TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol. 12, No. 12, December 2014, pp. 8238 ~ 8245
DOI: 10.11591/telkomnika.v12i12.6694

Speech Enhancement Research Based on FRFT

Jingfang Wang
School of Information Science and Engineering, Hunan International Economics University, Changsha, China, postcode: 410205
email: matlab_bysj@126.com

Abstract

As many traditional de-noising methods fail in the intensive noises environment and are unadaptable in various noisy environments, a method of speech enhancement has been advanced based on dynamic Fractional Fourier Transform (FRFT) filtering. The acoustic signals are framed. The renewing methods are put in FRFT optimal disperse degree of noising speech and this method is implemented in detail. By TIMIT criterion voice and Noisex-92, the experimental results show that this algorithm can filter noise from voice availably and improve the performance of automatic speech recognition system significantly. It is proved to be robust under various noisy environments and Signal-to-Noise Ratio (SNR) conditions. This algorithm is of low computational complexity and briefness in realization.

Keywords: *acoustic signal, fractional Fourier transform (FRFT), speech enhancement, filtering de-noising, auto-adaptive processing*

1. Introduction

With the development of communication technology, speech communication has become a major medium of communication for people to pass information more convenient. However, the widespread noise in nature makes the decline in speeh communication quality. In order to reduce the effect of noise on speech communications performance and improve the quality of speech communication, S. Boll et al [1] proposed a classical spectral subtraction algorithm (Spectral Subtraction, SS) in 1979, the algorithm is under the assumption of additive noise and short-term steady voice signal independent conditions, the estimated noise spectrum is subtracted from the noisy speech signal spectrum, speech denoised signal spectrum is obtained. However, due to its local stationarity assumption is not consistent with the actual situation, so the effect is not ideal, leading to a larger residual musical noise and other issues. On the basis of the traditional spectral subtraction, Berouti [2] increased the adjustment coefficient of the noise power spectrum and enhanced the minimum limit of the speech power spectrum, and it is to improve the performance of spectral subtraction. However, since the correction factor an it minimum are determined according to the experience, and there is poor adaptability in the method. In 1984, the minimum mean square error is introduced to a spectral subtraction by Y. Ephraim et al [3], it is possible to solve the music noise portion, the effect of the denoising is of improved. However, the distribution of the speech spectrum is required prior to be estimated in this algorithm, and calculation is relatively large. On the basis of the spectral subtraction, P. Lochwood et al [4] gave speech enhancement gain function adaptively based on speech signal to noise ratio, and nonlinear spectral subtraction algorithm (Nonlinear Spectral Subtractor, NSS) is proposed, although the SNR of speech is improved in the algorithmvoice, but the audio quality is not improved. To further reduce musical noise and improve voice clarity, people continue to make a variety of improvements based on traditional spectral subtraction algorithm [5-7], the quality of voice is improved better. However, when the signal to noise ratio is low or in a non-stationary noise environment, the performance of the conventional spectral subtraction tends to become poor. In 2002, S. Kamath et al [8] proposed a multi-band spectral subtraction based on iterative methods. This method takes into account the colored noise spectral inhomogeneity effects on speech, introducing the twiddle factor and the band segment treatment, while maintaining high voice quality, the background noise and music noise can be effectively eliminated under colored noise pollution. There are maximum posteriori estimation method based on speech generation model [9], and Kalman filters, etc. [10-12], the voice

Received May 19, 2014; Revised October 22, 2014; Accepted November 16, 2014

generation process can be modeled as a linear variable filter, the different excitation sources are used for different types of voice.

