Noise Removing for Time-Variant Vocal Signal by Generalized Modulation

Jian-Jiun Ding and Ching-Hua Lee
Graduate Institute of Communication Engineering, National Taiwan University
E-mail: djj@cc.ee.ntu.edu.tw, melomeloguy@gmail.com, TEL: +886-2-33669652

***Abstract*— Since the instantaneous frequencies of vocal signals always vary with time, it is inconvenient to use the conventional filter to remove the noise of vocal signals. In this paper, we propose a method that uses the generalized modulation to reshape and minimize the areas of the spectrograms of vocal signals. Instead of multiplying an exponential function with the first order phase, the generalized modulation is to multiply an exponential function whose phase is a higher order polynomial. With the proposed noise-removing algorithm based on generalized modulation, the signal part and the noise part of a vocal signal can be well separated and the effect of noise can be significantly reduced.**

I. INTRODUCTION

Noise removal is always an essential issue for vocal signal processing. Vocal signals are often interfered by the noise caused from surrounding environment, which degrades their intelligibility. There are several well-known techniques that can be used for noise reduction, such as the Wiener filter, the infinite impulse response, the finite impulse response filter, and the spectral subtraction method [1-5]. However, none of these methods considers that the spectrum of a vocal signal may vary with time and utilizes the time and the frequency domain characteristics simultaneously to design the filter. In this paper, we propose a new noise-removing algorithm based on the generalized modulation and time-frequency analysis.

Since the spectrum of the noise always widely spread in the time-frequency domain, the area of a signal in the time-frequency domain determines how much noise is remained after applying the filter [6]. With the proposed filter based on the generalized modulation, the area of the time-frequency distribution of a signal can be minimized and the effect of noise can hence be much reduced.

Basically, a simple way to remove noise is by the use of frequency selective filters, i.e., one can use proper lowpass filters to separate the signal from the noise. With lowpass filters, the noise part whose frequency is beyond the highest frequency of the signal can be removed. Nevertheless, a great deal of noise might still remain after the lowpass filtration if the signal bandwidth is large. Therefore, to further reduce noise, it is reasonable to think of modifying the signal into one with a narrower bandwidth before going to the lowpass filtering step. This can be made by several methods; for example, by the combination of analytic signal generation and the conventional modulation [7]. After obtaining the signal with a narrower bandwidth, a lowpass filter with lower cutoff frequency then can be apply and hence the amount of noise would be further reduced.

Although using a certain method can indeed remove more noise, we believe that it can be further attenuated. In this paper, instead of the conventional modulation, we introduce a new noise-removing algorithm based on the proposed generalized modulation, which can reshape and significantly minimize the areas of the spectrograms of vocal signals and hence achieve even better performances on noise reduction. Moreover, the fractional Fourier transform (FRFT) [8-12] is also applied in our noise-removing scheme.

The whole noise-removing algorithm is plotted as in Fig. 1. In the modification stage of our algorithm, a series of methods, including the generalized modulation, are used to reshape the signal spectrogram. After that, the signal part is separated from the noise part in the next stage by a proper lowpass filter. The final step is the recovery stage, in which we reconstruct the signal by demodulation and the inverse FRFT.

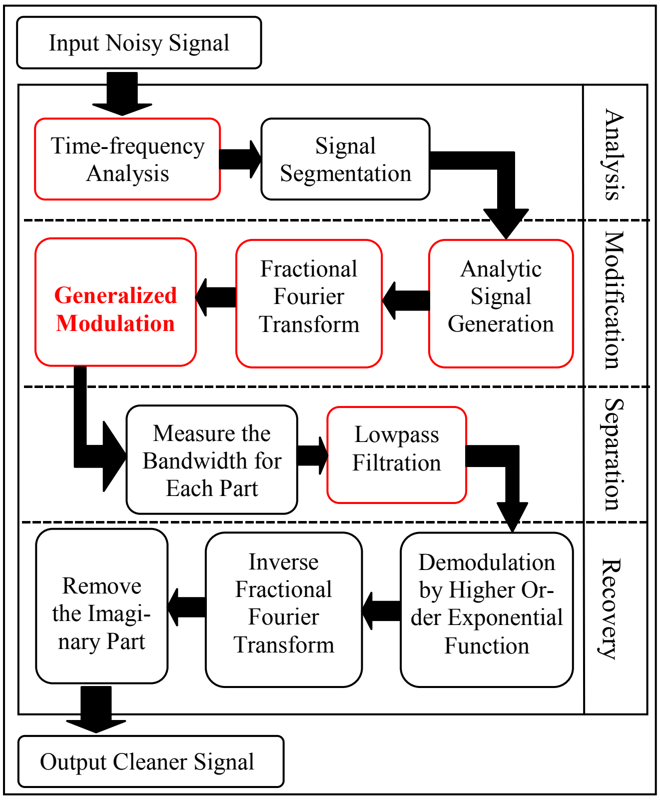

Fig. 1 Flowchart of the proposed noise-removing algorithm.

This paper is organized as follows. In Section II, the noise-removing approach based on the conventional modulation is briefly illustrated. In Section III, the proposed generalized modulation method is introduced. In Section IV, we combine the fractional Fourier transform (FRFT) with the proposed generalized modulation in the noise–removing algorithm. In Section V, the simulation results are presented. A conclusion is made in Section VI.

II. Noise Removing Based on Conventional Modulation

In this section, we discuss the noise-removing process based on the conventional modulation. The process consists of the time-frequency analysis, the analytic signal generation, and the conventional modulation.

Firstly, the time-frequency analysis is an approach for analyzing signals in a two-dimensional (2-D) way – simultaneously in the time domain and in the frequency domain [13]. Different from the conventional frequency analysis based on the Fourier transform, it utilizes the short-time Fourier transform (STFT) [14] for obtaining information in the 2-D time-frequency plain. For example, Fig. 2(a) is the STFT of the fundamental harmonic part of a human vocal signal. On the other hand, Fig. 2(b) shows the Fourier transform of the same signal. In Fig. 2(a), one can obtain the instantaneous frequency of the signal varying with time, rather than the only frequency information provided in Fig. 2(b).

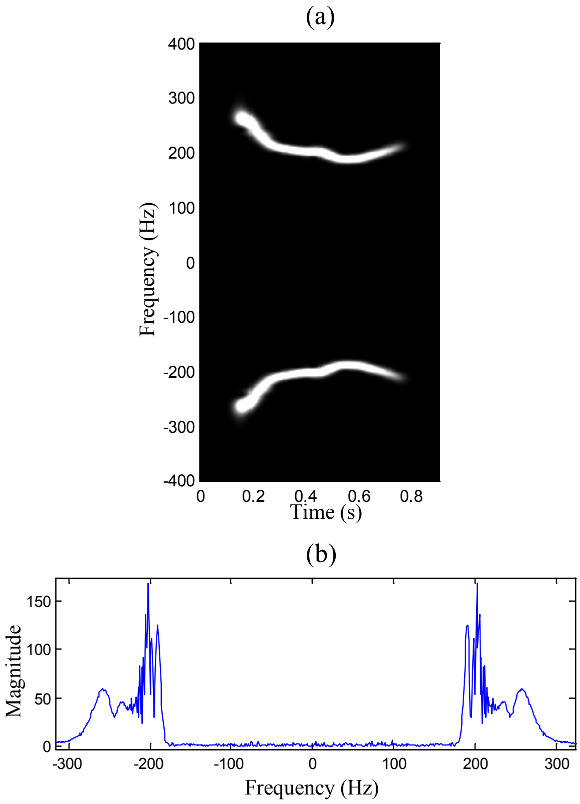

Fig. 2 (a) The STFT of the fundamental harmonic part of a human vocal signal. (b) The Fourier transform of the signal whose STFT is as in (a).

