IJCTA, 8(3), 2015, pp. 1063-1070

Speech Enhancement using Fractional Order Spectral Subtraction Method

A. Rutwik and B.T. Krishna

Abstract: Spectral subtraction method is proven to be one of the most efficient algorithm for the enhancement of speech. Fractional Fourier Transform (FRFT) is the generalization of the Fourier Transform. The discrete version of FRFT called as DFRFT is defined in several ways by different people. In this paper, a novel spectral subtraction method is proposed by replacing Fast Fourier Transform (FFT) with FRFT. The performance of the proposed system is evaluated in terms of Mean Squared Error(MSE) and Peak Signal to Noise Ratio (PSNR) for different values of fractional order, α and compared with the conventional techniques by using different definitions of DFRFT.

***Keywords**:* Speech enhancement; Spectral subtraction; Fractional order; Non Stationary; Eigen Value; Fourier Transform.

1. INTRODUCTION

Speech enhancement aims to improve the intelligibility and overall perceptual quality of degraded speech signal. There is wide variety of scenarios in which it is desired to enhance speech [10]. Spectral-subtractive algorithm [11] is used to find the estimate of clean signal by subtracting the noise from a noisy speech spectrum. It is considered as first one of its kind for the removal of unwanted signals.

FRFT is considered as fractional power of the classical Fourier transform. It provides an angular rotation in the frequency spectrum. Now a day, it is finding applications in the field of signal, image processing, control systems and other allied fields of engineering and science. The discrete version of FRFT is called as Discrete Fractional Fourier Transform (DFRFT) [2],[3]. There are several definitions of DFRFT exists in the literature [5]-[9]. Instead of FFT, a novel technique FRFT is replaced in this proposed spectral subtraction method [1],[4]. The performance of the novel system in terms of Peak Signal to Noise Ratio (PSNR) for different values of fractional order, α is studied and compared with the conventional technique. This process is repeated by considering different definitions of DFRFT.

This Paper is organized as follows. Spectral subtraction method explained in section 2. The novel technique fractional Fourier transform and various definitions of FRFT is presented in section 3. Proposed spectral subtraction method, derivation and its PSNR, MSE plots are described in Section 4. Experimental results are presented in section 5. Conclusion and future scope of the work is presented in section 6.

2. SPECTRAL SUBTRACTION METHOD

In this method clean signal is estimated by performing subtraction and Fourier transform operation. This algorithm is considered to be simplest one because of reduced computational complexity. The diagram of spectral subtraction method is as in Fig 1.

University College of Engineering Vizianagaram, JNTUK, Vizianagaram, Andhrapradesh, India-535003
e-mail: rutwik777@gmail.com (and) tkbattula@gmail.com

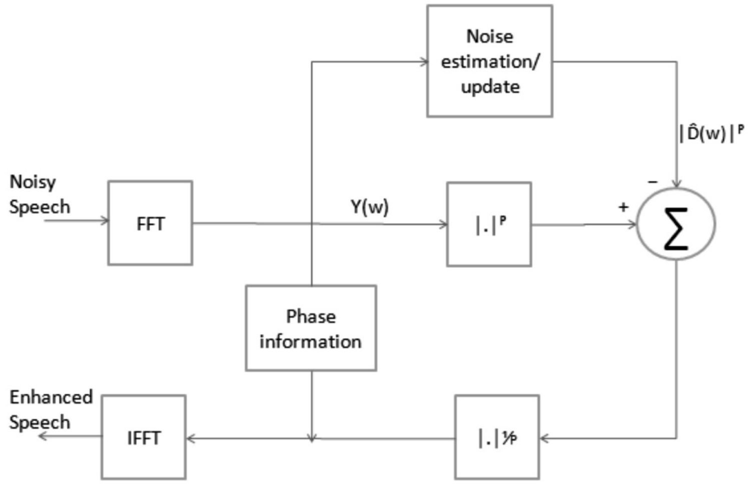

Figure 1: Spectral Subtraction method

The input to the system is a corrupted signal y(n)y(n), defined as

$$y(n) = x(n) + d(n) \tag{1}$$

where $x(n)$ is desired signal and $d(n)$ is noise. The above equation can be written in frequency domain as,

$$Y(w) = X(w) + D(w) \tag{2}$$

Expressing $Y(w)$ in polar form as,

$$Y(w) = |Y(w)|e^{j\emptyset_y(w)} \tag{3}$$

where $|Y(w)|$is the magnitude spectrum and $\emptyset_y(w)$ is the phase of the corrupted noisy signal. The estimate of the clean signal is