In 1995, Y. Ephraim, etc. build a new avenue in the time domain for voice enhancement (frame theory subspace) [13] for the first time, we propose speech enhancement algorithm based on subspace signal. Y. Ephraim initial work was mainly for white noise, in order to deal with the case of non-white noise, in 2000, Mrital [14] proposed a signal / noise KL transform, despite the enhanced signal had a small residual noise for each frame, but the non stability of residual noise between frames is disturbing. In 2001, A.Rezaye and S. Gazor [15] proposed an adaptive KLT method for processing non-stationary noise, they assume that the feature vectors for the speech signal can be approximated to covariance matrix diagonalization of non-stationary colored noise. In 2003, based on signal subspace decomposition and for lack of Rezayee method, speech enhancement algorithm is proposed by Y. Hu et al [16] for colored noise in the time domain and frequency domain. In the same year, A. Lev and Y. Ephraim [17] have also proposed approach for colored noise. However, the premise of the above methodsrequire that noise covariance matrix must be full rank, which is not applicable for narrowband noise. We studied FRFT (Fractional Fourier Transform) filter method, and after to test in various noise environment, the effect is good, computational cost is small in the proposed algorithm, it is simple and easy to achieve.

2. Fractional Fourier Transform (FRFT)

Fractional Fourier Transform (Fractional Fourier Transform, FRFT) is a recently developed, and it is a new time-frequency analysis tool, it is a generalized form of Fourier transform. Essentially, the signal is representation on the fractional Fourier domain, while it is the integration of the signal information in the time domain and frequency domain. This new mathematical tools not only is closely linked with the Fourier transform, but also there are also very meaningful contact with other time-frequency analysis tools, it has been widely used in optical systems analysis, filter design, signal analysis, solving differential equations, phase recovery, pattern recognition and other fields [18-20]. In recent years, application of fractional Fourier transform is mostly concentrated in the linear FM signal estimation, detection and filtering aspects.

FRFT can be interpreted as signal representation in the composition of the fractional Fourier domain when the counterclockwise rotation is done the time-frequency plane at any angle. FRFT is a generalized form of Fourier transform. FRFT signal is defined as [20].

$$X_{\alpha}(u)=\left\{F^{\alpha}\left[x(t)\right]\right\}(u)=\int_{-\infty}^{\infty}x(t)K_{\alpha}(t,u)\,dt \tag{1}$$

Where the transform kernel FRFT $K_{\alpha}(t,u)$ is:

$$K_{\alpha}(t,u)=\begin{cases}\sqrt{\dfrac{1-j\cot\alpha}{2\pi}}\exp\left(j\dfrac{t^2+u^2}{2}\cot\alpha-tu\csc\alpha\right), & \alpha\neq n\pi\\ \delta(t-u), & \alpha=2n\pi\\ \delta(t+u), & \alpha=(2n\pm1)\pi\end{cases} \tag{2}$$

Where $\alpha=p\pi/2$ is FRFT rotation. Signal $x(t)$ Return to:

$$x(t)=\{F^{-\alpha}[X(u)]\}(t)=\int_{-\infty}^{\infty}X(u)K_{-\alpha}(t,u)\,du \tag{3}$$

3. FRFT Dynamic Filtering

The energy accumulation of Fractional Fourier transform is related to transformation

order α, and its aggregation is strongly depends on the extent of its close to Fourier transform. Fractional Fourier transform of the speech signal has a certain energy in both voiced and unvoiced focus, the energy difference is different in their focus areas: voiced focused energy reflects the central region of the waveform on Fractional transform domain, the voiceless focusing energy is reflected at both ends of the region of the waveform. Fractional Fourier Transform of white noise is not the nature of the energy focus, the focus energy is poor on the noise, it can be used to de-noising in voice signal.

3.1. The Best Fractional Order α FRFT

For different segments and noise pollution signals, FRFT transformation is made in different fractional α, then the effective filtering is done. What is measured? It is common MMSE (minimum mean square error, MMSE), the energy focus degree is measured by the weighted variance in this paper.

Fractional α order Fourier transform of 2N point signal $x(t)$ is: $X_{\partial}(k), k = 1,2,\ldots,2N$, its half is taken because the center symmetric. Its normalized Probability is:

$$p_{\alpha}(k) = \frac{|X_{\alpha}(k)|}{\sum_{i=1}^{N} |X_{\alpha}(i)|} \quad k = 1,2,\ldots,N \tag{4}$$

$$EX = \sum_{k=1}^{N} k p_{\alpha}(k), \; Var(X,\alpha) = \sum_{k=1}^{N} (k - EX)^2 p_{\alpha}(k) \tag{5}$$

Taking α_i, $0 < \alpha_i < 1$, to calculate the weighted variance Var (X, α_i), then these data are fitted in cubic spline, and the minimum value Var (X, α_0) is sought in the Var (X, α), so α_0 is of the corresponding best fractional order.