Next, the analytic signal generation can be achieved by using the Hilbert transform [7]. For example, the analytic signal $x_a(t)$ of a signal $x(t)$ is :

$$x_a(t) = x(t) + jx_H(t) \quad (1)$$

where $x_H(t)$ is the Hilbert transform of $x(t)$. Note that an important property of the analytic signal is :

$$X_a(f) = \begin{cases} X(f) & \text{for } f > 0 \\ 0 & \text{for } f \le 0 \end{cases} \quad (2)$$

where $X_a(f)$ and $X(f)$ are the Fourier transforms of $x_a(t)$ and $x(t)$, respectively.

For example, if $x(t)$ represents the human vocal signal as in Fig. 2(a), after applying the analytic signal conversion as in (1), the STFT of the resultant signal $x_a(t)$ is shown as in Fig. 3(a). From Fig. 3(a), one can see that only the positive part of the signal remains after analytic signal conversion, just as the aforementioned property indicates.

Finally, the conventional modulation is performed by multiplying an exponential function with the first order phase [7]. For example, a modulation function $m_c(t)$ is in the form of :

$$m_c(t) = \exp(-j2\pi f_1 t) \quad (3)$$

where f_l is the modulation frequency. After the conventional modulation operation, a signal will be frequency-shifted to a lower (if $f_l > 0$) or to a higher (if $f_l < 0$) location by the amount of f_l Hz in the frequency domain. The modulated signal $y_1(t)$ of a signal $x(t)$ is expressed as :

$$y_1(t) = m_c(t)x(t) = \exp(-j2\pi f_1 t)x(t). \quad (4)$$

For example, if $x(t)$ represents the analytic human vocal signal as in Fig. 3(a), after applying the conventional modulation (with $f_l = 230$) as in (4), the STFT of the resultant signal $y_1(t)$ is shown as in Fig. 3(b). From Fig. 3(b), one can see that the signal part originally located around **180~280Hz** is shifted to a much lower frequency location around **-55~55Hz**.

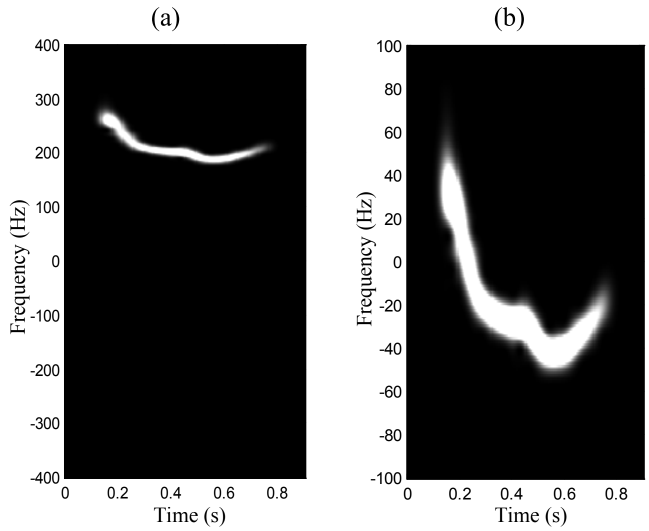

Fig. 3 (a) The STFT of the human vocal signal in Fig. 2(a) after analytic signal generation. (b) The STFT of the analytic signal in (a) after conventional modulation.

In order to discuss the noise-removing process, AWGN (additive white Gaussian noise) is then added to the human vocal signal. In Fig. 4(a) we show the STFT of this noisy signal. Since the added noise is AWGN, we can see that the noise is distributed over all frequencies. For measuring the error, we use NMSE (Normalized Mean Square Error) to represent the degree of the effect of noise. The NMSE is calculated as follows:

$$NMSE = \frac{\sum_i (y[i]-x[i])^2}{\sum_i x^2[i]} \tag{5}$$

where $x[i]$ is the original signal and $y[i]$ is the noisy version of the signal. Computation shows that the NMSE of the noisy signal is about **27.61%**.

If we attempt to use a lowpass filter for removing the noise right away, the cutoff frequency of the filter should be above **280Hz,** which is half the signal bandwidth (the bandwidth is about **580Hz** as in Fig. 4(a)). After the lowpass filtration, only the noise part within the signal bandwidth remains. Fig.4 (b) indicates the remaining part after using the lowpass filter. Computation shows the NMSE of remaining part is reduced to **1.83%**.

To further reduce the amount of noise by the same concept, we need to obtain as narrower signal bandwidth as we can. Thus, we apply the analytic signal generation and the conventional modulation to the noisy signal, and the resultant signal is as Fig. 5(a) shows. From Fig. 5(a), one can see that the signal bandwidth is now only around **110Hz,** which is a much smaller value compared with the one in Fig. 4. Now a lowpass filter with only **55Hz** cutoff frequency is needed to remove the noise while keeping the complete information of recovering the signal. Fig. 5(b) shows the remaining part. After the lowpass filtration and signal reconstruction, the NMSE of the recovered signal is computed to be only **0.92%**.

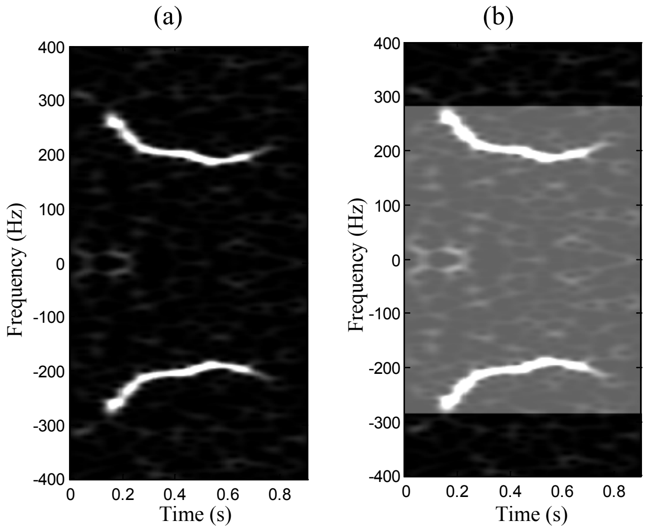

Fig. 4 (a) The STFT of the noisy human vocal signal **(NMSE=27.61%)**. (b) The remaining part of the signal in (a) after using a proper lowpass filter (cutoff frequency = **280Hz**) **(NMSE=1.83%)**.

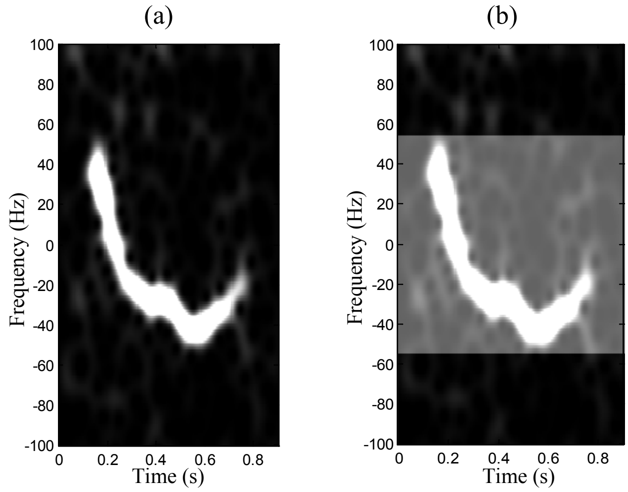

Fig. 5 (a) The STFT of the noisy signal after analytic signal generation + conventional modulation. (b) The remaining part of the signal in (a) after using a proper lowpass filter (cutoff frequency = **55Hz**) **(NMSE=0.92%)**.