$$\hat{X}(w) = [|Y(w)| - |\hat{D}(w)|]e^{j\emptyset_y(w)} \tag{4}$$

The enhanced speech signal can be obtained by simply taking the inverse Fourier transform of $\hat{X}(w)$. Multiplying $Y(w)$ by its complex conjugate $Y^*(w)$,

$$|Y(w)|^2 = |X(w)|^2 + |D(w)|^2 + X(w)D^*(w) + X^*(w)D(w)$$

$$= |X(w)|^2 + |D(w)|^2 + 2Re\{X(w)D^*(w)\} \tag{5}$$

Taking expectation on both sides of the equation with an assumption that $X(w)$ and $D(w)$ are uncorrelated to each other, then the above equation reduces to

$$|\hat{X}(w)|^2 = |Y(w)|^2 - |\hat{D}(w)|^2 \tag{6}$$

$$E[|\hat{X}(w)|^2] = E[|Y(w)|^2] - E[|\hat{D}(w)|^2] \quad 7)$$

This can be written as,

$$r_{\hat{x}\hat{x}}(m) = r_{yy}(m) - r_{\hat{d}\hat{d}}(m) \tag{8}$$

where $r_{\hat{x}\hat{x}}(m)$, $r_{yy}(m)$and $r_{\hat{d}\hat{d}}(m)$ are the autocorrelation sequences of the estimated clean signal, the noisy speech signal and the estimated noise signals respectively. Dividing Eqn(6) with $|Y(w)|^2$

$$\frac{|\hat{X}(w)|^2}{|Y(w)|^2} = 1 - \frac{|\hat{D}(w)|^2}{|Y(w)|^2} = H^2(w) \tag{9}$$

Where $H(w)$ is called as gain function and $0 \le H(w) \le 1$. A more generalized version of the spectral subtraction algorithm is given by [10]

$$\left|\hat{X}(w)\right|^p = |Y(w)|^p - \left|\hat{D}(w)\right|^p \tag{10}$$

Where p is power exponent, when p=1 yields the original magnitude spectral subtraction and p=2 yields the power spectral subtraction algorithm. The results obtained by considering an original noisy signal from www.noizeus.org named as 'sp01_car_sn15.wav' is shown in Fig.2.

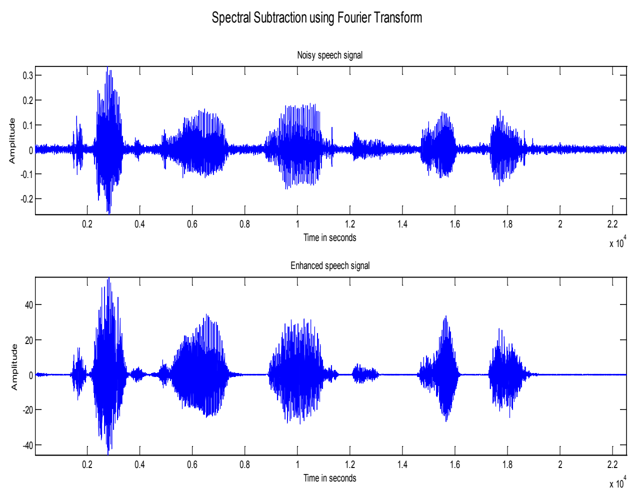

Figure 2: Output plots of spectral subtraction method

3. FRACTIONAL FOURIER TRANSFORM

The FRFT with an angle α of a signal $f(t)$ is defined as [2],

$$F_\alpha(u) = \int_{-\infty}^{\infty} f(t) K_\alpha(u,t) dt \tag{11}$$

where the transform kernel $K_\alpha(u,t)$ is given by

$$K_\alpha(u,t) = \begin{cases} \sqrt{\frac{1-j\cot\alpha}{2\pi}} e^{\frac{j(u^2+t^2)}{2}\cot\alpha - jut\operatorname{cosec}\alpha} & \alpha \neq k\pi \\ \partial(u-t) & \alpha = 2k\pi \\ \partial(u+t) & \alpha = (2k+1)\pi \end{cases} \tag{12}$$

where $\alpha = \frac{a\pi}{2}$ is rotation angle in the phase plane and a is a constant ranging from 0 to 4.