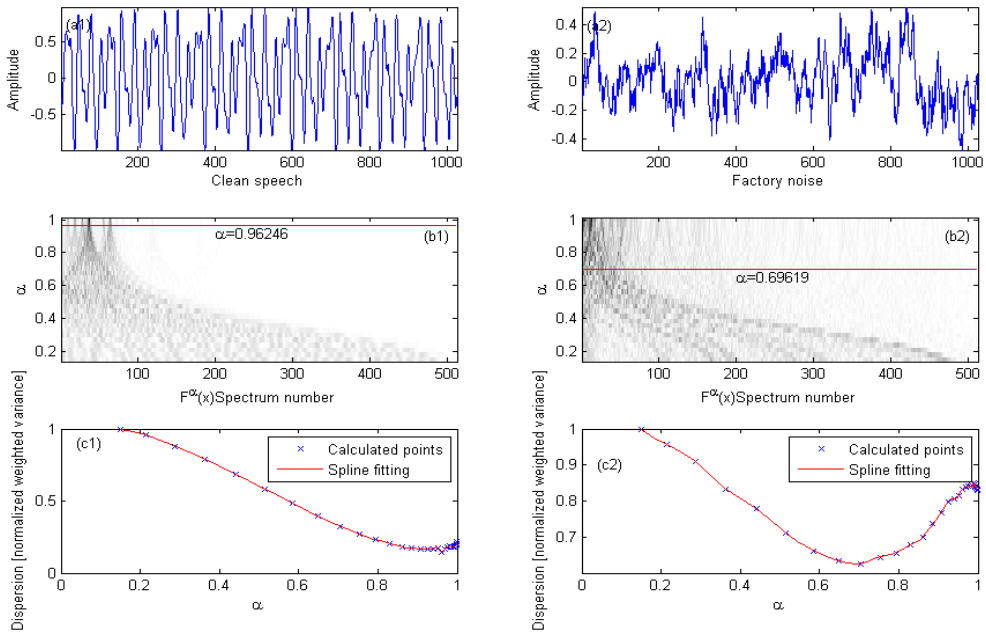

Figure 1. According to the weighted variance, best FRFT Fractional order α contrast of speech and noise

In Figure 1, (a1) is speech signal, (b1) is its α-order fraction Fourier transform, it changes in the energy process of gradual accumulation from the time domain to frequency

domain; (c1) are Var (X, α) and α Trends and the cubic spline fitting by Formula (5). (a2) is a factory noise, (b2) is corresponding to the α-order fractional Fourier transform, it changes in the energy process of gradual accumulation from the time domain to frequency domain; (c2) are its Var (X, α) and α Trends and the cubic spline fitting. Speech energy is gathered well in the field of FRFT, noise energy accumulation of FRFT domain is bad.

3.2. FRFT Domain Filter Design

Because speech energy is gathered well in the FRFT domain, noise FRFT domain energy aggregation is poor, then FRFT domain magnitude is higher, and noise margin is low, the amplitude is used to achieve better efficiency, how to select cutting threshold?

First n_0 frames is set for the noise frame, The mean amplitude of these n_0 frames in FRFT domain were MV_i,i=1,2,…, n_0。

$$MV = \frac{\sum_{i=1}^{n_0} MV_i}{n_0} \tag{6}$$

The current frame FRFT Amplitude:

$$p_\alpha(k) = | X_\alpha(k) | \qquad k = 1,2,\ldots,N$$

Threshold:

$$T = \max\{median(p_\alpha(k), k = 1,2,\ldots,N), a * MV\}, a > 1 \tag{7}$$

Filters:

$$H(k) = sign(\max\{p_\alpha(k) - T, 0\}) \quad Sign\ function \tag{8}$$

Filtered signal recovery:

$$\hat{x}(n) = F^{-\alpha}\{HF^{\alpha}[x(n')](k)\}(n) \tag{9}$$

4. Experimental Evaluation

Background noise is selected from Noisex-92 database [21], we test FRFT filter by TIMIT speech database standard, the sampling frequency fs = 16kHz, speech file KDT_003.WAV is accessed in library, and its wave is in (a) of Figure 2. In the course of the speech sub-frame, to take 32ms in frame length, ie, frame length M = [0.32fs] points.

