Threshold Filtering for Phoneme Pronunciation Signals Based on FrFT

Zhenyan Fan, Jun Yu, Zhongxiao Li, Xiaodong Zhuang*
Electronic Information College
Qingdao University
Qingdao, China
zhenyan_fan@163.com

Nikos E. Mastorakis
English Language Faculty of Engineering
Technical University of Sofia
Sofia, Bulgaria
mastor@tu-sofia.bg

***Abstract*—The Fractional Fourier Transform (FrFT) is applied to the denoising of noisy speech. The optimal transform order of FrFT for single phoneme is determined by using weighted variance method. Then the soft-hard threshold compromise denoising algorithm is put forward. This method removes the amplitude of noise from noisy phoneme signals in FrFT-domain. Signals are reconstructed by inverse FrFT to get original speech. The experimental results show that this method can effectively remove noise from signals and get a good auditory effect, and this algorithm is of low computational complexity.**

***Keywords*—FrFT, Weighted Variance, Threshold Filtering, SNR**

I. INTRODUCTION

With the development of communication technology, voice communication has become a major medium for people to communicate. However, the noise in nature has led to a decline in the quality of voice communication. Therefore, in order to reduce the impact of noise on voice communication performance and improve voice communication, denoising technology for speech has always been a hot research topic[1]. In recent years, FrFT is widely used to denoising and become a practical and useful method for filtering[2].

In FrFT the signal in time and frequency domain represents more clearly as overlapping is very small and more suitable for non stationary type of signal. The performance is estimated for various values of fractional order (α) and then the results are compared with the standard techniques by accessing different definitions of FrFT. Speech enhancement has been done by using spectral subtraction method and de-noises the speech by conventional method . It showed that these methods are unable and inadaptable in the intense noise environment. Therefore the method for speech enhancement based on FrFT filtering is suggested. The method has been implemented by putting the noising speech in disperse degree of FrFT[3]. In this paper, FrFT is applied to threshold filtering for noisy phoneme signal and it gets a good effect.

II. FRACTIONAL FOURIER TRANSFORM

The FrFT is a time-frequency analysis method[1]. As a generalized form of Fourier transform, the FrFT of signals can be explained as a counterclockwise rotation of the signal axes in the time-frequency plane around the origin. When the rotation angle is π/2, it turns to be Fourier transform. Define x(t) as a square integrable function in time domain of t , then its FrFT in p order is a near integral operation[4-8]:

$$X_p(u)=\int_{-\infty}^{\infty}K_p(t,u)x(t)dt \quad .1$$

Among them, p refers to the FrFT order. $K_p(t,u)$ is called the kernel function of FrFT, its expression is:

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

Plug $\alpha=p\pi/2$ into the formula, so

$$X_p(u)=\int_{-\infty}^{\infty}K_p(t,u)x(t)dt$$
$$=\begin{cases}\sqrt{\dfrac{1-j\cot\alpha}{2\pi}}e^{j\frac{u^2}{2}\cot\alpha}\displaystyle\int_{-\infty}^{\infty}x(t)e^{j\frac{t^2}{2}\cot\alpha}e^{-jut\csc\alpha}dt & \alpha\neq n\pi\\ \delta(t-u) & \alpha=2n\pi\\ \delta(t+u) & \alpha=(2n+1)\pi\end{cases} \quad ..$$

Where ,$\alpha=p\pi/2$ is the angle of time axis and u axis. When $\alpha=\pi/2$, F_p is DFT. p and α is in the period of 4 and 2π respectively, so $p\in(-2,2]$ and $\alpha\in(-\pi,\pi]$, are enough[9]. Energy aggregation degree of FrFT for signals is different with the change of the order *p*. So it is necessary to determine an optimal order first.

III. THRESHOLD FILTERING OF PRONUNCIATION SIGNAL BY FRFT

A. Determination of Optimal Order

Energy aggregation degree of FrFT is related to the order *p*. The energy aggregation of the voiced pronunciation is reflected in the central region of the waveform on the fractional transform domain. The energy aggregation of the devoiced pronunciation is reflected in the two ends of the waveform. Fractional Fourier transform has no energy aggregation properties for noises, and it has poor aggregation for noise, so FrFT can be used for speech signal denoising[8].

Different fractional order p of FrFT is given for different signal segments and noise pollution degree. Weighted variance method is proposed to measure the energy aggregation degree.

