Active Voice Detection Using Ridgelet Transform

Hanaa Mohsin Ahmed*, Ph.D (Asst. Prof.)

Abstract

Regularly, voice activity detection is the main crucial parameter in any control system using command of voice speech, spatially in environments with noise. The uses voice activity detection method affects both, complexity of computational and the overall performance of any control system using command of voice speech. This paper handles the problem of automatic word boundary detection in both cases (the noiseless and noisy backgrounds), by proposed voice activity detection of speech signals based on using Ridgelet Transform (RT). It uses the Inverse Ridgelet Transform (Ridgelet Transform multiply Ridgelet Transform), which gives the accuracy in environments with noise. The experimental results point up the effectiveness of the used method in low signal to noise ratio environments, for command voice speech signal detection capability.

Keywords: Ridgelet Transform, Voice Activity Detection, Isolated Word/Sentence, Command Control, Noise.

*University of Technology

1. Introduction

Generally, in environments with noise, distinguishing voiced speech signal from unvoiced speech signal and determine start and end point of voice speech signal, is a demanding requirement in any control system using command of voice speech. Since this operation is the most important step of any control system using command of voice speech, because of its direct impact on improve the overall ratio under different types of background noise and decrease the computing power lost induced by incorrect speech detection. A feature parameter that can adequately specify the characteristics of a voice speech and be robust in environments with noise is very crucial, [1-4].

Researchers point up that, even in environments without noise, errors in any control system using command of voice speech is due mainly to the accurate active voice speech detection [1], and [4].

In environments with noise, almost the existing methods so for can't always spot the endpoint of the voice speech with appropriate accurate.

In this paper RT is proposed for speech voice detection, in environments with noise. In the following subsections, background knowledge is presented in section 2. Second, the proposed Ridgelet based voice activity detection is presented in sections 3. In section 4 and section 5, shows the Experiment results, and conclusion.

2. Theoretical Part

2.1 The Transform of Ridgelet

The continuous representation of RT for a given signal $s \in L^2(\mathbb{R}^2)$, is the scalar product of s with product of $\psi_{a,b,\Theta}(x)$. The two-dimensional Ridgelet faction is defined using one-dimensional wavelet function ψ such as [5]:

$$\psi_{a,b,\Theta}(\chi) =$$

$$a^{-1/2}\psi\left(\frac{x_1 cos\Theta + x_2 sin\Theta - b}{a}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(1)}$$

Where:

$\chi = (x_1,x_2)$ □ $(\mathbb{R}^2)$, and the directional parameter,Θ □ $[0,2\pi]$,$a,b \in (\mathbb{R})$, a is dilation, and b is translation. $\psi_{a,b,\Theta}(\chi)$, oriented at the AngeleΘ, and is constant at the line $(x_1\ cos\Theta + x_2\ sin\Theta)$. Transverse to all these ridges is a Wavelet. Ridgelet coefficients can be defined as [5, 6]:

$$\mathbf{R\Theta} = \int \boldsymbol{\psi}_{\mathbf{a,b,\Theta}} \mathbf{f(x_1,x_2)dx_1dx_2}, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \mathbf{(2)}$$

Also the RT can be represented by using the Radon Transform (RA), as depicted in Figure (1). The RA of a signal $f(x_1,x_2)$ is defined as [5, 6]:

$$\mathbf{RA(\Theta,t)} = \int \mathbf{f(x_1,x_2)\delta(x_1cos\Theta + x_2sin\Theta - t)\,dx_1dx_2} \ldots\ldots\ldots (3)$$

Where δ is the Dirac distribution, the angular variable Θ is constant and t is varying. So RT is indeed the application of a one-dimension Wavelet Transform (WT) to the slices of the RA [5, 6].