Objectively, the performance of the algorithm is comprehensively analyzed from several aspects of the speech waveform, spectrogram, SNR improvement. Effect of denoising algorithm is analyzed quantitatively by using SNR.

$$SNR = 10 \log_{10}\left(\frac{\sum_{t=1}^{N} signal^2(t)}{\sum_{t=1}^{N} noise^2(t)}\right) \tag{10}$$

Experiment 1: The original voice is in Figure 2(a), the original speech were mixed with white noise (white), pink noise (pink), fighters (f16_cockpit) noise, plant (factory) noise, noisy vocal (babble) source from noise Noisex-92 libraries, the results were compared before and after this article FRFT method, their speech waveforms are shown in Figure 2, speech and the noisy speech are in left part of Figure 2, the filtered speech are in the right, the horizontal axis is the time (seconds) in each thumbnail, the ordinate is the amplitude; (a), (a1) are compared

before and after the filtering of the original speech; (b), (b1) are comparative voice filtering before and after the mixing of white noise (white) ; (c), (c1) are contrast before and after the filtering of the mixed pink noise (pink) voice; (d), (d1) are comparing before and after the noisy speech filtering of the mixed aircraft noise voice (f16_cockpit), (e), (e1) are comparison before and after voice filter when it is mixed with varying noise sources - plant noise (factory), (f), (f1) are comparison before and after voice filtering when it is mixed with varying noise sources - loud voices (babble). White noise (white), pink noise (pink), fighters (f16_cockpit) noise are stationary noise sources, while the factory (factory) noise, loud voices (babble) are non-stationary noise sources. The right in Figure 2 is the corresponding dynamic fractional α's changes.