* Xiaodong Zhuang is the corresponding author. This work is supported by the National Science Foundation of China (41804110).

$2N$ points FrFT, whose order is p, of signal x(t) is $X_p(k), k=1,2,\cdots,2N$. Because it is symmetry about the center point, its probability normalization is showed as (4).

$$P_p(k)=\frac{|X_p(k)|}{\sum_{i=1}^{N}|X_p(i)|}$$

Where k=1,2,...,N, and $|\bullet|$ is absolute value.

Mean is defined as (5).

$$EX=\sum_{k=1}^{N}kP_p(k)$$

Weighted variance is showed as (6).

$$Var(X,p)=\sum_{k=1}^{N}(k-EX)^2P_p(k)$$

Then $p_i(0<p_i<1)$ is chosen to compute weighted variance $Var(X,p_i)$, and the p_0 responding to the minimum value of $Var(X,p)$ is the optimal order[1].

B. Soft-Hard Threshold Compromise Denoising Algorithm

The energy aggregation degree of speech is better than that of noise and the amplitude of speech is larger than the amplitude of noise in FrFT-domain[1]. In view of the disadvantages of soft threshold and hard threshold, the soft-hard threshold compromise method is put forward to denoising.

The definite of Soft-Hard Threshold Compromise Denoising Algorithm is[10]

$$\hat{X}_p(k)=\begin{cases} sign(X_p(k))\bullet(|X_p(k)|-\alpha\lambda), & |X_p(k)|\geq\lambda \\ 0, & |X_p(k)|<\lambda \end{cases},\quad 0\leq\alpha\leq 1$$

By reference to[11], for practical situations where the noise level is unknown, one can apply hard-soft threshold compromise denoising algorithm with threshold

$$\lambda=r*\sigma\sqrt{2\ln(n)/n}$$

where the variance estimate of noise is $\sigma=\mathrm{MAD}(X_p(k))/0.6745$, with MAD is the median absolute value of $X_p(k)$.

C. Measuring Parameters

SNR(signal to noise ratio) is defined as the ratio between the power of signal and the power of noise. It is usually expressed in logarithmic decibel scale[12]

$$SNR=10\log_{10}\frac{\sum_{i=1}^{N}x^2(i)}{\sum_{i=1}^{N}n^2(t)}$$

where $x(i)$ and $n(i)$ are the speech segments and noise segments. The larger the SNR is, the smaller the noise is.

IV. Experiments

Some single phonemes, such as /e/, /o/, /u/ and /v/, are applied into this experiments. In addition, some noises, such as white.wav and babble.wav, are obtained from noisex92 library. Their sample frequency is 8KHz. First the speech and noise are framed to calculate the weighted variance for determining the optimal order. The results are shown from Fig. 1. to Fig.6.

