Speech Enhancement based on Fractional Fourier transform

JINGFANG WANG
School of Information Science and Engineering
Hunan International Economics University
Changsha, China, postcode:410205
e-mail: 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.

Key-Words: - Acoustic signal, Fractional Fourier Transform(FRFT), Speech Enhancemen, de-noising, auto-adaptive processing, Dynamic filtering

1.INTRODUCTION

With the development of communication technology, voice communication has become a major communication medium for people to transmit information more convenient. However, the widespread nature of noise makes the voice communication quality has declined. Therefore, to reduce the noise on the performance of voice communications, improve the quality of voice communications, voice denoising for technology has become a hot research topic. S. Boll [1] in 1979 presented the classic spectral subtraction algorithm (Spectral Subtraction, SS). The algorithm assumes that short-term stationary additive noise and speech signal independent of the conditions, through the spectrum from the noisy speech signal by subtracting the estimated noise spectrum, resulting in denoised speech signal spectrum. But because of its assumption of local stability is not consistent with the actual situation, so the results are unsatisfactory, leading to larger residual musical noise and other issues. So Berouti [2] in the traditional spectral subtraction based on the increase of the size of the adjustment coefficient of the noise power spectrum and the enhanced speech power spectrum to increase the minimum limit the performance of spectral subtraction. However, due to its correction factor and the minimum value is determined based on experience, poor adaptability of the method. Y. Ephraim 1984 [3] introduced the minimum mean square error to the spectral subtraction, can be part of the solution to the music noise and improve the denoising results. But the algorithm requires prior estimates because the distribution of speech spectrum, and thus larger than the calculation. P. Lochwood et al [4] on the basis of spectral subtraction, SNR of speech signals based on adaptive speech enhancement gain function, nonlinear spectral subtraction algorithm is proposed (Nonlinear Spectral Subtractor, NSS), although the algorithm to improve the voice signal to noise ratio, but the audio quality has not improved. Finally, in order to further reduce the musical noise, improving voice clarity, people continue to put forward a variety based on the traditional spectral subtraction improved algorithm [5-7], better voice quality improved. However, when the signal to noise ratio in low or non-stationary noise, the performance of the traditional spectral subtraction tends to become poor. To this end, 2002 S. Kamath et al [8] based on iterative multi-band spectrum subtraction method. The method takes into account colored noise on the speech spectrum of the inhomogeneity, the introduction of spin sub-band processing factors, while maintaining high voice quality at the same time, can effectively eliminate the

noise pollution under the colored background noise and music noise. Also based on speech production model maximum a posteriori estimation [9], Kalman filters [10-12], it is the voice of the generation process can be equivalent to a linear time-varying filters for different types of voice using different excitation sources.

1995 Y. Ephraim et al [13] constructed the first time in the time domain for a new speech enhancement approach (sub-space frame theory), a signal subspace based speech enhancement algorithm. The basic idea is to noisy speech signal space by some method into two orthogonal subspaces, one is the voice signal plus noise subspace, also called the signal subspace, because in this sub-space is primarily based on the signal based; the other is the noise subspace, the noise subspace contains only noise components. Therefore, the estimated clean speech can be a signal in the noise subspace removed, leaving only the signal subspace of the signal. After removing the noise subspace in the signal plus noise subspace to estimate the voice signal filtering. Y. Ephraim, however initial work mainly for white noise, in order to deal with non-white noise, 2000 Mrital [14] proposed a signal / noise KL transform, although the enhanced signals after each frame has a smaller residual noise, However, the non-frames between the stability of the residual noise disturbing. 2001 A. Rezaye and S. Gazor [15] proposed an adaptive KLT approach for handling non-stationary noise, they assumed that the feature vector for the speech signal can be approximated by the non-stationary colored noise covariance matrix diagonalization. But this is a sub-optimal approach, it does not exist fast algorithm, which is the inadequacy of the method. Therefore, 2003 Y. Hu et al [16] method for Rezayee deficiencies in the signal subspace decomposition is proposed based on in the time domain and frequency domain speech enhancement for colored noise algorithm. In the same year A. Lev and Y. Ephraim [17] have also proposed approach for colored noise. However, the premise of the above methods are required noise covariance matrix must be full rank, which is not applicable for narrow-band noise. Around the voice signal edge enhancement fought study the scientific and technological work, we study the FRFT (Fractional Fourier Transform) filter, and after a good variety of noise test results, the proposed algorithm to calculate the cost of a small, simple and easy to implement.