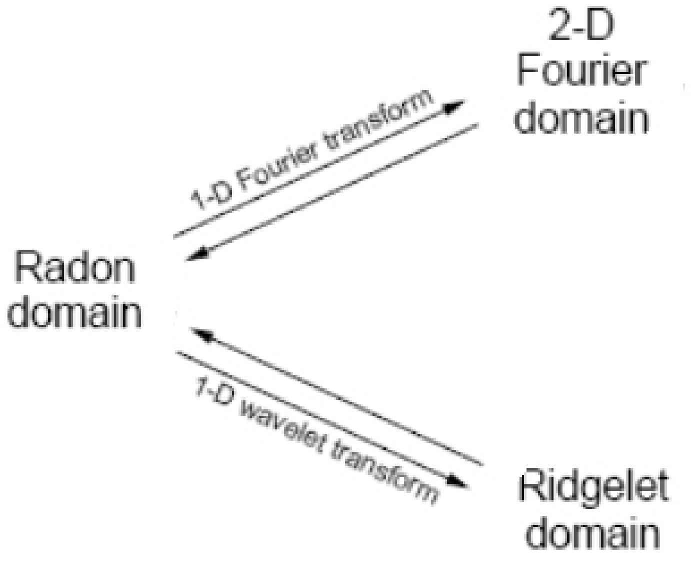


Figure (1): The relation between RT and RA

2.2 Algorithm of Discreet Ridgelet Transform

The main steps of applying discreet Ridgelet Transform is by: Apply the RA (two dimension signal). Then apply the on-dimension scalar WT to the resulting angular lines in order to obtain the Ridgelet coefficients.

2.3 Algorithm of applying Radon transforms

The main steps of applying RA by: apply the Fast Fourier Transform (FFT) to two-dimension signal. Then substitute the sampled values of the FFT that obtained on the square lattice with sampled values on a polar lattice. Finally apply one-dimension Inverse Fast Fourier Transform (IFFT) on each angular line.

2.4 Algorithm of applying Inverse Ridgelet Transform

The main steps of applying inverse discreet Ridgelet by: apply one-dimension inverse scalar wavelet transform (IWT) on the Ridgelet coefficients. Then apply the Inverse Radon transform (IRA).

2.5 Algorithm of applying Inverse Radon transforms

The main steps of applying IRA by: apply one-dimension IFFT. Then substitute the sampled values of the IFFT that obtained on the polar lattice with sampled values on a square lattice. Apply one-dimension inverse FFT on each angular line.

3. The proposed Ridgelet Based Voice Activity Detection

First of all apply wave recording step, that is determine as input duration of recording time (one second), sampling rate (eleven KHz), number of channel (one), and data format (sixteen bits), then record wave data using (duration of recording time multiplied by sampling rate, no. of channel, data format) . Second step is applying the proposed voice activity detection to recorded wave; then find the inverse integer Ridgelet transform, to the result of multiply integer Ridgelet transform to integer

Ridgelet transform, the mean of result is subtract from the original input signal, and apply as input threshold to signal noise removal using Ridgelet transform. Finally; apply Endpoint Detection step by: determine the position of the non-zero elements position. The Start point is first element position over the original signal, and the end is last element position, of the previously determine non-zero elements over the original signal.

4. Experiment Results

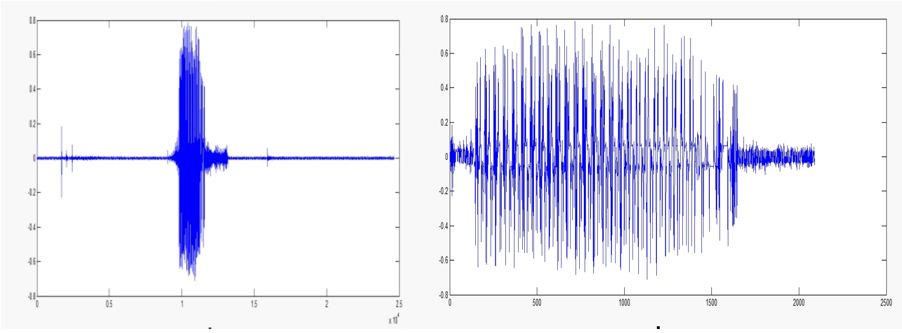
4.1 Test Samples Set:

Generally, for all experiments used in this paper, it has collected voice speech samples by using a sound_card in a wave file format. These recorded voice speech samples have used, sampling_frequency of eight_KHz, eight_bits to quantify, sixteen_bit sampling accuracy. The Mat Lab is used as programming tool.

