DENOISING SPEECH OF MARATHI NUMERALS USING SPECTRAL SUBTRACTION

RAMTEKE G.D.[1]*, NIMBHORE S.S.[2] AND RAMTEKE R.J.[1]

[1]Department of Computer Science, North Maharashtra University, Jalgaon, MS, India.
[2]Department of Computer Science & IT, Dr.B.A.M.University, Aurangabad, MS, India.
*Corresponding Author: Email- shyam.ramteke@gmail.com

Abstract- This paper presents the denoising technique based on spectral subtraction for speech synthesis of the Marathi numerals. Numerals are recorded through mice and normalized the signals with PRAAT tools. Different form of speech signals were analyzed by added noise in original speech signals. The voice activity detection (VAD) algorithm estimate the noise spectrum of original speech signal. Spectral Subtraction technique is adopted to reduce the noise. It exploits the ability of actively unwanted signals of speech. Spectral Subtraction method has reduced noise and improved the quality of speech signals. This paper concentrates on the application of quality speech signals for speech synthesis and the results found to be satisfactory.
Keywords- Speech Enhancement, Noise Estimation, VAD, Spectral Subtraction Method, Digital Signal Processing

Introduction

In speech signal noise reduction is difficult task in signal processing. The degradation of speech by additive acoustic noise is a common phenomenon. In an office environment, the noise is created through the various components of printers, typewriters, and other speakers in the surrounding area. Noise is one kind of sound that is unexpected or undesired signals [1].

In general environments such as telephone calls, telephone booth, or some background noise in the speech signals is picked up by microphone [2]. These characteristics of the noise components in the speech signals are estimate by extracting statistical parameters of silence interval is found in the speech signals. Spectral subtraction reduced the noisy speech signals to improve quality of speech. All these speech enhancement techniques assume the speech and noise is uncorrelated. It depicted the application of VAD-algorithm used in the speech recognition, voice compression, noise estimation, and echo cancellation. The VAD algorithm includes energy level detection, zero crossing rates, periodicity, spectral energy distribution, timing, and adaptive noise modeling. However, experimental results show some noisy speech signals were affected of hearing.

In Simple, there are five major stages in Noise reduction processing:

1. Analysis Speech
2. Segmentation
3. Estimate Noise
4. Reduce Noise
5. Clean Speech

Analysis Speech is the primary stage before starting noise reduction. Initially, it obtains basic structure of speech signals. The noise speech signal will affect in speech analysis. Although some applications can use directly the speech signals, these signals are too complex and unproductive speech. Therefore, frequently need to divide signals in frame and estimation the noise, reduction of noise and reconstruct the speech signals.

VAD Algorithm

The concepts of Voice Activity Detection (VAD) is to determine whether a frame of the captured analysis speech signal represents active voice signals, inactive voice signals, or silence signals. This method includes those based on energy thresholds, pitch detection, zero-crossing rate, and periodicity measure.

Silence has the least amount of energy and representation of the background noise of the environment [3].
The VAD algorithms accuracy dramatically affects the noise suppression level and amount of speech distortion. The noise estimate in spectral subtraction uses the VAD to decide when to update the noise reference in the absence of speech. The speech signal is compared with the threshold depending on the noise level. Pitch thresholds are computed accurately value for voice active signals and noise active signals level changes considerably before the next noise level re-calibration instant. Generally, output of VAD algorithms is binary decision on a frame-by-frame basis having frame duration 20-30 msec. A segment of speech is declared to contain voice activity (VAD='1') if calculate threshold value otherwise it is declared a noise (VAD ='0'). The first step of the algorithm is to buffer the data into the kth frame x (nf, k), and converts time domain into the frequency domain.

$$X(wf,k) = FFT(x(nf,k))$$

Next, the noise spectrum for k=1 are initialized.

$$N(wf) = X(wf,k)$$

$$\mu_N = \frac{1}{Nl}\sum_{wf=0}^{Nl-1} N(wf)$$

The frequency index is wf, noise estimate is N (wf) & mean of the noise estimate is μ_N .