Although the signal bandwidth can be significantly reduced by the analytic signal generation and the conventional modulation operation, it is still possible for us to further narrow the signal bandwidth and hence achieve an extremely low error, by the proposed generalized modulation which we introduce in the next section. In [15], we use the generalized modulation for reducing the number of sampling points of a signal. In fact, it can also be applied for noise reduction.

III. Generalized Modulation for Noise Reduction

The goal of the generalized modulation operation is to make the bandwidth of each signal part as small as possible (Remember that narrower bandwidth means that less amount of noise will remain after being filtered).

The conventional modulation operation is to multiply the signal by a linear phase exponential function, as in (4). Here, instead of (4), we perform the generalized modulation operation and multiplying $x(t)$ by a higher order exponential function, $m_g(t)$:

$$x(t) \xrightarrow{\text{modulation}} y(t) = m_g(t)x(t) \tag{6}$$

where

$$m_g(t) = \exp[-j2\pi(a_n t^n + a_{n-1}t^{n-1} + \cdots\cdots + a_1 t + a_0)]. \tag{7}$$

Note that the phase of $m_g(t)$ is an n^{th} order polynomial. Since the instantaneous frequency of $m_g(t)$ is the derivative of the phase with respect to time, the instantaneous frequency, f_{in}, can be derived as :

$$\begin{aligned} f_{in} &= \frac{1}{2\pi}\frac{d}{dt}\arg\left(m_g(t)\right) \\ &= \frac{d}{dt}[-(a_n t^n + a_{n-1}t^{n-1} + \cdots\cdots + a_1 t + a_0)] \\ &= -[na_n t^{n-1} + (n-1)a_{n-1}t^{n-2} + \cdots\cdots + a_1]. \end{aligned} \tag{8}$$

If $STFT_x(t, f)$ and $STFT_y(t, f)$ are the STFTs of $x(t)$ and $y(t)$, respectively, then

$$STFT_y(t,f)$$
$$\cong STFT_x(t, f+f_{in})$$
$$= STFT_x(t, f-[na_n t^{n-1}+(n-1)a_{n-1}t^{n-1}+\cdots\cdots+a_1]) \,. \quad (9)$$

Therefore, with the generalized modulation in (7), one can adjust the "shape" of the time-frequency distribution of a signal more flexibly. Namely, we can "reshape" and thus "minimize" the areas of the spectrogram of the signal. In other words, the generalized modulation has higher ability to reduce the bandwidth requirement for a signal and the amount of noise that would remain after lowpass filtration.

For example, for the human vocal signal whose STFT is as in Fig. 3(a), its central frequency is plotted as in Fig. 6(a). Note that the central frequency varies with time. Then, in Fig. 6(b), we use a 6th order polynomial as follows to approximate the central frequency curve in Fig. 6(a):

$$P_6(t) = -124500t^6 + 336810t^5 - 360660t^4 + 193720t^3 - 53944t^2 + 6987t - 64 \,. \quad (10)$$

Such a polynomial is called the "fitting polynomial" since it approximates the central frequency and "fits" it well. The approximation is performed by Legendre polynomial expansion [16]. That is, if the central frequency of the signal is $h(t)$, then the n^{th} order fitting polynomial used for approximating $h(t)$ can be determined from:

$$P_n(t) = \sum_{k=0}^{n} a_k \varphi_k(t) \,, \quad (11)$$

$$\text{where} \quad a_k = \int_{t_0}^{t_0+T} h(t)\phi_k(t)dt \,, \quad (12)$$

$$\varphi_k(t) = \sqrt{\frac{2}{T}} L_k\left(\frac{t-t_0-T/2}{T/2}\right). \quad (13)$$

$[t_0, t_0+T]$ is the support of $h(t)$, and $\{ L_k(t) \mid k = 0, 1, 2, \ldots\}$ is the Legendre polynomial set that is orthonormal in the interval of $t \in [-1, 1]$. Since

$$\int P_6(t)dt = -17785t^7 + 56135t^6 - 72132t^5 + 48429t^4 - 17981t^3 + 3494t^2 - 64t \,, \quad (14)$$

from (7)-(9), after obtaining and integrating the fitting polynomial, we then apply the generalized modulation to the analytic signal by

$$x_a(t) \xrightarrow{\text{modulation}} y_2(t) = m_g(t)x_a(t) \quad (15)$$

where $x_a(t)$ represents the analytic signal as in Fig. 3(a) and

$$m_g(t) = \exp[-j2\pi \int P_7(t)dt]$$
$$= \exp[-j2\pi(-17785t^7 + 56135t^6 - 72132t^5 + 48429t^4 - 17981t^3 + 3494t^2 - 64t)] \,. \quad (16)$$

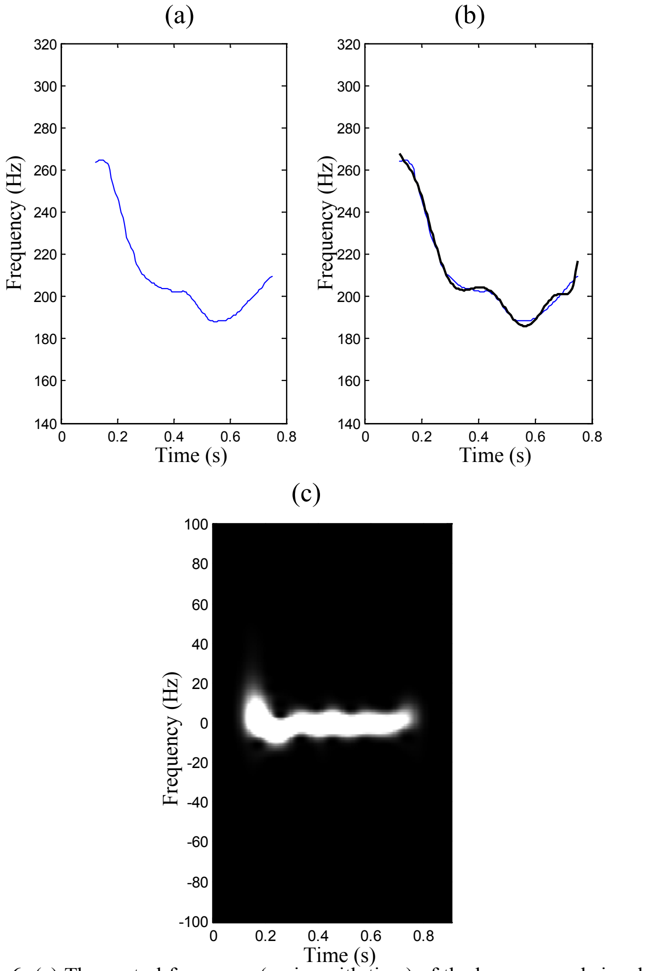

Fig. 6 (a) The central frequency (varies with time) of the human vocal signal whose STFT is as in Fig. 3(a). (b) Using a 6th order polynomial to approximate the central frequency of the vocal signal. (c) The STFT of the signal in Fig. 3(a) after generalized modulation.