2.FRACTIONAL FOURIER TRANSFORM

Fractional Fourier Transform (Fractional Fourier Transform, FRFT) is developed in recent years, a new time-frequency analysis tool, which is a generalization of Fourier transform. In essence, the signal in the fractional Fourier domain representation, while integration of the signal in time domain and frequency domain information. This new mathematical tools not only closely linked with the Fourier transform, but also with other time-frequency analysis tool is also very meaningful connection, has been widely used in optical system analysis, filter design, signal analysis, solving differential equations, phase retrieval , pattern recognition [18-20]. In recent years the application of fractional Fourier transform, most studies focused on estimation of linear FM signals, detection and filtering aspects.

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

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

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

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

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

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

3.DYNAMIC FILTERING FRFT

Because the location of voice communication has a very large variability, the inevitable will be from the surrounding environment, and even the interference of the other speaker, due to the existence of various kinds of interference, making the call performance greatly reduced, in order to improve call quality, Denoising have been proposed for speech enhancement algorithm, to extract the original voice pure as possible. Fractional Fourier transform of the energy accumulation and transformation of order α is related to its aggregation strongly depends on the extent of its close to Fourier transform; fractional Fourier transform in the speech signal are voiced and unvoiced focus some energy Nature, the energy difference is the different focus areas: energy focused on the voiced transform domain in the fractional reflected in the central region of the waveform, the energy of the voiceless focus is reflected both ends of the waveform area. Fractional Fourier transform white noise did not focus on the nature of the energy,

noise energy is focused on the poor, can be used for speech signal denoising.

2.1 The best fractional order α FRFT

Different segments of the signal and noise pollution levels give different α's FRFT fractional transformation, then the effective filtering. So by what measure is the common MMSE (minimum mean square error, MMSE), where we focus the energy weighted measure of the degree of variance.

2N point signal Fourier transform fractional α order total is: $X_{\hat{\alpha}}(k), k = 1,2,\ldots,2N$, Symmetric about the center because it is taking half. Probability normalized it:

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

$$EX = \sum_{k=1}^{N} k p_\alpha(k),$$

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

α_i between taking a weighted variance Var (X, α_i), then this set of data for cubic spline fitting, and then seek the Var (X, α) α_0 is the minimum value of the corresponding fractional for the best .

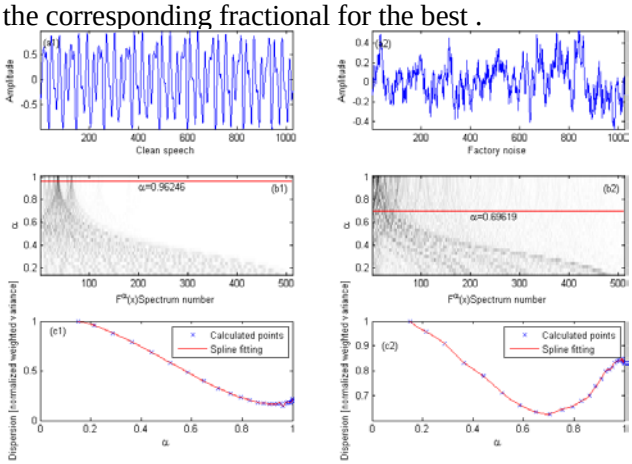

Figure 1. weighted speech and noise, the fractional variance α to find the best contrast FRFT

Figure 1 (a1) to take a voice signal, (b1) the fraction of α-order Fourier transform, from the time domain to frequency domain changes in the energy process of gradual accumulation; (c1) by type (5) Var (X, α) and α in Trends and the cubic spline fitting. (a2) to take a factory noise, (b2) corresponding to the α-order fractional Fourier transform, from the time domain to frequency domain changes in the energy process of gradual accumulation; (c2) Var (X, α) and α and cubic trends Article fitting. Voice in the field of energy gathered FRFT good, bad noise energy accumulation FRFT domain.