If Speech active value is 1 then VAD will be set as 1.
If Speech inactive value is 0 then VAD will be set as 0.

Special Subscription Technique

Speech enhancement normally assumes that the noise source is additive and not correlated with the clean speech signal. One of the most popular methods for reducing the effect of background (additive) noise is spectral subtraction. Spectral subtraction (SS) uses a Discrete Fourier Transform (DFT) to obtain estimates of the noise spectrum, which are then reduced from the spectral magnitude in those areas where noise is present. The spectrum of noise X(f) is estimated during speech inactive periods and reduced from the spectrum of the current frame N(f) resulting in an estimate of the spectrum Y(f) of the clean speech signals[4].

$$Y(f) = X(f) + N(f)$$

The noisy signal y(m), the original signal x(m), the noise n(m) and f is the frequency variable. In this technique, the incoming signal x (m) is buffered and divided into segments of N sample length. Each segment is windowed, using a Hamming window, then transformed via DFT to N spectral samples.
The windowed signal is given by;

$$y_w(m) = w(m)y(m)$$
$$y_w(m) = w(m)[x(m) + n(m)]$$
$$y_w(m) = x_w(m) + n_w(m)$$

The windowing operation, where the convolution identity of operator * denoted. The signals are windowed. We use of the subscript w for windowed signals.

$$Y_w(f) = W(f) * Y(f)$$
$$Y_w(f) = X_w(f) + N_w(f)$$

The spectral subtraction can be calculated by following equation;

$$\hat{X}_f^b = Y_f^b - \alpha(f)\overline{N_f^b}$$

Where $\hat{X}_f^b$ is an estimate of the signal magnitude spectrum to the power of b & $\overline{N_f^b}$ is the time-averaged magnitude of noise spectra to power the b. the exponent b=1, and for power spectral subtraction, b=2. The parameter $\alpha(f)$ controls the amount of noise subtracted from the noisy signal.
Discrete Fourier Transform and Inverse Discrete Fourier Transform are calculated using following equation

$$x = FFT(Y_w(f))$$
$$Y(f) = IFFT(x * q)$$

There exist many refinements of the original method that improve the quality of the enhanced speech. The mismatch between the training conditions and the testing conditions has a deep impact on the accuracy of these systems and represents a barrier for their operation in noisy environments.
Non-speech frame-dropping (FD) is also a frequently used technique in speech recognition to reduce the number of insertion errors due to the noise that can be a serious error source under high mismatch training/testing conditions [5].
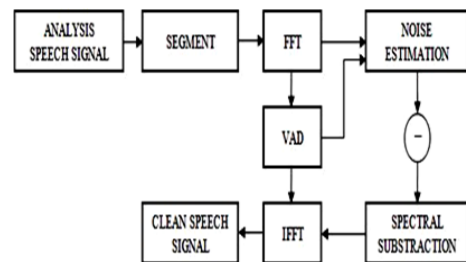

Fig. 1- Block diagram of Spectral Subtraction

Discussion and Results

Experimental Work

The experimental work is carried out with 10 Marathi numerals are recorded to spoken by 10 speakers. So, it create the sampling corpus with 37 speakers are spoken the each Marathi numerals. Table-1 shows the experimental database carried out for acoustic speech signals for analysis.

Table 1- Acoustic Analysis

Sampling frequency	22.050 (KHz)
Acoustic Parameter	FFT Spectrum (256 pts)
Frame length	30 ms
Frame Shift	10 ms
Analysis window	Hamming Window
Mixture Noise	Transient Noise, Acoustic Noise etc.
Speaker	Male (26) Female (11) Total 37

Experimental Result

The experimental result is shown in Table-2 Average Accuracy with respect type of noise such as Transient type the number of Male Speakers 26 and 11 female Speakers it having average accuracy is 88.33% and Acoustic type only 3 male Speakers it having the average accuracy is 60.35%. Table-3 has shown the result of Clean Accuracy for Transient Type Particular 1-10 Marathi digit spoken by 1 speaker. Fig. 2 Show the noise reduction with transient Noise type of one Marathi digit waveform. Fig 2.1 shows a sample of Noise Reduction with Acoustic Noise Type of original Signals with transient Noise Type .and Fig 2.2 Shows clean speech signal of one sample of Marathi digit.