The STFT of the resultant signal $y_2(t)$ is plotted as in Fig. 6(c). From Fig. 6(c), one can see that the signal is reshaped to almost a straight line, and the bandwidth becomes much narrower than the one in Fig. 3(b), which is derived by the conventional modulation.

For illustrating our noise-removing process, we apply the analytic signal generation and the generalized modulation to the noisy human vocal signal in Fig. 4(a). The STFT of the resultant signal is plotted as in Fig.7 (a). Again, the narrower the bandwidth is, the less the noise is contained. Now, since the signal bandwidth is only about **36Hz**, we can use a lowpass filter with extremely low cutoff frequency (about only **18Hz**) to filter out the whole signal while keeping the information we need to recover it. In Fig. 7(b), the remaining part is as indicated. After the lowpass filtration and reconstruction, the NMSE of the noise-removed signal is computed and is shown to be only **0.41%**, which is a much smaller value compared with all the error values mentioned previously. Therefore, with the proposed generalized modulation operation, the bandwidth of a signal can be much reduced and the noise-removing performance is obviously improved.

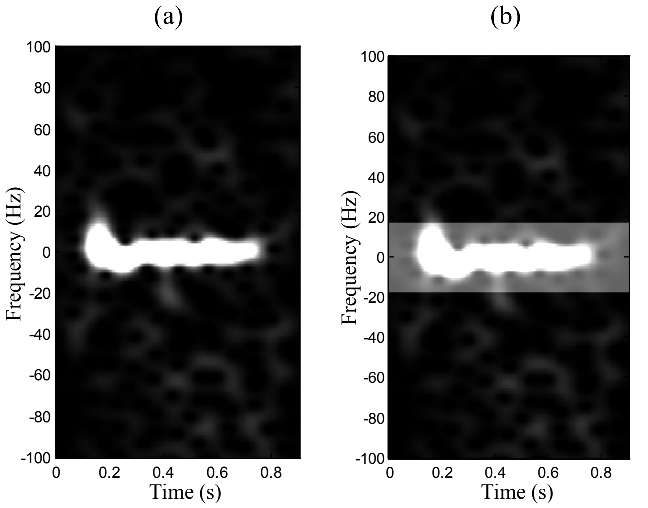

Fig. 7 (a) The STFT of the noisy human vocal signal in Fig. 4(a) after analytic signal generation + generalized modulation. (b) The remaining part of the signal in (a) after using a proper lowpass filter (cutoff frequency = **18Hz**) **(NMSE=0.41%).**

IV. Combining Generalized Modulation with the Fractional Fourier Transform

In [8-12], the fractional Fourier transform (FRFT) was adopted to rotate the time-frequency distribution of a signal and improve the sampling efficiency. In this paper, we find that, combining the proposed generalized modulation with the FRFT, a very narrow signal bandwidth can be achieved and the noise-removing performance can be further improved.

The definition of the FRFT is:

$$O_{FRFT}^{\alpha}\left[x(t)\right] = \sqrt{1-j\cot\alpha}\int_{-\infty}^{\infty} e^{j\pi u^2 \cot\alpha - j2\pi u t \csc\alpha + j\pi t^2 \cot\alpha}\, x(t)\, dt \quad . \qquad (17)$$

FRFT is useful for signal decomposition and segmentation. It can be viewed as performing the Fourier transform $2\alpha/\pi$ times. From [9], one can see that the FRFT has very close relations with the time-frequency distribution. If

$$y(t) = O_{FRFT}^{-\alpha}\left\{O_{FRFT}^{\alpha}\left[x(t)\right]H(u)\right\} \qquad (18)$$

where $H(u) = 1$ for $u < u_0$ and $H(u) = 0$ for $u > u_0$, then the FRFT filter in (18) is equivalent to placing a separating line in the time-frequency domain. The angle between the line and f-axis is α and the distance between the line and the origin is u_0.

Besides signal decomposition and segmentation, the FRFT plays another role in our noise-removing scheme: To assist the generalized modulation in reducing the signal bandwidth by "rotating" the time-frequency distribution in advance.

Note that, for the human vocal signal in Fig. 3(b), it has very large negative slope around 0.2 second, which causes the modulated signal as in Fig. 6(c) to be out-of-flatness around 0.2 second. Therefore, it is more proper to rotate the time-frequency distribution of the signal before performing generalized modulation for the signal, in order to alleviate the effect of large slope on the modulated signal. The rotation in the time-frequency domain can be done by the FRFT.

In Fig. 8(a) and Fig. 8(b), we show the STFTs of $y_1(t)$ and $y_3(t)$, respectively, where $y_1(t)$ is defined as in (4), which is generated from the human vocal signal by performing the analytic signal conversion and the conventional modulation, and $y_3(t)$ is the FRFT of $y_1(t)$:

$$y_3(t) = O_{FRFT}^{-0.2}\left[y_1(t)\right]. \qquad (19)$$

Then, according to the 6[th] order polynomial that can approximate the central frequency of $y_3(t)$, one can perform the following generalized modulation operation for $y_3(t)$:

$$y_3(t) \xrightarrow{\text{mod}\,ulation} y_4(t) = m_g(t)\,y_3(t) \qquad (20)$$

where

$$m_g(t) = \exp[-j2\pi(-2920t^7 + 9845t^6 - 13076t^5 + 8708t^4 - 2963t^3 + 388t^2 + 19t)]\,. \qquad (21)$$

The central frequency and the fitting polynomial are shown as in Fig. 8(c).The STFT of $y_4(t)$ is plotted as in Fig. 8(d). In Fig. 8(d), the signal becomes flatter and smoother around 0.2 second, and hence the bandwidth is further reduced.

For illustrating our noise-removing process, we apply our method to the noisy human vocal signal in Fig. 4(a). The STFT of the resultant signal is shown as in Fig. 9(a). Now, we can see that the bandwidth is further reduced (to only **24Hz**) and we can use a lowpass filter with only **12Hz** cutoff frequency to remove the noise, as Fig. 9(b) shows. After the lowpass filtration and reconstruction, the NMSE of the recovered signal is computed to be only about **0.41%**.

In Fig. 10, we restate our noise-removing algorithm with the aid of several diagrams. Since the analytic operation can remove the negative part, as in Fig. 10(b), and the FRFT can rotate the time frequency distribution of a signal, as in Fig. 10(c), they are helpful for reducing the signal bandwidth. However, even in Fig. 10(c), the rectangular region circled by the dash lines still contains a lot of non-signal parts. Thus, it is proper to use the proposed generalized modulation scheme to "re-shape" the time-frequency distribution of the signal. After the proposed generalized modulation is applied, the signal bandwidth and hence the amount of the noise contained in the signal can both be minimized, as in Fig. 10(d).