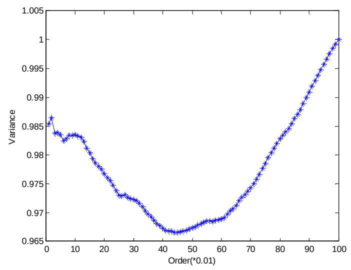

Fig.1. The Var(X, p) of pronunciation /e/

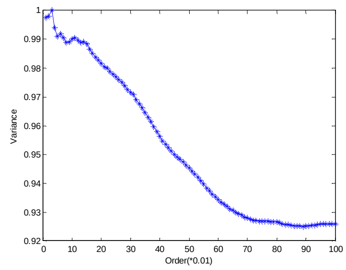

Fig.2. The Var(X, p) of pronunciation /o/

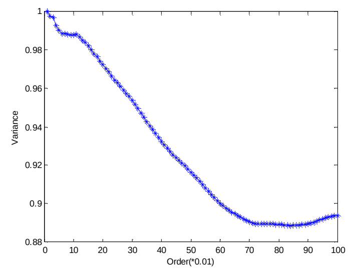

Fig.3. The Var(X, p) of pronunciation /u/

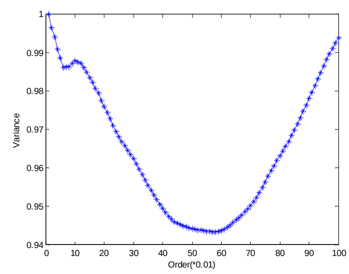

Fig.4. The Var(X, p) of pronunciation /v/

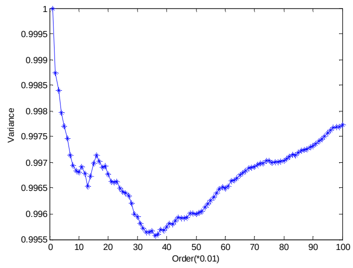

Fig.5. The Var(X, p) of white noise

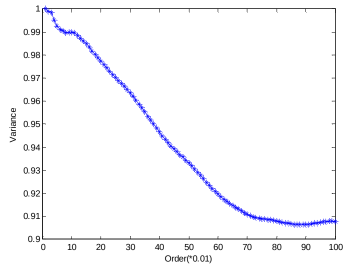

Fig.6. The Var(X, p) of babble noise

From Fig.1. to Fig.6., it can be seen that the minimum variance corresponds to the optimal order. The optimal orders of single phonemes and noises are represented in Table Ⅰ.

TABLE I. THE OPTIMAL ORDER OF SIGNALS

Signals	**Optimal order**
/e/	p=0.45
/o/	p=0.89
/u/	p=0.84
/v/	p=0.58
white noise	p=0.36
babble noise	p=0.89

Then the noises are added into every single phonemes speech to get noisy speeches. The noisy speeches are framed first. FrFT is used to every frame and the threshold λ is calculated by reference to (8). Then the threshold is applied on the spectrum received from FrFT to reduce the noise level in the noisy speeches. After thresholding, inverse fractional Fourier transform is applied to the get the denoised speech signal. The results are shown from Fig.7. to Fig.14. Denoising effect is evaluated by using SNR and the results are shown in Table Ⅱ.

