2-D Speech Enhancement based on Curvelet Transform using Different Window Functions

A.K. Verma
Department of Electronic & Communication Engineering, Rajasthan Institute of Engineering and Technology Jaipur, India.

A.R. Verma
Department of Electronic & Communication Engineering, G. B. Pant Engineering College Pauri-Garhwal, India

Manoj Kumar
Department of Electronic &Communication Engineering, G. B. Pant Engineering College Pauri-Garhwal, India

ABSTRACT
In this paper, an improved method based on Curvelet Transform using different window functions is presented for the speech enhancement. The window function is used for pre-processing of speech signals. In this method, instead of using two-dimensional (2-D) discrete Fourier Transform, Curvelet transform is employed with spectral magnitude subtraction method.

General terms
Spectral Substraction Method, Curvlet transform.

Keywords
Spectral subtraction method, Cosh, Exponential, Hamming, Hanning, Curvelet transform.

1. INTRODUCTION
An explosive advances in recent years in the field of digital computing have provided a remarkable progress to the field of speech processing. The speech processing is an application of signal processing to acoustic signals using the knowledge offered by researchers in the field of hearing sciences. The speech processing systems are most commonly used in many fields such as voice communication and voice recognition [1-8].

Generally, speech enhancement techniques are classified into two categories: time domain techniques and transform domain techniques. Time domain techniques were based on conventional filtering approach such as linear predictive coding based filtering, Hidden Markov Model and Kalman filtering. Hidden Markov Model is used firstly for speech enhancement [9-17]. A marked research progress have been made in many transformation methods such as Discrete Cosine Transform (DCT),

Fast Fourier Transform (FFT), Karhunen-Loeve Transform (KLT) Discrete Wavelet Transform (DWT), Ridgelet and Curvelet transform which are extensively used in data compression, detection and classification. For noisy speech decomposition, many transformation techniques are used such as KLT is an Eigen value decomposition technique, DCT, DWT, Ridgelet and Curvelet transform which is computationally efficient [18-22].

2. OVERVIEW OF CURVELET ANALYSIS
Curvelet are defined at scale 2^{-j}, orientation θ_l and position $X_k^{j,l} = R_{\theta l}^{-1}(2^{-j} k_1, 2^{-j/2} k_2)$ by translation and rotation of a mother Curvelet ψ_j as

$$\varphi_{j,l,k}(X) = \varphi_j(R_{\theta_l}(X - X_K^{j,l})) \qquad (1)$$

where R_{θ_l} is the rotation by θ_l radians. θ_l is the equi-spaced sequence of rotation angles $\theta_l = 2\pi 2^{-[j/2]} l$, with integer l such that $0 \leq \theta_l \leq 2\_$ (note that the number of orientations varies as 1/√scale). k = (k_1, k_2) ∈ Z^2 is the sequence of translation parameters. The waveform φ_j is defined by means of its Fourier transform $\hat{\varphi}_j(v)$, written in polar coordinates in the Fourier domain.

$$\hat{\varphi}(r,\theta) = 2^{-3j/4} \hat{\omega}(2^{-j} r) \hat{v}\left(\frac{2^{\lfloor j/2 \rfloor}\theta}{2\pi}\right) \qquad (2)$$

The support of $\hat{\varphi}_j$ is a polar parabolic wedge defined by the support of $\hat{\omega}$ and $\hat{v}$, the radial and angular windows (both smooth, nonnegative and real-valued), applied with scale-dependent window widths in each direction. $\hat{\omega}$ and $\hat{v}$ must also satisfy the partition of unity property [23-25].
In continuous frequency v, the CurveletG2 coefficients of data f(x) are defined as the inner product.

$$c_{j,l,k} := < f, \varphi_{j,l,k} > = \int_{R^2} \hat{f}(v) \hat{\varphi}_j(R_{\theta_l} v) e^{i X_K^{j,l} \cdot v} dv \qquad (3)$$

This construction implies a few properties: (i) the CurveletG2 defines a tight frame of $L_2(R^2)$, (ii) the effective length and width of these Curvelet obey the parabolic scaling relation (2^{-j} = width) = (length = $2^{-j}/2$)2, (iii) the Curvelet exhibit an oscillating behavior in the direction perpendicular to their orientation. Curvelet as just constructed are complex-valued. It is easy to obtain real-valued Curvelet by working on the symmetrized version.

We have introduced in this paper two new transforms, the Ridgelet transform and the Curvelet transform. Several other transforms are often used in astronomy, such the Fourier transform, the isotropic `a trous wavelet transform and the bi-orthogonal wavelet transform. The choise of the best transform may be delicate. Each transform has its own domain of optimality:

The Fourier transform for stationary process

The a trous wavelet transform for isotropic features. The bi-orthogonal wavelet transform for features with a small anisotropy, typically with a width equals to half the length.