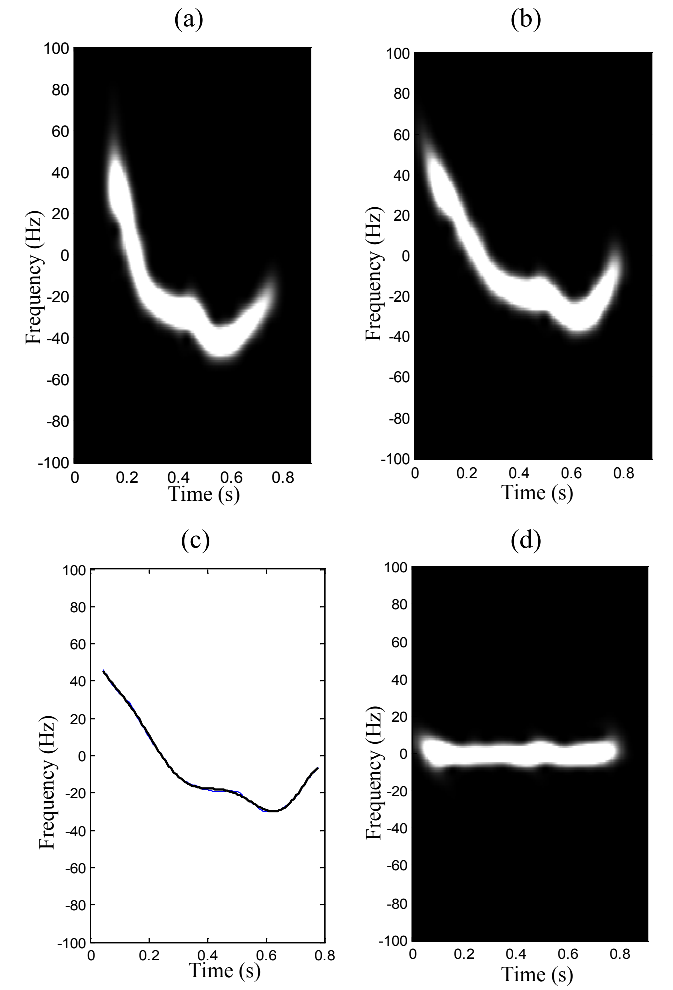

Fig. 8 (a) The STFT of the human vocal signal in Fig. 2(a) after analytic signal generation + conventional modulation. (b) After performing the FRFT, the STFT in (a) is rotated. (c) Using a 6th order polynomial (black line) to approximate the central frequency (blue line) of the signal in Fig. 6(b). (d) The STFT of the signal in (a) after the FRFT + generalized modulation.

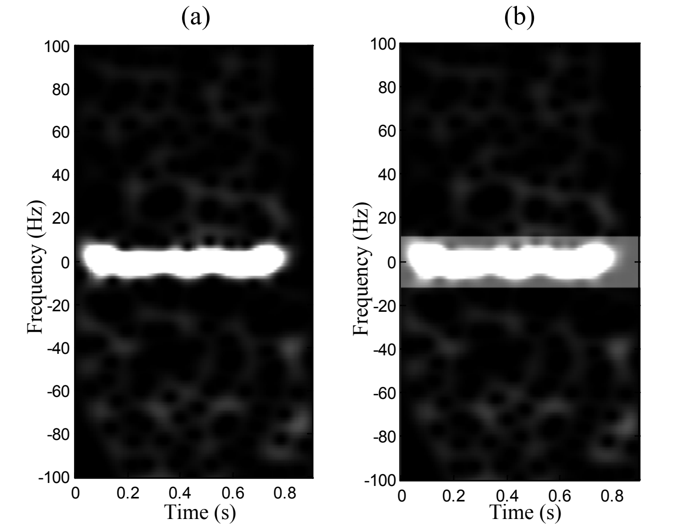

Fig. 9 (a) The STFT of the noisy human vocal signal in Fig. 4(a) after conventional modulation + FRFT + generalized modulation. (b) The remaining part of the signal in (a) after using a proper lowpass filter (cutoff frequency = 12Hz) **(NMSE=0.41%).**

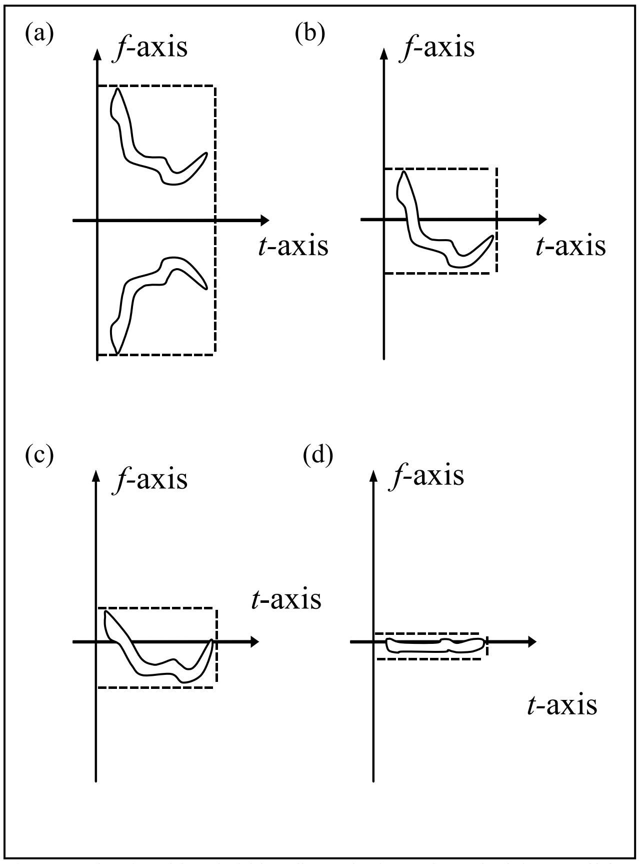

Fig. 10 The signal bandwidth reflects the lower bound of the cutoff frequency of the low-pass filter, which is used to reduce the noise while keeping complete information for recovering the signal. For each noise-removing scheme, the part that remains after the lowpass filtration is depicted by the area of the block circled by the dot lines. (a) The original signal. (b) Analytic signal conversion + conventional modulation. (c) Analytic signal conversion + FRFT + conventional modulation. (d) Analytic signal conversion + FRFT + proposed generalized modulation.

V. SIMULATION RESULTS

In our simulations, we add the AWGN with different average powers to the human vocal signal to see the performances of different noise-removing schemes under various noise conditions. The noise reduction performances are measured by the NMSE. We show the MSEs of the noisy signal without any modifications (**Original Error**), the noisy signal after a proper lowpass filter (**LPF**), the noisy signal modified by the noise-removing scheme based on the conventional modulation (**Analytic + Conventional Modulation + LPF**), the noisy signal modified by the noise-removing algorithm based on the proposed generalized modulation (**Analytic + Proposed Generalized Modulation + LPF**), and the noisy signal modified by the proposed noise-removing algorithm as illustrated in Fig. 11 **(Analytic + FRFT + Proposed Generalized Modulation + LPF)**. The result in **Fig. 11** shows that the proposed noise-removing method (**Analytic + FRFT + Proposed Generalized Modulation + LPF**) can indeed significantly reduce the amount of noise and achieve the best accuracy.