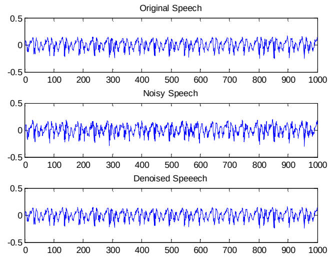

Fig.7. The denoising results of /e/ and white noise

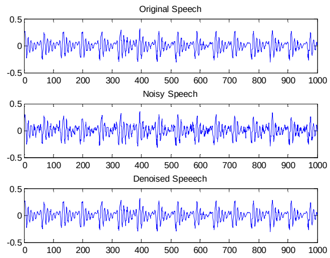

Fig.8. The denoising results of /o/ and white noise

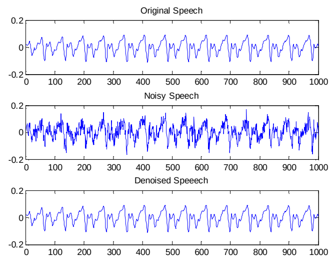

Fig.9. The denoising results of /u/ and white noise

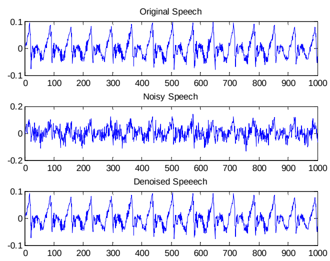

Fig.10. The denoising results of /v/ and white noise

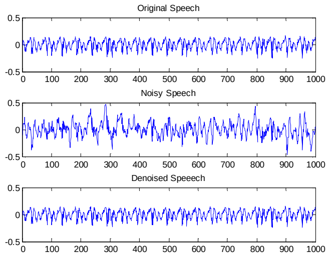

Fig.11. The denoising results of /e/ and babble noise

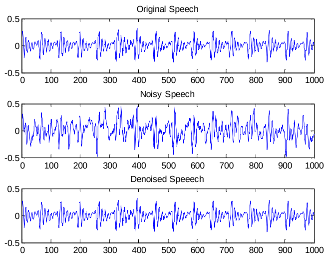

Fig.12. The denoising results of /o/ and babble noise

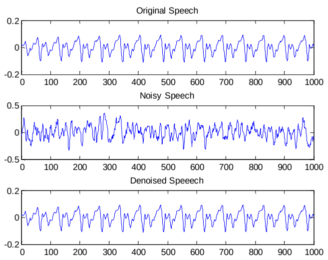

Fig.13. The denoising results of /u/ and babble noise

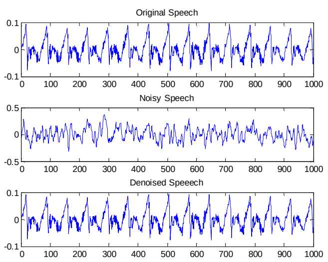

Fig.14. The denoising results of /v/ and babble noise

TABLE II. THE RESULTS OF SNR

Signals	SNR before filtering (dB)	SNR after filtering (dB)
/e/ and white noise	9.1832	28.8725
/o/ and white noise	8.8886	38.7629
/u/ and white noise	4.1753	49.2135
/v/ and white noise	-1.9438	29.9689
/e/ and babble noise	-2.9079	28.8725
/o/ and babble noise	-3.3351	38.7629
/u/ and babble noise	-7.6120	49.2135
/v/ and babble noise	-13.5863	29.9689

Form Fig.7. to Fig.14. , it can be obviously seen that the noise in noisy speech can be effectively removed in FrFT-domain by soft-hard threshold compromise denoising algorithm. And it also has a good auditory effect. In table 2, it displays that the values of SNR after filtering are larger than that before filtering. The results show that the thresholding filtering in FrFT-domian can effectively remove the noise from noisy speech.

V. CONCLUSION

The work is based on FrFT technique to remove the noise from pronunciation signal. Some single phonemes and noise are selected to filter in FrFT-domain. The optimal order p is firstly determined by using weighted variance. Then noisy signals are filtered by using soft-hard threshold compromise denoising algorithm in FrFT-domain. Through this way, the noise is removed effectively and get a good auditory effect. This method offers many advantages over the other denoising method due to less perturb to the original content of the signal. Hence other technique may be hybrid with this technique.

REFERENCES

[1] J. F. Wang, H. Y. Xu, “Speech enhancement based on FrFT filtering”, Computer Engineering and Applications, 48th ed., vol. 12. pp. 129-134, July 2012.

[2] Y. H. Bu, P. B. Wang, “Filtering Based on Fractional Fourier Transform”, Ship Electronic Engineering, 36th ed., vol. 4. pp. 38-40, June 2016.

[3] P. Kumar, S. Kansal, “Noise Removal in Speech signal using Fractional Fourier Transform”, IEEE International Conference on Information Communication, Instrumentation and Control (ICICIC-2017) Conference. India, August 2017.

[4] J. Xia , J. Y. Zhang, X.B. Li, “Weak signal detection based on FrFT-domain adaptive filtering algorithm”, Cie International Conference on Radar. IEEE. China, October 2017.

[5] H. Li, Y.L. Qin, Y.P. Li, H.Q. Wang, X. Li, “A FrFT Based Algorithm for Doppler Frequency Rate Estimation from LFM Coherent Pulse Train”, Journal of Electronics & Information Technology, 11th ed., vol. 32. pp.2718-2723, March 2011.

[6] X. Q. Zhang, Y.C. Yao, M. W. Qin, "Composite Chirp Signal Analysis based on FRFT", Communications Technology, 5th ed., pp.474-477, May 2014.

[7] H. Yin, C. Nadeu, V. Hohmann, "Pitch- and Formant-Based Order Adaptation of the Fractional Fourier Transform and Its Application to Speech Recognition", Eurasip Journal on Audio Speech & Music Processing, vol. 1, pp.14, November 2009.

[8] Z.T. Yuan, W.D. Hu, "Parameter estimation of LFM signals using FrFT interpolation", 11rd ed., vol. 25. pp.1726-1731, November 2009.

[9] N. Pang, P. J. Xu, "A new acceleration estimation method based on special FrFT in MIMO radar," IET International Radar Conference 2015, pp. 1-5, Hangzhou, October 2015.

[10] W. L. Zhao, "Speech Signal Denoising of Threshold Based on Wavelet Transformation", Chongqing University of Science & Technology (Natural Science Edition), 7th ed., vol. 4. pp. 73-75, December 2005.

[11] D. L. Donoho, "De-noising by soft-thresholding", IEEE Transactions on Information Theory, 41st ed., vol. 3. pp. 613-627, May 1995.

[12] G. Ling, T. Yamada, S. Makino, "Performance estimation of noisy speech recognition using spectral distortion and SNR of noise-reduced speech", Tencon IEEE Region 10 Conference. IEEE, China, pp. 1-4, January 2014.