- The Ridgelet wavelet transform for anisotropic features with a given length (i.e. block size).
- The Curvelet transform for anisotropic features with different length and width equals to the square of the length.

3. CURVELET TRANSFORM BASED METHODOLOGY FOR SPEECH ENHANCEMENT

The noise taken for the speech enhancement is additive noise model to model background noise; the noise and the speech both are uncorrelated. The following equation shows the additive noise model,

$$y(t) = x(t) + n(t) \tag{4}$$

Where y (t) is the observed noisy speech, x (t) is the clean speech and n (t) additive background noise. The observed speech signal is divided into the overlapping frame. The length of each frame is 256 samples. The overlap taken between the two consecutive frames is from 50% or 75%. The overlap between frame proposed by Soon and Koh is taken 75%.That means each frame is shift by previous frame by 64 samples [8]. To maintain the order of continuity of first and last frame, every frame is multiply by a window function.

The dot product of window and block is windowed speech block. Thus Ridgelet transform can then be applied onto the speech block. After applying Ridgelet transform then applying the spectral

magnitude subtraction method [8] can be applied to the case. In this scheme the magnitude of the Ridgelet transform coefficient is attenuated by a threshold which is dependent on the expected noise magnitude. Negative resultant values are clipped to zero. The attenuated magnitude is then combined with the noisy phase before the inverse Ridgelet transform operation is carried out.

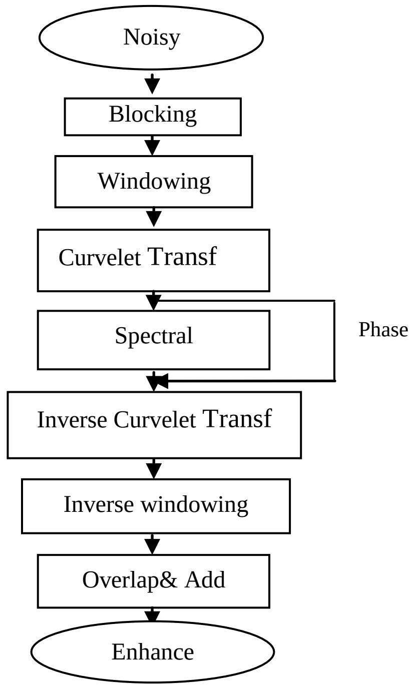

Fig.1. Block Diagram of spectral magnitude subtraction method

Let the transformed noisy speech block be represented by Y and the enhanced speech block be represented by $\hat{X}$ .The enhancement process can explain by this equation.

$$\left|\hat{X}(u,v)\right| = \max(\left|Y(u,v)\right| - E[\left|N(u,v)\right|], 0) \tag{5}$$

Where $E[\left|N(u,v)\right|]$ expected mean represent of the noise and enhance signal represented by this equation

$$\hat{X}(u,v) = \left|\hat{X}(u,v)\right| \exp^{(i\theta_Y(u,v))} \tag{6}$$

The enhanced speech block in the time domain is obtained through invers Ridgelet Transform. The final enhanced speech is obtained by the reversing the blocking and framing process [8] finally speech enhanced signal is represented by this equation

$$\hat{x}(t) = 0.25(f_L(j) + f_{L-1}(j+64) + f_{L-2}(j+128) + f_{L-3}(j+192)) \tag{7}$$

Where, $L = \left\lfloor \frac{t}{64} \right\rfloor$ And $j = t - 64L$

4. RESULT AND DISCUSSIONS

Table 1 Performance of different sample based on Curvelet Transform using CosH window, Table 2 Performance of different window base on Curvelet Transform, Table 3 Performance of different Transform using five samples through CosH window. Thus illustrate the performance of different windows i.e. CosH window, Exponential, Hanning, hamming window for the speech enhancement using different transform. It’s shown from the below results the Curvelet transform gives better performance than other transform in terms of noise measurement parameter. Here, the results are evaluated based on different fidelity parameters such as signal-to-noise ratio (*SNR*), maximum error (*ME*) and mean square error (*MSE*), these are define as:

- Signal to noise ratio (SNR):

$$SNR = 10\log_{10}\left(\frac{\text{energy of input signal}}{\text{energy of the reconstructed error}}\right) = 10\log_{10}\left\{\frac{\sum x^2(n)}{\sum |x(n)-y(n)|^2}\right\} \quad (8)$$