The background noise have used five common known noise, including white-noise, speaking background-noise, factory-noise, high speed car-noise and Pink-noise. For different Signal to Noise Ratio (SNR), gives detection of endpoint to test samples set that have recorded, and evaluation with the correct rate for detection. The end points of the test samples set are allocated with marks, manually, first then compared with results from the proposed method for accurate allocation requirements.

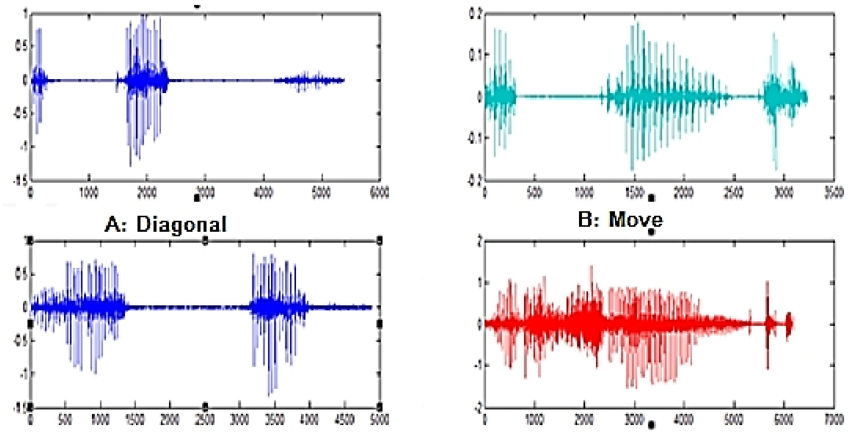
4.2 Experimental Results:

In this paper, the proposed RT active voice detection method is examined with a five different command words, {Reverse, Diagonal, Move, Left, Right}. Figure (2), is the example (input and output) of proposed algorithm to left command speech. The results of applying the proposed algorithm to a four different command words are shown in Figure (3).



A: input speech command B: Output speed command

Figure (2): Example of applying the proposed Algorithm



A: Diagonal

B: Move

C: Reverse

D: Right

Figure (3) (A-D): The results of applying the proposed method to four different command speech

4.3 Analysis of the results:

Table 1, present comparison results of the detected accuracy for the proposed algorithm, in case of five different types of noise, and SNR. As can be deduce from Table 1, that the proposed algorithm for, using Ridgelet as accurate parameter for threshold, to extract the active voice speech from un-active voice speech and detection the endpoints.

Table 1: SNR of five different types of noise

Noise Level Noise Type	-5	0	5	10	15
White	**70.96**	**77.77**	**81.66**	**86.56**	**95.11**
Speaking-background	**60.61**	**65.92**	**76.42**	**81.72**	**88.96**
Factory	**76.95**	**84.44**	**88.97**	**92.99**	**96.33**
High-speed	**76.43**	**78.75**	**82.58**	**86.18**	**90.98**
Pink	**71.36**	**78.27**	**80.96**	**85.39**	**94.31**

4.4 Comparison of the methods:

VAD is also used in speech recognition, and speech coding to detect speech signal from non_speech signal. The most crucial characteristics of AVD are those related to reliability, robtness, simplisity, real time and accurse. Many features are used for AVD from these are: short time energy, zero crossing rate, autocorrelation, spectrum, also multiples of these features are applied [7]. For the AVD methods, proposed method is compared with three previous methods based command speech, such as: Rabiner's method [8]: which used the first 10 frame as threshold to compute the, AVD for speech signal. Qiangm method [9]: which, he used the threshold value between [2, 10]. Frequency Domain method [10]: which, used energy threshold in the frequency domain to compute the, AVD for speech signal. The proposed method uses the normalized distribution of pre-computed Ridgelet transform as threshold to compute the, AVD for speech signal.