The inverse Fractional Fourier Transform is defined as

$$f(t) = \int_{-\infty}^{\infty} F_\alpha(u) K_{-\alpha}(u,t) du \tag{13}$$

Although FRFT has attracted the attention of many researchers its definition is vague. There are several interpretations and understandings about the operation of FRFT and hence several definitions has been proposed in literature [5,6,7,9], few of them are listed below.

i) H.M. Ozaktas definition[6]***:*** According to this FRFT is given by,

$$F_\alpha[f(x)] = \{F^\alpha f\}(x) = \int_{-\infty}^{\infty} B_\alpha(x,x')f(x')dx' \tag{16}$$

$$B_\alpha(x,x') = A_\emptyset exp\{i\pi[x^2\cot(\emptyset) - 2xx'\csc(\emptyset) + x'^2\cot(\emptyset)]\} \tag{17}$$

$$A_\emptyset = \frac{exp\left\{-i\pi\frac{sgn[\sin(\emptyset)]}{4} + i\pi/2\right\}}{|\sin(\emptyset)|^{1/2}} \tag{18}$$

where $\phi = \frac{\alpha}{2}$ and i is the imaginary unit. The kernel approaches $B_0(x,x') \equiv \delta(x-x')$ and $B_{\pm 2}(x,x') \equiv \delta(x+x')$ for $a=0$ and $a=\pm 2$ respectively.

ii) R. Tao definition: The definition of the FRFT can be written as [9]

$$\{F^\alpha f\}(u) = A_\emptyset e^{j\pi u^2 cot\emptyset}\int_{-\infty}^{\infty} e^{-j2\pi uxcsc\emptyset}[e^{j\pi x^2 cot\emptyset}f(x)]dx \tag{19}$$

The above definition simplifies the difficulties involved in implementing FRFT in hardware form and is widely used while implementing with FPGA.

iii)S.-C. Pei definition [8]***:*** The kernel of FRFT is defined as

$$K_\alpha(u,t) = \begin{cases} \sqrt{\frac{1-jcot\alpha}{2\pi}}e^{\frac{j(u^2+t^2)}{2}cot\alpha - jutcosec\alpha} & \alpha \neq k\pi \\ \partial(u-t) & \alpha = 2k\pi \\ \partial(u+t) & \alpha = (2k+1)\pi \end{cases}$$

$$= \sum_{n=0}^{\infty} e^{-jn\alpha}H_n(t)H_n(u) \tag{20}$$

where α indicates the rotation angle of transformed signal for the FRFT. $H_n(t)$ is the n^{th}-order normalized Hermite function with unit variance. Using the kernel of the FRFT, the FRFT of the signal $x(t)$ by angle α is computed as

$$x_\alpha(u) = \int_{-\infty}^{\infty} x(t)K_\alpha(t,u)dt$$

$$= \sum_{n=0}^{\infty} H_n(u)(e^{-jn\alpha}\int_{-\infty}^{\infty} x(t)H_n(t)dt) \tag{21}$$

The rotational properties of FRFT will not change by using Hermite polynomial definition.

iv) C. Candan definition[7]***:*** According to this definition,

$$F^\alpha[m,n] = \sum_{k=0}^{N-1} p_k[m](\lambda_k)^a p_k[n] \tag{22}$$

Where $p_k(n)$ is a set of orthonormal eigen vectors, (λ_k) is eigen values.

4. FRFT BASED SPECTRAL SUBTRACTION METHOD

The speech signal X is composed of clean speech S and additive noise N and can be written as,

$$X = S + N \tag{23}$$

where the S and N are mutually independent. Applying DFRFT to Eqn(23)

$$X_\alpha = S_\alpha + N_\alpha \tag{24}$$

If W_O represents optimal filter, ΔW_O represents random perturbation factor then,

$$W_\alpha = W_O + \beta \Delta W_O \tag{25}$$

$$\hat{S}_\alpha = W_\alpha X_\alpha = (W_O + \beta \Delta W_O) X_\alpha \tag{26}$$