2.2 FRFT domain filter design

Voice in the field because of the energy gathered FRFT good, bad noise energy accumulation FRFT domain; then FRFT Amplitude high, the noise margin is low, the use rate cuts to achieve a better rate of wave, the question how to select the cutting threshold.

n_0 frame before setting the noise frame, the frame before the frame n_0 average rate FRFT domain were MV_i,i=1,2,…, $n_{0\circ}$

$$MV = \frac{\sum_{i=1}^{n_0} MV_i}{n_0} \quad (6)$$

The current frame FRFT Amplitude:

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

Threshold:

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

Filters:

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

Filtered signal recovery:

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

4. EXPERIMENTAL EVALUATION

Background noise taken from Noisex-92 database [21], following the standard TIMIT speech database we tested, the sampling frequency fs = 16kHz, access to library KDT_003.WAV in Figure 2 (a). In the speech sub-frame process, the frame size to take 32ms, the frame length M = [0.32fs] points.

Objectively, from the speech waveform, spectrogram, signal to noise ratio to improve several aspects of the performance of the algorithm a comprehensive analysis. With signal to noise ratio (10) to quantitatively analyze the effect of noise reduction algorithm..

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

Experiment 1, the original voice Figure 2 (a), the original voice and noise Noisex-92 library, the different nature of the white noise (white), pink noise (pink), fighter (f16_cockpit) noise, factory (factory) noise, noise Vocal (babble) sources were mixed, with this iterative Wiener filter speech waveform before and after filtering the results of comparison in Figure 1, left the Ministry of noisy speech, right is the filtered voice, every small diagram of the horizontal axis is time (Seconds), vertical axis is amplitude; (a), (a1) before and after filtering the original speech, (b), (b1) was mixed with white noise (white) speech before and after filtering, (c), (c1) is Mixed with pink

noise (pink) before and after voice filtering, (d), (d1) for the melee machine (f16_cockpit) before and after the noise audio filtering, (e), (e1) for the mixed time-varying noise sources - plant noise (factory) Speech before and after filtering, (f), (f1) for the mixed time-varying noise sources - loud voices (babble) speech before and after filtering. White noise (white), pink noise (pink), fighter (f16_cockpit) noise is stationary noise sources, plant (factory) noise, noisy voices (babble) non-stationary noise sources.

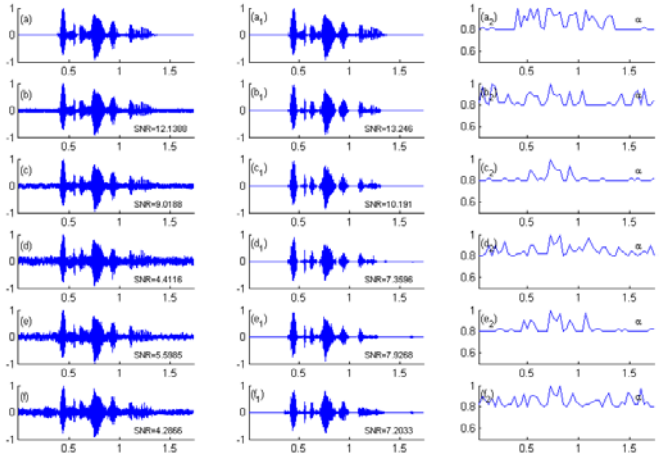

Figure 2. FRFT speech waveform before and after filtering the results of comparison

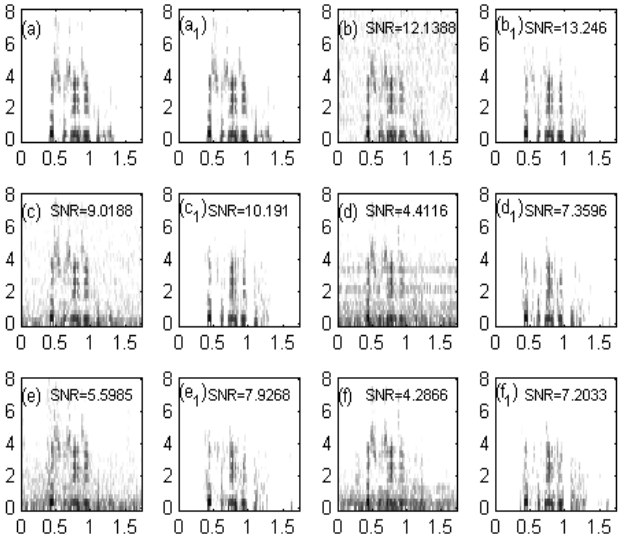

Figure 3. FRFT domain filtering the results compared before and after filtering spectrogram