Mean square error (*MSE*)

$$MSE = \frac{1}{2}\sum_n |x(n)-y(n)|^2 \quad (9)$$

Maximum error

$$ME = \max|x(n)-y(n)| \quad (10)$$

Table.1 Performance of different sample based on Curvelet Transform using CosH window

Curvelet Transform	Input SNR (db)	Output SNR (db)	MSE	ME
Sample1	-10	36	0.0111	0.5431
Sample2	-5	28	0.0103	0.5051
Sample3	0	41.5	0.0121	0.3921
Sample4	5	38	0.0101	0.2032
Sample5	15	40	0.0104	0.1004
Sample6	20	49.5	0.0161	0.03210
Sample7	25	48	0.0231	0.2214

Table.2 Performance of different window base on Curvelet Transform

Window	Input SNR (db)	Output SNR (db)	MSE	ME
Hamm	4.14	18.67	0.1001	0.2321
Hann	4.14	12.89	0.0121	0.4601
Black man	4.14	24.63	1.1861	0.0324
Exponential	4.14	29.41	0.0854	1.0931
CosH	4.14	36.49	0.0967	1.4321

5. GRAPHICAL RESULT OF SPEECH ENHANCEMENT

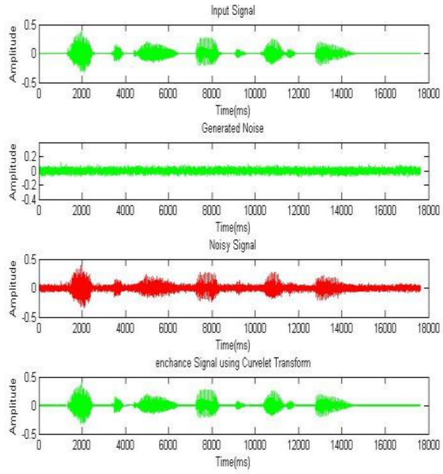

Fig. 2 a) Input signal (i.e. Speech signal) b) Generated noise (6db) c) Noisy signal d) Enhance signal using Curvelet Transform

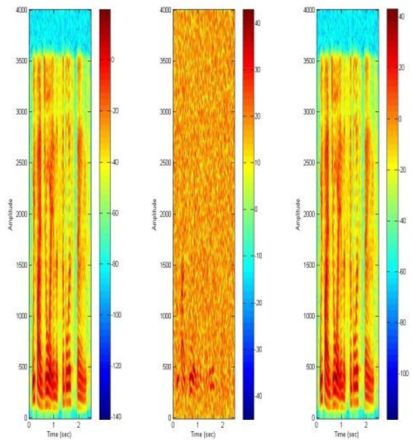

Fig. 3 a) Input signal (i.e. Speech signal) b) Noisy signal c) Enhance signal using Curvelet Transform.

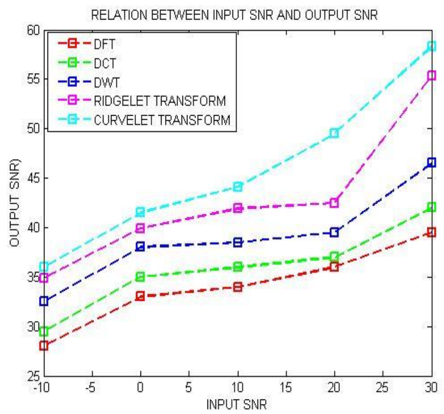

Fig. 4 Relation between input SNR and output SNR

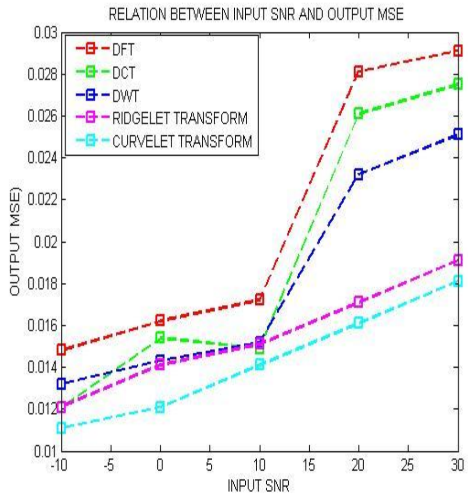

Fig. 5 Relation between input SNR and output MSE

6. CONCLUSION

We use different windows and transform through spectral magnitude subtraction for speech enhancement. The result is tested on number of noise measurement parameters Minimum Square Error (MSE), Signal to Noise Ratio (SNR), Maximum error (ME). The CosH window and Curvelet transform gives the better result in all aspects of noise parameter. It is clear from this definition of the Curvelet transform that the transform is separable and can be implemented. This paper work is to analysis performance of Curvelet transform for the enhancement of the speech signal Curvelet transform gives better performance than other transform.