5. Conclusion

With this research paper, we have presented propose for command voice speech detection, that is used in any control system, based on using RT. The proposed method uses the normalized distribution of pre-computed Ridgelet transform as threshold as a first step and used to compute the, AVD for speech signal. The results of using different noise types, point up its capability to detect, the endpoint of the active voice speech signal. The test results of using different SNR for the proposed method, point up its effectiveness in accurate detection, and efficient proposed algorithm for locating the end-points of an utterance in a background of noise.

References

[1] J. Zhang, F. Jiang, H. LIU, Study on endpoint detection based on multi-characteristic jointed in noisy environment, Computer Engineering and Applications, 2009 (45): 114-116.

[2] J.F. Kaiser, On a simple algorithm to calculate the energy of a signal, IEEE International Conference on Acoustics, Speech and Signal Processing [A] (1990), pp. 381–384

[3] L. Jie, Z. Ping, J. Xinxing, D. Zhiran, Speech Endpoint Detection Method Based on TEO in Noisy Environment, Procedia Engineering, 29, 2012, 2655-2660

[4] J. XinXing and S. Xu, “Speech Recognition Based on Efficient DTW Algorithm and Its DSP Implementation,” Procedia Eng., vol. 29, pp. 832–836, Jan. 2012.

[5] G.Y. Chen, B. Kégl, “Image de-noising with complex ridgelets", Pattern Recognition Val No. 40, Page 578 – 585, 2007.

[6] H. Qiangui , H. Boya, C. Sheng, “Adaptive Digital Ridgelet Transform and its Application in Image Denoising”, Digital Signal Processing, Volume 52, May 2016, Pages 45–54

[7] A. Akila. and E. Chandra, Comparative Study of Endpoint Detection Algorithms Suitable for Isolated WordRecognition, International Journal of Information Technology Bharati Vidyapeeth’s Institute of Computer Applications and Management, New Delhi (INDIA), 2014.

[8] L.R.Rabiner and M.R.Sambur, “An Algorithm for Determining the endpoints of isolated utterances”, The Bell System Technical Journal, Vol. 54, No. 2, February 1975, pp 297-315

[9] Qiang He “On Prefiltering and Endpoint Detection of Speech Signal”, ICSP,1998

[10] Kirill Sakhnov, “Approach for Energy-Based Voice Detector with Adaptive Scaling Factor”, IAENG International Journal of Computer Science, Vol. 36, No. 4, November 2009.

كشف ألاشارة الصوتيه بأستخدم تحويل الرجيليت

أ.م.د.هناء محسن احمد*

المستخلص

من المعروف ان تحديد بدايه ونهايةالصوت هو المعلمة الرئيسية الحاسمة في أي نظام للسيطرة بأستخدام اوامر التعبير الصوتي، وخاصة عندما تكون البيئه ذات ضوضاء .فيؤثر على الطريقة الستخدامة لتحديد بدايه ونهايةالصوت على حد سواء لكلا من ، التعقيد الحاسوبي و الأداء العام لأي نظام للسيطرة بأستخدام اوامر التعبير الصوتي.يتناول هذا البحث مشكلة تحديد بدايه ونهايةالصوت التلقائي لكلا الحالتين (بوجود الضوضاء و بعدم وجود الضوضاء) ، فمن خلال المقترح المقدم لتحديد بدايه ونهايةاشارات الصوت, المبني على استخدام تحويل الرجيليت(TransformRidgelet(RT . بأستخدام معكوس تحويل الرجليت مظروبا في(تحويل الرجيلييت في تحويل الرجلييت) ، الذي يعطي دقة عاليه في بيئات ذات الضوضاء . وتشير النتائج التجريبية الى فعالية الطريقة المستخدمة في القدرة على كشف إشارة الكلام وبالاخص في البيئات ذات نسبة الضوضاء الواطي.

*الجامعةالتكنولوجية