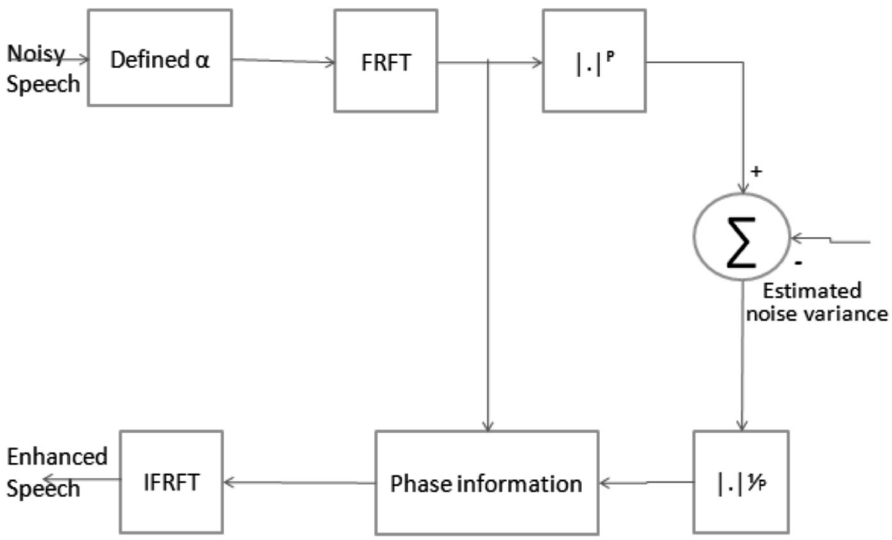

Figure 3: FRFT based spectral subtraction method

The cost function, J is minimum mean square error (MMSE) and is given by,

$$J = \sigma^2 = E(|S - \hat{s}|^2) = E(|S_\alpha - \hat{s}_\alpha|^2) \tag{27}$$

From Eqn(26) and Eqn (27),

$$\frac{\partial \hat{s}_\alpha}{\partial \beta} = \frac{\partial}{\partial \beta}(W_\alpha X_\alpha) = \frac{\partial}{\partial \beta}[(W_O + \beta \Delta W_O) X_\alpha] = \Delta W_O X_\alpha \tag{28}$$

$$\frac{\partial J}{\partial \beta} = \frac{\partial}{\partial \beta} E(|S - \hat{s}|^2) = 2\Delta W_O (W_\alpha [R_{XX}]_\alpha - [R_{SX}]_\alpha) \tag{29}$$

W_α can be obtained by equating $\left.\frac{\partial J}{\partial \beta}\right|_{\beta=0} = 0$

$$W_\alpha = \frac{[R_{SX}]_\alpha}{[R_{XX}]_\alpha}$$

$$J = \sigma^2 = E(|S_\alpha - \hat{s}_\alpha|^2) J = \sigma^2 = E(|S_\alpha - \hat{s}_\alpha|^2)$$

$$= [R_{SS}]_\alpha - 2W[R_{SX}]_\alpha + |W|^2 [R_{XX}]_\alpha \tag{30}$$

From Eqn (24) $$S_\alpha = X_\alpha - N_\alpha$$

$$\hat{S}_\alpha = W_\alpha X_\alpha = W_\alpha(S_\alpha + N_\alpha) = W_\alpha[FRFT(S+N)] \tag{31}$$

Time domain value is obtained by taking inverse transform to Eqn (31)

$$\hat{S}(t) = FRFT^{-1}(\hat{S}_\alpha) = FRFT^{-1}[W_\alpha FRFT(S+N)] \tag{32}$$

ŝ(t) is the enhanced output signal. For a system with input signal *i* of length *l* and output signal *o* of length *m*, the figure of merits Mean Square Error(MSE) and Peak Signal to Noise Ratio(PSNR) are defined as [4,10,11]

$$MSE = \frac{1}{lm}\sum[o-i]^2 \tag{33}$$

$$PSNR = 10\log_{10}\left(\frac{MAX_i^2}{MSE}\right) \tag{34}$$

Under ideal conditions the MSE should be 0 and PSNR to be ∞.

5. EXPERIMENTAL RESULTS

To check the efficacy of the proposed method, above mentioned four definitions of FRFT have been taken into consideration. The corrupted signal that has been collected from noizeous data base is applied to the proposed algorithm. Variation of MSE and PSNR with respect to fractional order α is as shown in Fig.4 and Fig.5 and comparison of PSNR and MSE values of various definitions of FRFT is shown in Table 1.