Table 2- Average Accuracy with respect to Type of Noise

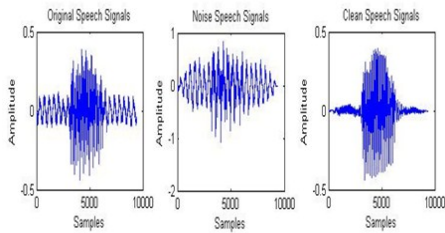| Noise Type	Number of Speakers		Average Accuracy
	Male	Female	
Transient	26	11	88.35
Acoustic	3	-	60.35

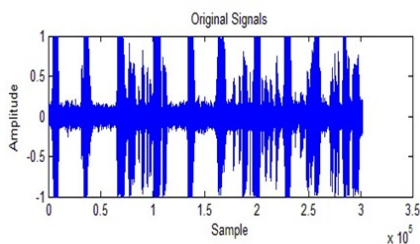**Fig. 2-** Show Noise Reduction of Transient Noise Type

Fig. 2.1- Noise Reduction of Acoustic Noise Type With Original Signals
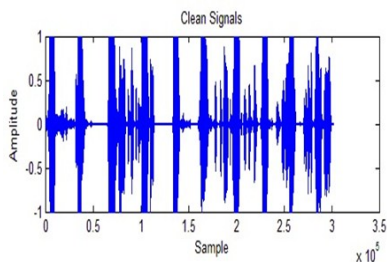

Fig. 2.2- Show Clean Signals

Table 3- Clean Accuracy for transient type-Number of Speakers 37 (Male-26 & Female-11)

English Digit	Pronunciation of Marathi Language	Marathi Digit	Clean Accuracy	
			Male	Female
1	Ek	१	90	89
2	Don	२	84.67	84.23
3	Teen	३	92.67	91.23
4	Chaar	४	93.34	93.12
5	Paach	५	81.34	82.45
6	Sahaa	६	94.34	92.78
7	Saat	७	88	85
8	Aath	८	90.67	91.23
9	Naoo	९	87.67	85.56
10	Dahaa	१०	86.34	83.45

In this paper, the noise reduction method based on spectral subtraction by using VAD algorithm is introduced. The experimental result shown in Fig. 2; Noise Reduction (Transitive and Acoustic Noise Type) effectively reduced the background noise of Marathi digit signals in comparison with commonly used spectral subtraction algorithm. The experimental result has shown the transient type of noise reduction average accuracy is 88.35% out of 37 (male and female) recorded samples of Marathi digits. The experimental result for Acoustic type is 60.35%. The details of the transient type experimental results are shown in Table 3. Accuracy calculated by frequency of Marathi 1-10 digits speech signals. It has found the clean accuracy maximum rate is to be 94.34% for Male and 93.12% for Female. Speech synthesis gets clean speech signals. The further work can be continuing for non-stationary noise such as industry and other speaker's speech signal.

References

[1] Zhixin Chen (2011) *IJACSA* 2(8).

[2] Wilfred N. Mwema and Elijah Mwangi, *A Spectral Subtraction Method for Noise Reduction in Speech Signals*.

[3] Jeffery J. Faneuff and D. Richard Brown III, *Noise Reduction and Increased VAD Accuracy Using Spectral Subtraction*.

[4] Kamil K. Wojcicki, Benjamin J. Shanon and Kuldip K. Paliwal, *Spectral Subtraction with Variance Reduced Noise Spectrum Estimates*.

[5] Ekaterina Verteletskaya and Boris Simak, *Enhanced Spectral Subtraction Method for Noise Reduction with minimal speech distortion*.

[6] Fujimoto M. and Ariki Y., *Noisy Speech Recognition Using Noise Reduction Method Based on Kalman Filter*.

[7] Kazuma Fujioka and Yoshikazu Miyanaga, *A New Noise Reduction Method of Speech Signal with Running Spectrum Filtering*.