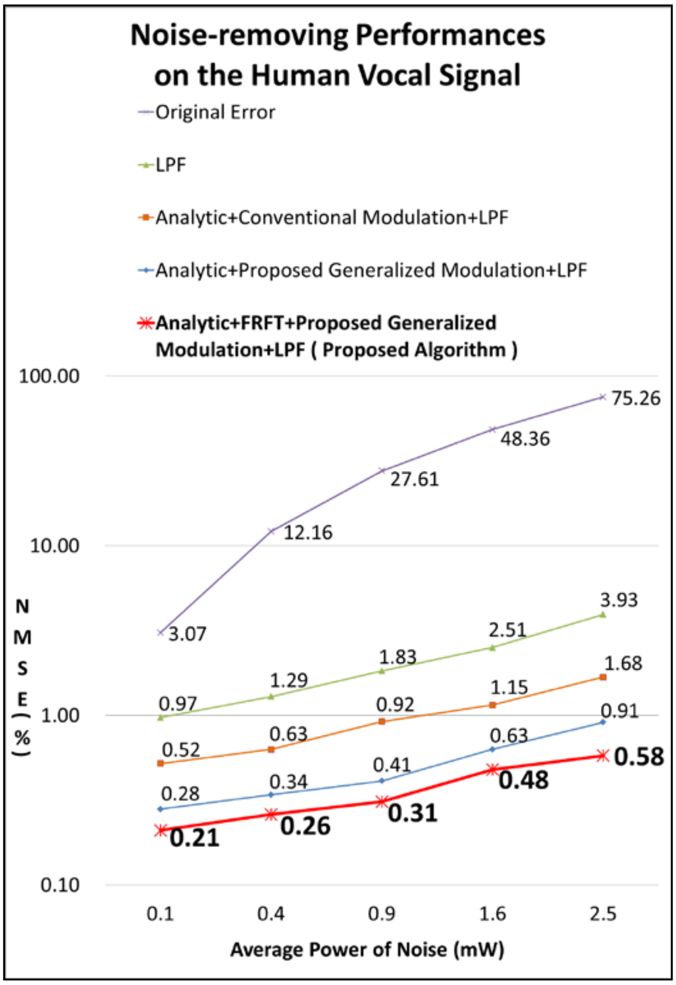

Fig. 11 Comparison among the performances of different noise-removing schemes on the human vocal signal.

In Fig. 12, Fig. 14, and Fig. 16, another three simulations are performed. The input signals for Fig. 12 and Fig. 14 are another two human vocal signals. The input signal for Fig. 16 is a whale voice signal acquired from [17], which is also a time-variant signal. There simulation results are shown in Fig. 13, Fig. 15, and Fig. 17, respectively. All of the simulation results show that the proposed noise–removing algorithm based on the generalized modulation is very helpful for reducing the effects of noise on time-variant signals

VI. Conclusion

A new noise-removing algorithm is proposed, which is the combination of the STFT, analytic signal conversion, the FRFT filter, and the generalized modulation operation. With the proposed algorithm, the area of the signal spectrogram is reshaped and the bandwidth of the signal can be further minimized. With an appropriate lowpass filter, considerable amount of noise can be removed and a cleaner signal is obtained. Simulation results show that the proposed method significantly reduces the effect of noise on the speech signal. In addition to speech signals, we also show that our method also performs well on other kind of time-variant signals, such as the whale voice signal. In conclusion, the proposed noise-removing algorithm based on the generalized modulation can well separate the signal part and the noise part of a time-variant signal and a better noise-reducing performance can then be achieved.

Acknowledgment

This work was supported by National Science Council, Taiwan, under grants 100-2221-E-002-206-MY3.

References

[1] J. Chen, J. Benesty, Y. Huang, and S. Doclo, “New insights into the noise reduction Wiener filter,” *IEEE Trans. Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1218-1234, July 2006.

[2] B. A. Shenoi, *Introduction to Digital Signal Processing and Filter Design*, Wiley-Interscience, N. J., 2006.

[3] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, London: Prentice-Hall, 3rd ed., 2010.

[4] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE Trans. Acoustic Speech Signal Processing*, vol. 27, no. 2, pp. 113-120, Apr. 1979.

[5] R. Martin, “Noise power spectral density estimation based on optimal smoothing and minimum statistics,” *IEEE Trans. Acoustic Speech Signal Processing,* vol. 9, no. 5, pp. 504-512, July 2001.

[6] S. C. Pei and J. J. Ding, “Fractional Fourier transform, Wigner distribution, and filter design for stationary and nonstationary random processes,” *IEEE Trans. Signal Processing*, vol. 58, no. 8, pp. 4079-4092, Aug. 2010.

[7] S. Haykin, *Communication Systems*, New York: John Wiley & Sons, 5th ed., 2009.

[8] H. M. Ozaktas, Z. Zalevsky, and M. A. Kutay, *The Fractional Fourier Transform with Applications in Optics and Signal Processing*, New York: John Wiley & Sons, 2000.

[9] S. C. Pei, M. H. Yeh, and C. C. Tseng, “Discrete fractional Fourier transform based on orthogonal projections,” *IEEE Trans. Signal Processing*., vol. 47, no. 5, pp. 1335-1348, May 1999.

[10] M. A. Kutay, H. M. Ozaktas, O. Arikan, and L. Onural, “Optimal filter in fractional Fourier domains,” *IEEE Trans. Signal Processing*, vol. 45, no. 5, pp. 1129-1143, May 1997.

[11] Z. Zalevsky and D. Mendlovic, “Fractional Wiener filter,” *Appl. Opt.*, vol. 35, no. 20, pp. 3930-3936, Jul. 1996.

[12] S. C. Pei and J. J. Ding, “Relations between Gabor transforms and fractional Fourier transforms and their applications for signal processing,” *IEEE Trans. Signal Processing*, vol. 55, no. 10, pp. 4839-4850, Oct. 2007.

[13] K. Grochenig, *Foundations of Time-Frequency Analysis*, Boston: Birk-hauser, 2001.

[14] M. J. Bastiaans, “Gabor’s expansion of a signal into Gaussian elementary signals,” *Proc. IEEE*, vol. 68, pp. 594-598, Apr. 1980.

[15] J. J. Ding, S. C. Pei, and T. Y. Ko, “Higher order modulation and the efficient sampling algorithm for time variant signal,” *European Signal Processing Conference*, pp. 2143-2147, Bucharest, Romania, Aug. 2012.

[16] M. R. Spiegel, *Mathematical Handbook of Formulas and Tables*, New York: McGraw-Hill, 1990.

[17] The whale voice data can be available from http:// oalib.hlsresearch.com/Whales/index.html.