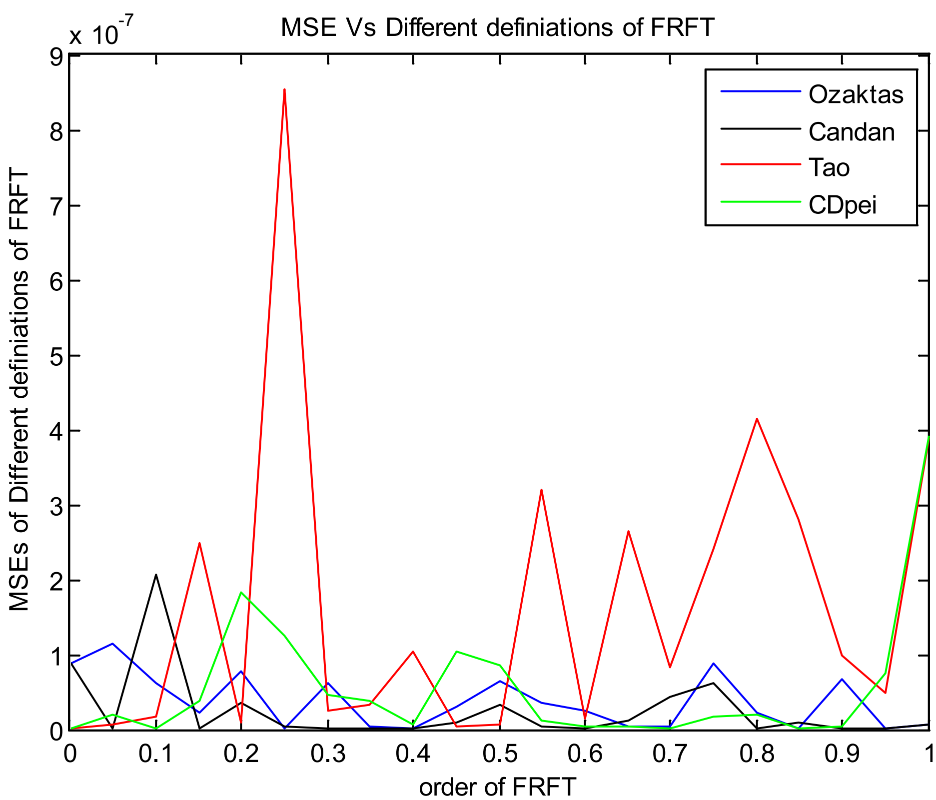

Figure 4: MSE plots of different definitions of FRFT

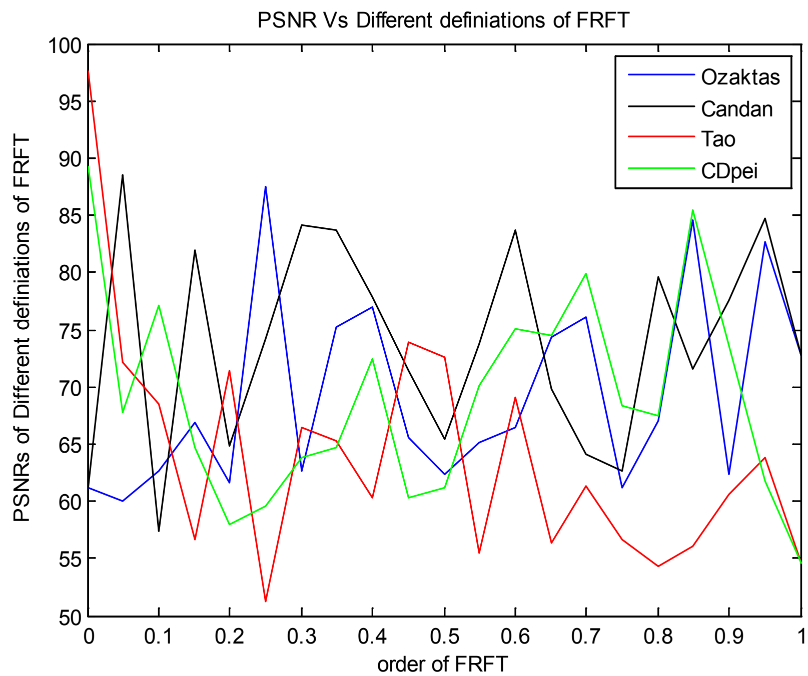

Figure 5: PSNR plots of different definitions of FRFT

Table 1
Comparison of PSNR and MSE values of various definitions of FRFT

Order α	*Parameter*	*FFT*	*FRFT*			
			Ozaktas	*R.Tao*	*S.C.Pei*	*Candan*
0.05	MSE	7.11×10^{-7}	1.13×10^{-7}	1.57×10^{-6}	1.88×10^{-8}	4.14×10^{-7}
	PSNR	73.285	59.99	54.26	67.79	54.36
0.1	MSE	7.11×10^{-7}	6.25×10^{-8}	2.07×10^{-7}	2.20×10^{-9}	1.54×10^{-8}
	PSNR	73.285	62.57	57.36	77.10	68.50
0.2	MSE	7.11×10^{-7}	7.55×10^{-8}	3.67×10^{-8}	1.83×10^{-7}	8.13×10^{-9}
	PSNR	73.285	61.64	64.89	57.90	71.43
0.5	MSE	7.11×10^{-7}	6.55×10^{-8}	3.19×10^{-8}	8.64×10^{-8}	6.19×10^{-9}
	PSNR	73.285	62.37	65.49	61.17	72.61
0.7	MSE	7.11×10^{-7}	2.77×10^{-9}	4.38×10^{-8}	1.16×10^{-9}	8.22×10^{-8}
	PSNR	73.285	76.10	64.11	79.88	61.39
0.8	MSE	7.11×10^{-7}	2.22×10^{-8}	1.25×10^{-9}	2.05×10^{-8}	6.92×10^{-9}
	PSNR	73.285	67.05	79.55	67.41	72.13
0.9	MSE	7.11×10^{-7}	6.63×10^{-8}	1.95×10^{-9}	4.99×10^{-9}	9.96×10^{-8}
	PSNR	73.285	62.32	77.63	73.55	60.55

6. CONCLUSION

Removal of noise using fractional order spectral subtraction method is the main objective of this paper. Initially conventional spectral subtraction method, basics of FRFT is studied. Next generalization of the spectral subtraction method is proposed. The performance of the proposed method is checked by varying different values of fractional order, α. From plotted graphs between MSE versus α, PSNR versus α, it can be observed that, for $\alpha<0.3$, the performance of ozaktas, candan, pei and tao definitions has produced same results. The error produced by candan FRFT for $0.3<\alpha<1$ is small compared to other definitions.

