


Fig.6 The spectrogram generated by the enhanced Morlet transform with $\sigma = 1$.

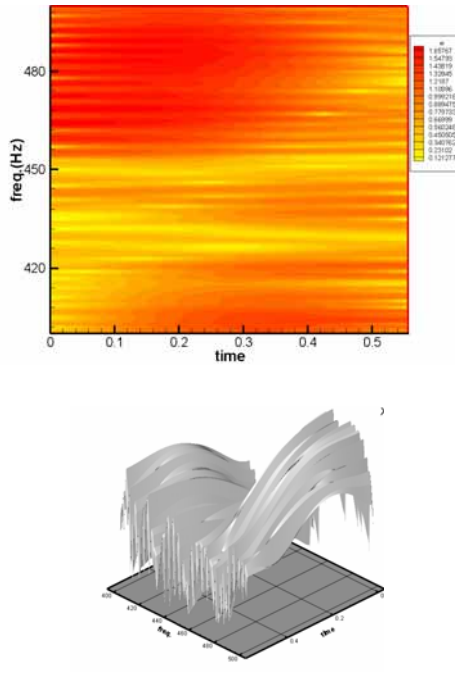


Fig.7 The spectrogram generated by the enhanced Gabor transform with fixed $a = 0.002$ second and $k = 1$.

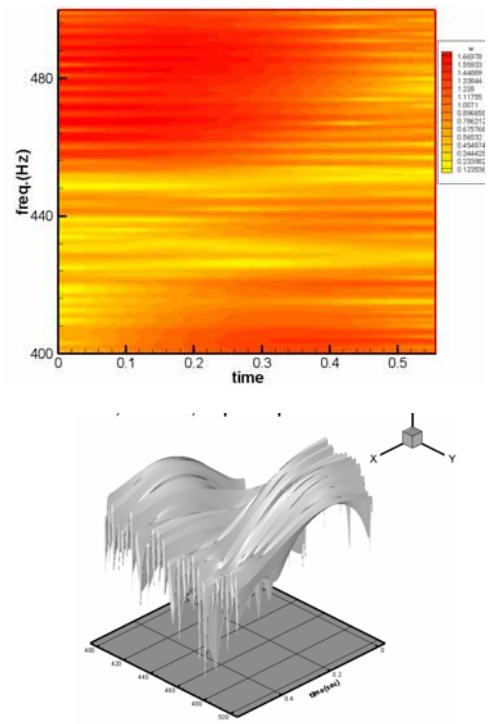
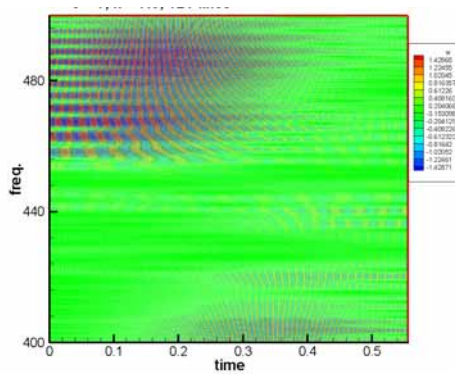


Fig.8 The spectrogram generated by the enhanced Gabor transform with variable a ($k = 1$) and $\sigma = 1$.

(9a)



(9b)

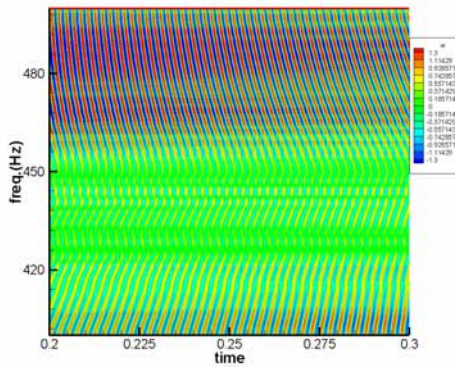


Fig.9 The real part plot of the spectrogram generated by the enhanced Gabor transform with variable a ($k = 1$) and $\sigma = 1$: (a) overall data plot, and (b) detailed data plot.