Figure 3 Spectrogram before and after filtering algorithm comparing the results of a small map of each horizontal axis is time (seconds), vertical axis is frequency (kHz); (a) the original speech spectrogram, (b), (b1) is Mixed with white noise (white) compared speech spectrogram before and after filtering, (c), (c1) was mixed with pink noise (pink) compared speech spectrogram before and after filtering, (d), (d1) for the melee machine (f16_cockpit) Voice spectrogram before and after filtering noise contrast, (e), (e1) is mixed with time-varying noise source - the factory (factory) compared speech spectrogram before and after filtering,, (f), (f1) for the mixed time-varying noise sources - Loud voices (babble) compared speech spectrogram before and after filtering, time-varying noise sources - loud voices (babble) mixed in the voice frequency band, the general method of hard work, the algorithm reached the same good results. The right subscript 2 is the corresponding graph α of fractional dynamics.

Calculated signal to noise ratio before filtering SNRin, filtered signal to noise ratio SNRout, white noise in mixed (white), pink noise (pink), fighter (f16_cockpit), factory noise (factory), noisy voices (babble) that Noise Filtering five signal to noise ratio of 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 of these 4 groups were added to the speech signal intensity were enhanced Gaussian noise from the input SNR SNRin mixed-signal, respectively: -4.556, -9.019, -18.41,-28.63dB. 4 different experiments, the speech signal under SNRin mean filter (n = 3,5), wavelet filter (db2 wavelet, decomposition level n = 3) and FRFT filtering denoising SNR. The results shown in Table 2. Can be seen from the table, in strong background noise, filtering based denoising FRFT domain is superior to the traditional denoising 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 direct comparison, the results of Table 2 plotted in the denoising curve.

Figure 3 shows, when the background noise, strong and, based FRFT domain filtering is better than the average noise reduction filtering and denoising, and it is enhanced with the denoising effect of noise was little changed, while the average filter and wavelet to Noise Noise is enhanced with the effect of noise decreased rapidly.

5. CONCLUDING REMARKS

In this paper, the noisy signal by using the fractional Fourier transform, Fourier transform based on fractional degree of the signal and everything is measured and denoising. Fourier transform traditional denoising method, compared the proposed method to the signal and noise fractional Fourier transform domain, 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 different SNR with noise signals, the proposed method are an optimal fractional order, can make the best denoising effect.

In this paper, a variety of non-Gaussian noise and strong background noise, sound signals difficult to extract the real problem, a filtering method using a FRFT. Standard with the TIMIT speech database and Noisex-92 noise database with the experimental results show that using this method in both time domain and frequency domain characteristics of a good local and superior to the traditional voice signal feature extraction methods. At the same time, through with white noise (white), pink noise (pink), fighter (f16_cockpit) noise, factory noise (factory), loud voices (babble), and have a strong voice denoising Gaussian background noise simulation experiments The results show that this method significantly improves the signal to noise ratio, and significantly better than the traditional mean filtering and wavelet denoising.

For non-stationary noise, from the perspective of noise filtering FRFT speech denoising algorithm is proposed. Fast tracking algorithm using non-stationary noise, noise smoothing frame updated to better estimate environmental noise; experiments show that the proposed algorithm can effectively suppress background noise and improve voice quality after denoising. This method has simple, real-time high, and strong anti-noise characteristics, and background noise for the strong and weak signal detection denoising provides a new way. This filter can also FRFT energy domain the main gathering point for further study of narrow-band filter.

REFERENCES

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction"[J]. *IEEE Trans. ASSP*, vol.27, No.2, pp. l13-120, 1979.