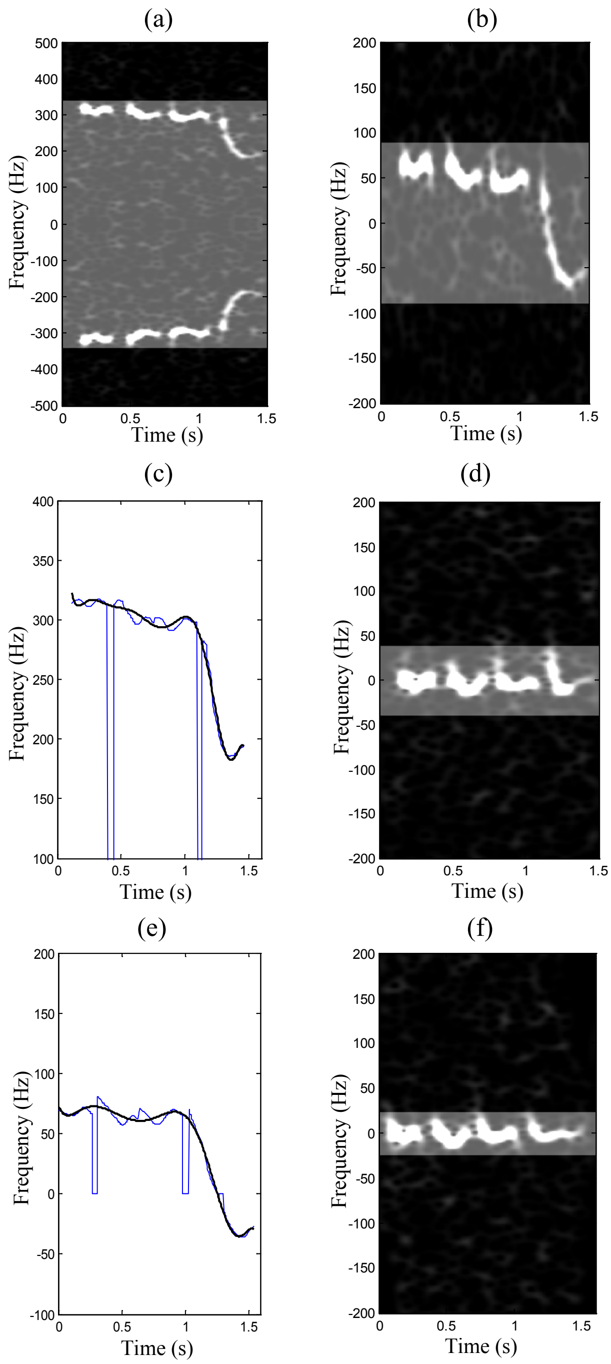

Fig. 12 (a) Lowpass filtered part of Human Vocal Signal 2. (b) Lowpass filter output of the signal in (a) after analytic signal generation + conventional modulation. (c) The fitting polynomial (black line) of the center frequency (blue line) of the signal in (a). (d) Lowpass filter output of the signal in (a) after analytic signal generation + generalized modulation. (e) The fitting polynomial (black line) of the center frequency (blue line) of the signal in (a) after analytic signal generation + conventional modulation + FRFT. (f) Lowpass filter output of the signal in (a) after analytic signal generation + conventional modulation + FRFT + generalized modulation.

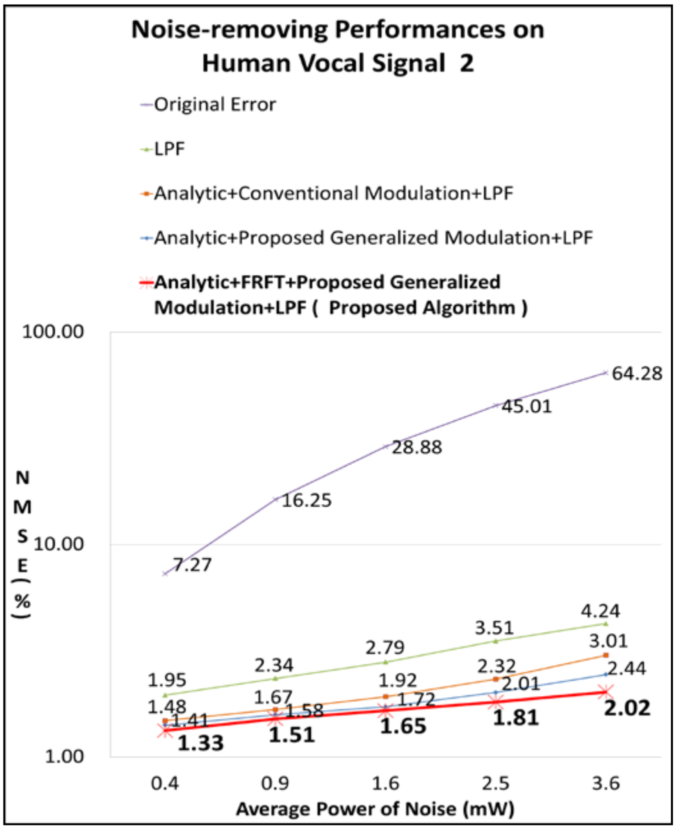

Fig. 13 Comparison among the performances of different noise-removing schemes on Human Vocal Signal 2.

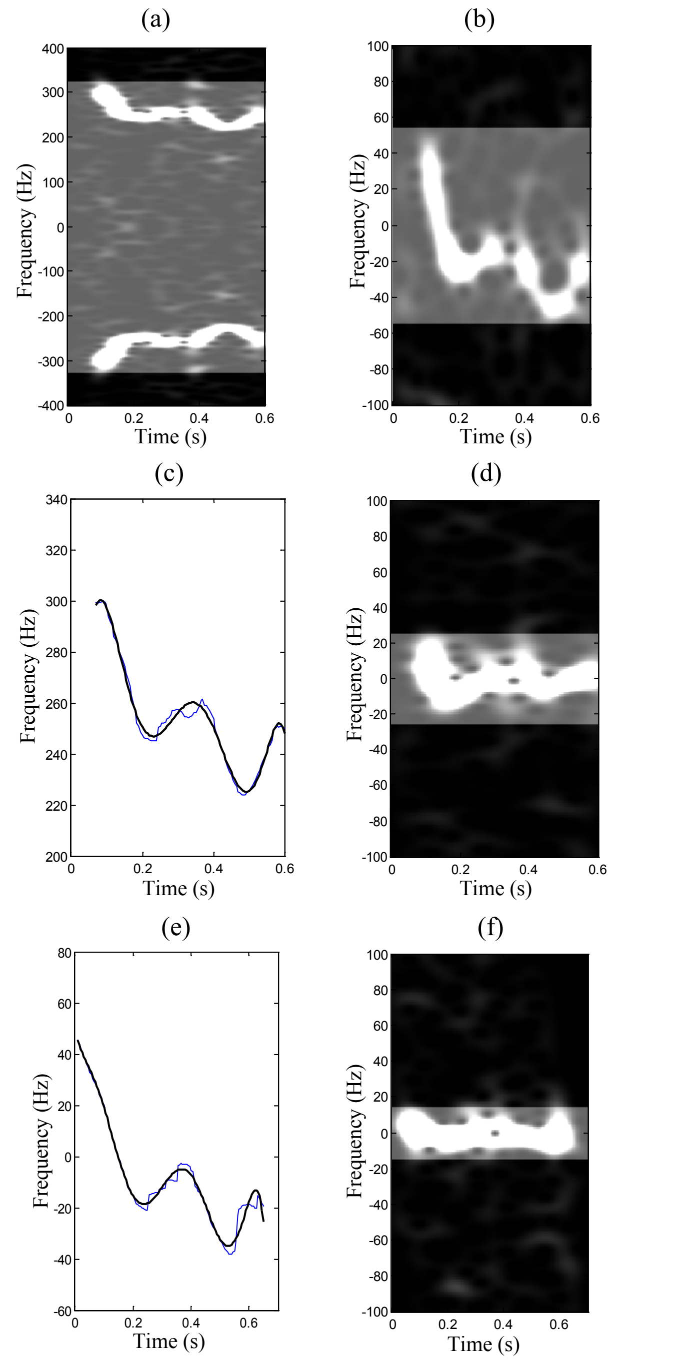

Fig. 14 (a) Lowpass filtered part of Human Vocal Signal 3. (b) Lowpass filter output of the signal in (a) after analytic signal generation + conventional modulation. (c) The fitting polynomial (black line) of the center frequency (blue line) of the signal in (a). (d) Lowpass filter output of the signal in (a) after analytic signal generation + generalized modulation. (e) The fitting polynomial (black line) of the center frequency (blue line) of the signal in (a) after analytic signal generation + conventional modulation + FRFT. (f) Lowpass filter output of the signal in (a) after analytic signal generation + conventional modulation + FRFT + generalized modulation.