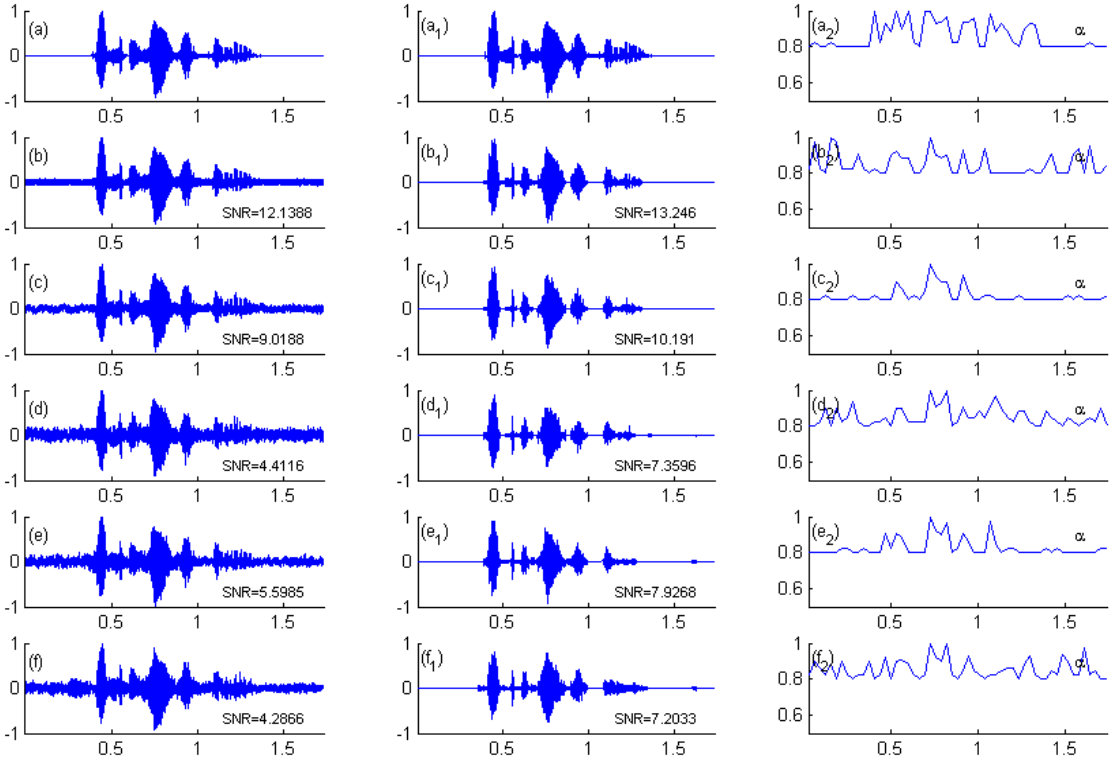

Figure 2. Speech waveform comparison before and after FRFT filtering

Spectrogram comparison results are in Figure 3 before and after filtering of this algorithm, the horizontal axis of each small graph is time (seconds), the vertical axis is frequency (kHz); (a) is the original voice spectrogram, (b), (b1) are the speech spectrogram comparison for mixing white noise (white) before and after filtering; (c), (c1) are the spectrogram contrast of speech which is mixed with pink noise (pink) before and after filtering; (d), (d1) are the spectrogram contrast of the the mixed aircraft noise (f16_cockpit) before and after filtering; (e), (e1) are speech spectrogram contrast when it is mixed with varying noise sources - plant (factory) before and after filtering; (f), (f1) are the speech spectrogram contrast when it is mixed with varying noise sources - noisy voices (babble) before and after filtering , time-varying noise sources - loud voices (babble) is mixed in voice frequency band, the general approach is hard work, good results are also reached in this algorithm.

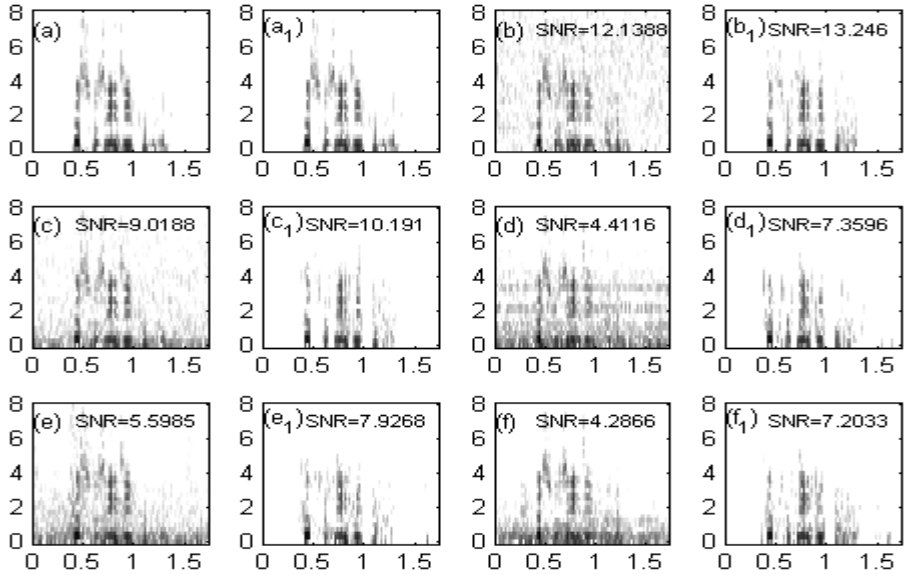

Figure 3. spectrogram comparison before and after FRFT filtering

Signal to noise ratio SNRin is calculated before filtering, the filtered signal to noise ratio is SNRout, speech is mixed respectively with white noise (white), pink noise (pink), fighter (f16_cockpit), factory noise (factory), noisy voices (babble), that five signal to noise ratio after these noisy speech is filtered by the algorithm: $\frac{SNR_{out}-SNR_{in}}{SNR_{in}}\times 100\%$ is Increased by 8.36%, 11.50%, 40.06%, 29.37%, 40.49% (see Table 1).