7. REFERENCE

[1] Virag, N. Single channel speech enhancement based on masking properties of the human auditory system, IEEE Trans on Speech and Audio Processing, 7, 126-137, (1999).

[2] Yang, L., Shuangtian, L. DCT speech enhancement based on masking properties of human auditory System. Proc, Inst of Acoust, Chinese Acad of Scien, 450-453,(2010).

[3] Lei, S.F, Tung, Y.K., Speech enhancement for nonstationary noises by wavelet packet transform and adaptive noise estimation, Proceedings of Int. Symposium on Intelligent Signal Processing and Communication Systems, 41-44 (2005).

[4] Jr, J.D., Hansen, J., Proakis, J. Discrete-Time Processing of Speech Signals, NY: IEEE Press, 2000.

[5] Levitt, H. Noise reduction in hearing aids: An overview, Journal of Rehabilitation Research and Development, 38(1), 111-121, (2001).

[6] Berouti, M., Schwartz, R, Makhoul, J. Enhancement of speech corrupted by acoustic noise, Proc. IEEE Int. Conf. on Acoust., Speech, Signal Procs., (1979), 208-211.

[7] Boll, S. F. Suppression of acoustic noise in speech using spectral subtraction, IEEE Transaction Acoust., Speech, Signal Processing, 27, 113-120, (1979).

[8] Soon, I.Y., Koh, S.N., Speech enhancement using 2D Fourier transform, IEEE Transactions on speech and audio processing , 11717-724 ,(2003).

[9] Vaseghi, S.V. Advanced digital signal processing and noise reduction, Wiley, 2nd edition, 270-290, (2000).

[10] Ephraim, Y., Malah, D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator, Acoustics, Speech, and Signal Processing 32(6) 1109-1121, (1984).

[11] Paliwal, K. K., Basu, A. A speech enhancement method based on Kalman filtering, in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, 177-180, (1987).

[12] Huang, J., Zhao, Y. A DCT-based fast signal subspace technique for robust speech recognition, IEEE Trans, on Speech and Audio Proc, 8(6), 747-751, (2000).

[13] Ephraim, Y. "Statistical-Model-Based Speech Enhancement Systems," Proceeding of the IEEE, 80(10), 1526-1555, (1992).

[14] Scalart, P., Filho, J.V. Speech enhancement based on a priori signal to noise estimation, in Proc. ICASSP, 2, 629–632, (1996).

[15] Makhoul, J. A Fast Cosine Transform in One and Two Dimensions, IEEE Trans on Acoust, Speech, and Signal Proce, assp-28, no.1, 27-34, (1980).

[16] Mittal, U., Phamdo, N. Signal/noise KLT based approach for enhancing speech degraded by colored noise, IEEE Trans. Speech Audio Processing, vol. 8,159–167, (2000).

[17] Jafer, E., Mahdi, A.E. Wavelete-based Perceptual Speech Enhancement Using Adaptive Threshold Estimation, In eurospeech 569-572, (2003).

[18] Bahoura, M., Rouat, J. Wavelet Speech Enhancement Based on the Teager Energy Operator.IEEE Sig. Proc. Letters, vol. 8, 10-12, (2001).

[19] Kaiser,J.F., Schafer, R.W., On the use of the IO-sinh window for spectrum analysis. IEEE Trans Acoustics Speech Signal Process 28:105–107, (1980).

[20] Saramaki, T. A class of window functions with nearly minimum sidelobe energy for designing fir filters. In: Proceedings of the IEEE international symposium circuits and systems, Portland, Ore, USA p. 359–62 (1989).

[21] K. Avci, A. Nacaroglu, A New Window Based on Exponential Function, IEEE Conference on Research in Microelectronics and Electronics, (2008) 69-72.

[22] A. Kumar, B. Kuldeep, Design of M-channel cosine modulated filter bank using modified Exponential window. Journal of the Franklin Institute 349, 1304–1315, (2012).

[23] C. Plapous, C. Marro and P. Scalart, Improved signal-to-noise ratio estimation for speech enhancement IEEE Transaction. Audio, Speech, And Language Processing, 14, (2006).

[24] Plapous, C., Marro, C., Scalart, P., Mauuary, L. A two-step noise reduction technique, InProc. IEEE International conference. Acoustic. Speech, Signal Process. Montréal, QC, Canada, 1, (2004), 289–299.

[25] Walker, J.S. A Primer on Wavelets and Scientific Applications 1999.