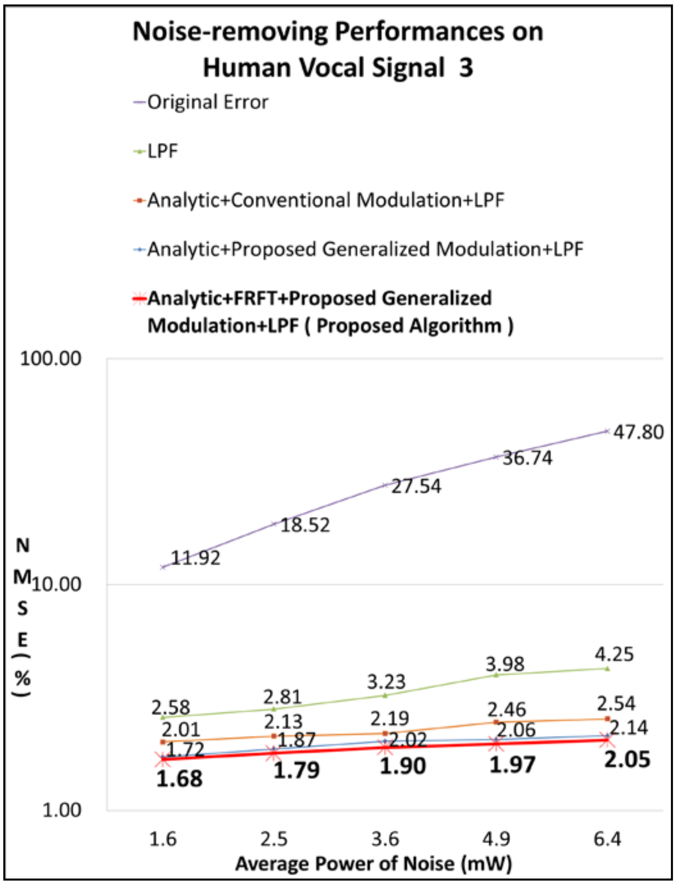

Fig. 15 Comparison among the performances of different noise-removing schemes on Human Vocal Signal 3.

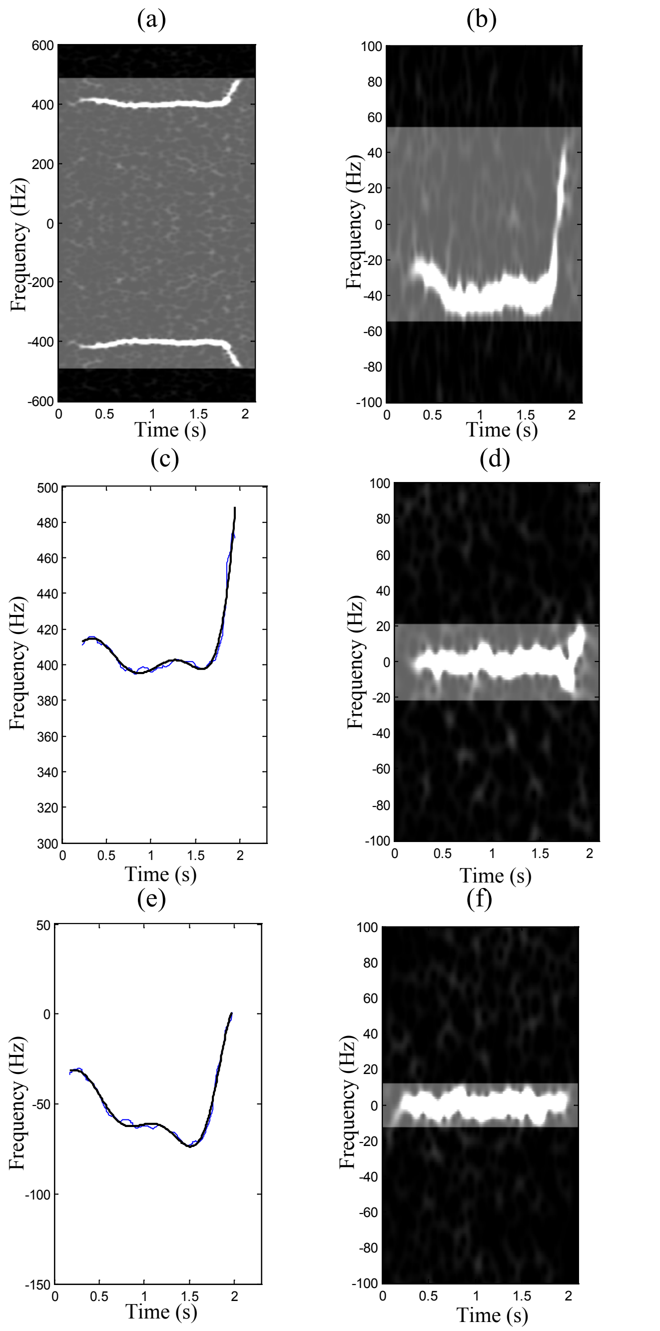

Fig. 16 (a) Lowpass filtered part of a whale voice signal. (b) Lowpass filter output of the signal in (a) after analytic signal generation + conventional modulation. (c) The fitting polynomial (black line) of the center frequency (blue line) of the signal in (a). (d) Lowpass filter output of the signal in (a) after analytic signal generation + generalized modulation. (e) The fitting polynomial (black line) of the center frequency (blue line) of the signal in (a) after analytic signal generation + conventional modulation + FRFT. (f) Lowpass filter output of the signal in (a) after analytic signal generation + conventional modulation + FRFT + generalized modulation.

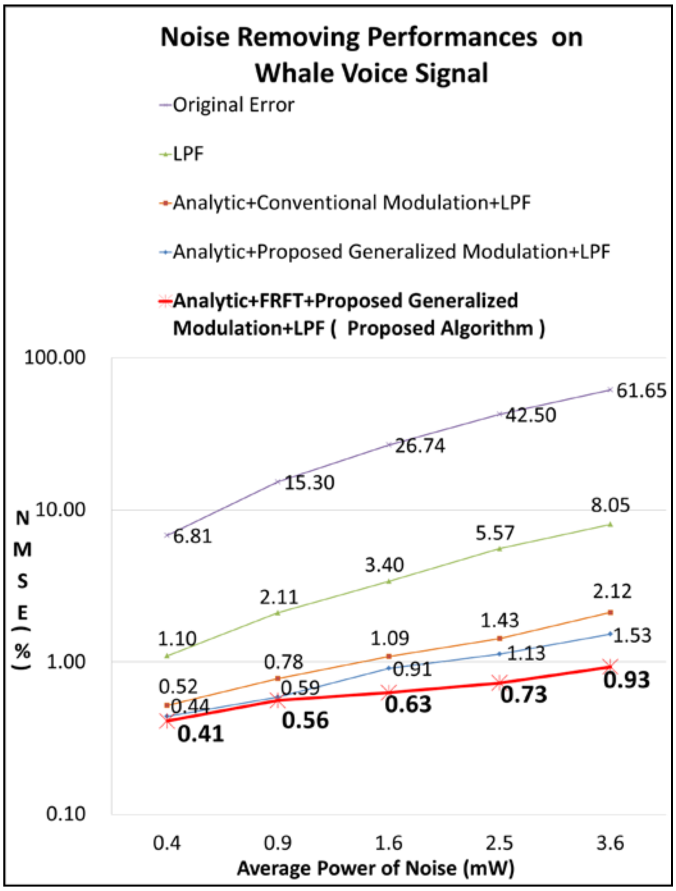

Fig. 17 Comparison among the performances of different noise-removing schemes on a whale voice signal.