Tab.1 The results of speech whit four different kinds of noise

Compare items	white	pink	f16	factory	babble
SNR_{in}/dB	12.1388	9.0188	4.4116	5.5985	4.2866
SNR_{out}/dB	13.2460	10.1910	7.3596	7.9268	7.2033
$(SNR_{out}-SNR_{in})/SNR_{in}$(%)	8.36%	11.50%	40.06%	29.37%	40.49%

Experiment 2: the above speech signals were added to 4 groups of Gaussian noise in order to enhance the intensity, the input SNR of the resulting mixed signal SNRin is respectively: -4.556, -9.019, -18.41, -28.63dB. In experiment, 4 kinds of different SNRin speech signal are used under the mean filtering (n = 3,5), wavelet filter (db2 wavelet decomposition level n = 3) and FRFT domain filtering denoising SNR. The results are shown in Table 2. It is seen from the table, in strong background noise, de-noising method based on FRFT domain filtering is superior to conventional de-noising methods.

Tab.2 The results of speech whit four different kinds of SNR_{in}

SNR_{in}/dB	SNR_{out}/dB			
	Mean filter		Wavelet	FRFTFRFT
	n=3	n=5		
-4.556	-4.764	-6.074	-1.761	4.781
-9.019	-6.966	-7.335	-2.662	3.446
-18.41	-13.18	-11.56	-5.836	-1.027
-28.63	-21.74	-18.43	-11.67	-8.591
Complexity	Low	Low	Middle	High

In order to facilitate visual comparison, the denoising results in Table 2 is plotted, the results are shown in Figure 4. Among them: after speech signal is added by strong Gaussian noise, SNRin and SNRout SNR are compared in three kinds of de-noising method in Figure 4.

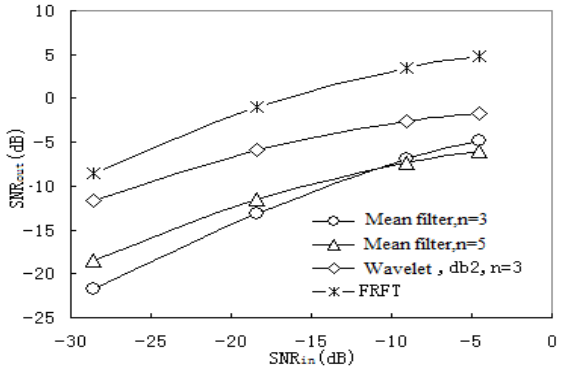

Figure 4. SNRout speech signal comparison in three kinds of de-noising method

From Figure 4, when the background noise is strong, denoising effect is based on FRFT domain filtering better than average filtering and wavelet de-noising, and de-noising effect is little changed with the enhancement of noise, while in the mean filter and wavelet de-noising methods, the de-noising effect are decreased rapidly with noise enhancement.

5. Conclusions and Outlook

In this paper, through the fractional Fourier transform application of noisy signals, signal measured vergence and its de-noising method are proposed based on fractional Fourier transform. Compared with the traditional Fourier transform de-noising method, fractional Fourier transform domain is made to the signal and noise by the proposed method, so that the signal and noise do not overlap as much as possible, so as to achieve better denoising effect. The results show that, for noisy signal with different SNR, there is a best Fractional order in the proposed method, de-noising effects make the best.

Because there are a variety of practical problems of non-Gaussian noise and strong background noise, it is difficult to extract the sound signal, a FRFT domain filtering methods is proposed in this paper. Experiments are done with standard TIMIT speech database and Noisex-92 noise library, the experimental results show that the use of this method all have good local features in the time domain and frequency domain, and it is superior to the traditional extraction method of sound signal feature. Meanwhile, denoising simulations are made with noisy speech which it is containing white noise (white), pink noise (pink), fighters (f16_cockpit) noise, factory noise (factory), noisy voices (babble), and has a strong voice Gaussian background noise, the simulation results show that this method significantly improves the signal to noise ratio, and it is significantly better than the traditional mean filtering and wavelet de-noising methods.

For non-stationary noise environments, speech denoising algorithm is proposed from the perspective of noise FRFT domain filtering. Non-stationary noise frame is smoothed and updated by the algorithm of fast tracking noise, it is better to estimate the ambient noise. Experiments show that the proposed algorithm can more effectively suppress background noise and improve voice quality after denoised. This method has simple calculation, real-time is high, noise immunity is strong, and it provides a new way to detect a weak signal de-noising and strong background noise. For this article filter on FRFT domain, the further study will be done on the narrowband filtering of the major energy aggregation points.