[2] M. Berouti, R. Schwartz, J. Makhoul, "Enhancement of Speech Corrupted by Acoustic Noise"[C]. *Proceeding of 1979 IEEE*, ICASSP, pp. 208-211,1979.

[3] Y. Epharim, D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator" [J]. *IEEE. Trans. Acoustic, Speech Signal Processing*, vol.32, No.6, pp. 1109-1121, 1984.

[4] P. Lochwood, J. Boundy, "Experiments with a Nonlinear Spectral Subtractor(NSS), Hidden Markov Models and Projection, for Robust Recongnition in Cars"[J]. *Speech Commun.*, vol.11, No.6, pp. 215-228, 1992.

[5] Y. Ephraim, "A minimum mean square error approach for speech enhancement"[J]. *Acoustics, Speech, and Signal Processing*, vo1.2, pp. 829 - 832, 1990.

[6] Liu Zhibin, Xu Naiping, "Speech enhancement based on minimum mean-square error short-time spectral estimation and its realization"[C]. *IEEE International conference on intelligent processing system*, pp. 1794-1797, Oct. 1997.

[7] R. Martin, "Speech enhancement using MMSE short time spectral estimation with Gamma distributed speech priors"[J]. in *Proc.IEEE Int.conf.Acoustics ,Speech,Signal Processing*, vol.1, pp. 253-256, 2002.

[8] S. Kamath, P. Loizou, "A multi-band Spectral Subtraction Method for Enhancing Speech Corrupted by Colored Noise"[C]. *Proceedings of ICASSP*[C]. Orlando USA, IV-4164, 2002.

[9] J. S. Lim and A. V. Oppenheim. "Enhancement and Bandwidth Compression of Noisy Speech"[J]. *Proc.of the IEEE*, vo1.67, No.12, pp. 1586-1604, Dec.1979.

[10] J. D. Gibson, B. Koo, and S. D. Gray, "Filtering of Colored Noise for Speech Enhancement and Coding" *IEEE Trans. Signal Processing*, vol.39, pp. 1732-1742, Aug. 1991.

[11] W. R. Wu and P. C. Chen, "Subband Kalman Filtering for Speech Enhancement"[J]. *IEEE Trans. On Circuits And Systemsli:Analog And Digital Signal Processing*, vol.45, pp. 1072-1083, Aug. 1998.

[12] S. Gannot, D. Burshtein, E. Weinstein, "Iterative and sequential Kalman filter-based speech enhancement algorithms"[J]. *IEEE Trans Speech and Audio Process*, vol.6,No.4, pp. 373-385, 1998.

[13] Y. Ephraim, H. L. V. Trees, "A signal subspace approach for speech enhancement"[J]. *IEEE Transactions on Speech and Audio Processing*, vol.3, No.4, pp. 251-266, 1995.

[14] U. Mrital, N. Phamdon, "Signal / noise KLT based approach for enhancing speech degraded by colored noise"[J] . *IEEE Trans on Speech and Audio Processing*, vol.8, No.3, pp. 159-167, 2000.

[15] A. Rezayee, S. Gazor, "An adaptive KLT approach for speech enhancement"[J]. *IEEE Tram Speech Audio Processing*, vol.9, No.2, pp. 87-95, 2001.

[16] Y. Hu, P. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise"[J]. *IEEE Trans on Speech and Audio Processing*, vol.11, No.4, pp. 334-341, 2004.

[17] H. Leva, Y. Ephraim, "Extension of the signal subspace speech enhancement approach to colored noise"[J]. *IEEE Signal Processing*, vol.10, No.4, pp. 104-106, 2003.

[18] Soo-Chang Pei, Jian-Jiun Ding .Relations between Fractional Operations and Time-Frequency Distributions, and Their Applications[J], IEEE TRANSACTIONS ON SIGNAL PROCESSING, vol. 49, No. 8, 1638-1655 , 2001 .

[19] Tao Ran, Bing Deng, etc. fractional Fourier transform in signal processing research [J]. Science in China Series F, 2006,49 (1) :1-25

[20] Tao Ran, Bing Deng, Wang Fractional Fourier transform and its application [M]. Beijing: Tsinghua University Press, 2009.

[21] Spib Noise data[EB/OL], http://spib.rice.edu/spib/select_noise.html.