The order of good performance in error performance is candan, ozaktas, pei and tao definitions. The tao definition has shown very poor performance at any value of α. It has been further observed that $0.3<\alpha<0.6$, the performance of all the FRFT's is relatively similar. Finding the optimum value of α, to produce lesser MSE and higher PSNR is still need to be addressed in future papers.

References

[1] Rajiv Saxena and KulbirSingh, "Fractional Fourier transform: A novel tool for signal Processing" *J. Indian Inst. Sci*., 85, 11–26, Jan.–Feb. 2005.

[2] H. M. Ozaktas, Z. Zalevsky and M. A. Kutay, "The fractional Fourier transform with applications in optics and signal processing" Wiley 2000.

[3] S. C. Pei and J. J. Ding, "Relations between fractional operations and time-frequency distributions and their applications" *IEEE Trans. Signal Processing*, 49, 1638–1655,2001.

[4] Zhu-Gao Ding and Feng-Qin Yu, "Improved Speech Denoising Algorithm Based on Discrete Fractional Fourier Transform," *Journal of Electronicscience and Technology of China,* Vol. 6, No. 1, March 2008.

[5] Ervin Sejdic, Igor Djurovic and L Jubisa Stankovic, "Fractional Fourier transform as a signal processing tool: An overview of recent developments" *Signal Processing*, Vol. 91 No. 6, June, 2011.

[6] H.M. Ozaktas, M.A. Kutay, and G. Bozdagi, "Digital computation of the fractional Fourier transform" *IEEE Trans. Sig. Proc*., 44:2141-2150, 1996.

[7] C. Candan, M.A. Kutay, and H.M. Ozaktas, "The discrete Fractional Fourier Transform" *IEEE Trans. Sig. Proc.,* 48:1329-1337, 2000.

[8] S.C. Pei, M.-H. Yeh, "Discrete fractional Cosine and Sine transforms" *IEEE Trans. Sig. Proc.,* 49(6):1198-1207, 2001.

[9] R. Tao, G. Liang, X. Zhao, "An efficient FPGA-based implementation of fractional Fourier transform algorithm" *J. Signal Processing Systems*, 60(1):47-58, 2010.

[10] Philipos C. Loizou, "Speech Enhancement Theory and Practice", Taylor and Francis, 2007.

[11] S. F. Boll "suppression of acoustic noise in speech using spectral subtraction," IEEE transactions on acoustics, speech, and signal processing, vol. assp-27, no. 2, april 1979.