References

[1] SF Boll. Suppression of acoustic noise in speech using spectral subtraction*. IEEE Trans. ASSP.* 1979; 27(2): I13-120.
[2] M Berouti, R Schwartz, J Makhoul. *Enhancement of Speech Corrupted by Acoustic Noise*. Proceeding of 1979 IEEE, ICASSP. 1979; 208-211.
[3] Y Epharim, D Malah. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator*. IEEE. Trans. Acoustic, Speech Signal Processing*. 1984; 32(6): 1109-1121.
[4] P Lochwood, J Boundy. Experiments with a Nonlinear Spectral Subtractor (NSS), Hidden Markov Models and Projection, for Robust Recongnition in Cars. *Speech Commun*., 1992; 11(6): 215-228.
[5] Y Ephraim. A minimum mean square error approach for speech enhancement. *Acoustics, Speech, and Signal Processing*. 1990; 2: 829 -832.
[6] Liu Zhibin, Xu Naiping. Speech enhancement based on minimum mean-square error short-time spectra*l estimation and its realization.* IEEE International conference on intelligent processing system. 1997; 1794-1797.
[7] R Martin. *Speech enhancement using MMSE short time spectral estimation with Gamma distributed speech priors*. Proc. IEEE Int. conf. Acoustics, Speech, Signal Processing. 2002; 1: 253-256.
[8] S Kamath, P Loizou. *A multi-band Spectral Subtraction Method for Enhancing Speech Corrupted by Colored Noi*se. Proceedings of ICASSP. Orlando USA, IV-4164. 2002.
[9] JS Lim, AV Oppenheim. *Enhancement and Bandwidth Compression of Noisy Speech.* Proc.of the IEEE. 1979; 67(12): 1586-1604.
[10] JD Gibson, B Koo, SD Gray. Filtering of Colored Noise for Speech Enhancement and Coding. *IEEE Trans. Signal Processing*. 1991; 39: 1732-1742.
[11] WR Wu and PC Chen. Subband Kalman Filtering for Speech Enhancement. *IEEE Trans. On Circuits And Systemsli: Analog And Digital Signal Processing*. 1998; 45: 1072-1083.
[12] S Gannot, D Burshtein, E Weinstein. Iterative and sequential Kalman filter-based speech enhancement algorithms. *IEEE Trans Speech and Audio Process.* 1998; 6(4): 373-38.
[13] Y Ephraim, HLV Trees. A signal subspace approach for speech enhancement. *IEEE Transactions on Speech and Audio Processing.* 1995; 3(4): 251-266.
[14] U Mrital, N Phamdon. Signal ⁄ noise KLT based approach for enhancing speech degraded by colored noise. *IEEE Trans on Speech and Audio Processing*. 2000; 8(3): 159-167.
[15] A Rezayee, S Gazor. An adaptive KLT approach for speech enhancement. *IEEE Tram Speech Audio Processing*. 2001; 9(2): 87-95.
[16] Y Hu, P Loizou. A generalized subspace approach for enhancing speech corrupted by colored noise. *IEEE Trans on Speech and Audio Processing*. 2004; 11(4): 334-341.
[17] H Leva, Y Ephraim. Extension of the signal subspace speech enhancement approach to colored noise. *IEEE Signal Processing*. 2003; 10(4): 104-106.
[18] Soo-Chang Pei, Jian-Jiun Ding .Relations between Fractional Operations and Time-Frequency Distributions, and Their Applications, *IEEE Transactions On Signal Processing*. 2001; 49(8): 1638-1655.
[19] Tao Ran, Bing Deng, etc. fractional Fourier transform in signal processing research. *Science in China Series F.* 2006; 49(1): 1-25
[20] Tao Ran, Bing Deng, Wang Yue. Fractional Fourier transform and its application. Beijing: Tsinghua University Press. China. 2009.
[21] Spib Noise data [EB/OL], http://spib.rice.edu/spib/select_noise.html