ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology
Volume 1, Issue 5, July 2012

Determination of Noise Levels in Using AMS Features of Noisy Speech Signal and Their Comparison

Mrs. Anusha Krishna Prakash,

Abstract— **Great difficulty in recognizing speech is under a noisy background. The signal to noise ratio plays a very important role in speech recognition techniques. The signal to noise ratio is the ratio of the signal estimate to noise estimate in a particular speech signal. It has many applications in speech recognition techniques and speech pattern classification schemes. This project helps in finding the signal to noise ratio estimates between the noises and reconstructed noise free speech segment**

In this phase the amplitude modulation spectrogram features are extracted from the noisy speech segment and they are classified into noise dominated and signal dominated features using baye's classifier. The classified features are reconstructed using overlap and add method. The signal to noise ratio is calculated between the reconstructed speech segment and the noises that the speech segment is corrupted.

Index Terms— **Amplitude modulation spectrogram, signal to noise ratio, baye's classifier and overlap and add method.**

I. Introduction

Hearing impairment refers to both complete and partial loss of the ability to hear. There are two types of hearing impairment, according to which part of the ear is affected. Conductive hearing impairment is a problem in the outer or middle ear. It is often medically or surgically treatable. A common example is chronic middle ear infection. Sensorineural hearing impairment is a problem with the inner ear, or, occasionally with the hearing nerve. It is usually permanent and requires rehabilitation such as the use of a hearing aid. Hearing impairment can impose a heavy social and economic burden on individuals, families, communities and countries.

Hearing impairment in children may delay development of language and cognitive skills, which may hinder progress in school. In adults, hearing impairment often makes it difficult to obtain, perform, and keep jobs. Hearing impaired children and adults are often stigmatized and socially isolated. The poor suffer more from hearing impairment because they cannot afford the preventive and routine care to avoid hearing loss nor the hearing aids to make the disability manageable. Hearing impairment also makes it more difficult for them to escape poverty by hindering progress in school or in the workplace and by isolating them socially. For countries, the cost of special education and lost employment due to hearing impairment can burden the economy. The hearing difficulties that normal hearing listeners facing in the world are aging factor, if they get any hearing ailments they lose their hearing efficiency. Some of the hearing difficulties create problems such that the listeners lose their ability to recognize the spoken words underlying the speech. Even at noisy atmospheres even normal persons fail to recognize the speech. In this phase the amplitude modulation spectrogram features are determined for the speech signal and the SNR values between the noisy signal and the reconstructed signal.

Manuscript received June 25, 2012.

Mrs. Anusha Krishna Prakash, *Department of Bio- Medical Instrumentation Engineering, Avinashilingam Institute for Home Science and Higher Education For Women University, Coimbatore, India,*

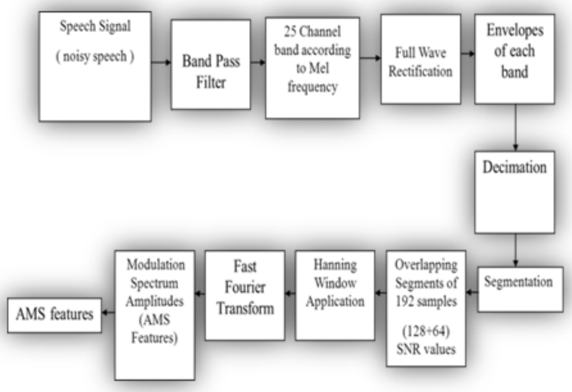

Fig 1. Block Diagram of Feature Extraction

II. Feature Extraction

The block diagram explains the procedure of feature extraction. At first a noisy speech signal and a clean speech signal are taken and they are passed through the band pass filter to get 25 channel bands according to Mel frequency. To these channel bands fast Fourier transforms are applied to get the envelopes of each band and to get the DC components of noisy speech signal and clean speech signal. Each envelope is decimated by the factor 3 and they are segmented. As a result of segmentation, we get the 128 samples and 64 overlapping samples after the application of Hann window to these samples and again applying the fast Fourier transforms we get modulation spectrum amplitudes which are the AMS features. Along with these AMS features we determine the delta feature values and we combine with those feature vectors to get the feature vectors which are 45 in number. By classifying these features using the Bayesian classifiers and the SNR is estimated between the two sets of speech segments i.e., SNR is estimated between the clean speech signal and noisy speech segment and the SNR is calculated between the clean speech segment and reconstructed speech segment.

Band pass filter is generally used to allow the signals falling within some band of frequencies. This filter will segregate the signal into 25 channel bands according to mel frequency spacing. Each band is full wave rectified to change it to DC components such that the DC components containing the noise can be retained or eliminated. This FFT is used to construct the envelope in each band. This is then decimated by the factor 3. This decimation is to reduce the bandwidth and to do reliable filtering in modulation domain. While applying the fast Fourier transforms we use various types windowing function here are some of the windows given below. The windowing technique used in this project is Hanning window. The Hann and Hamming windows, both of which are in the family known as "raised cosine" or "generalized Hamming" windows, are respectively named after Julius von Hann and Richard Hamming. The term "Hanning window" is sometimes used to refer to the Hann window. The reason for using this window is that when comparing to other windowing techniques it has better frequency resolution and spectral leakage and amplitude accuracy. Fast Fourier transforms are applied to each segment after padding it with the zeros. The main use of the fast Fourier transform is that it can calculate the modulation spectrum of each channel. To each band 15 triangular shaped windows are multiplied to get modulation spectral amplitudes. These amplitudes are nothing but the amplitude modulation spectrogram (AMS) features. These features have a frequency resolution of 15.6Hz.

.

III. Classification and Reconstruction

A. Bayesian classification

Classification of amplitude modulation spectrogram (AMS) features is done using Bayesian classifiers. Bayesian classifier involves classification of the features into either noise dominated features or signal dominated features. This is done by comparing the probabilities $P(\lambda_0|A_y(\tau,k))$ and $P(\lambda_1 | A_y(\tau,k))$.

Where $P(\lambda_0|A_y(\tau,k))$ is the noise dominated amplitude modulation spectrogram features. These features are not considered for reconstruction since they are the features of interest.

$P(\lambda_1|A_y(\tau,k))$ is the signal dominated amplitude modulation spectrogram features and these are features of interest and they are used for the reconstruction.

The probability $P(\lambda_0|A_y(\tau,k))$ is found by using the equation

$$P(\lambda_0|\mathbf{A}_Y(\tau,k)) = \frac{P(\lambda_0,\mathbf{A}_Y(\tau,k))}{P(\mathbf{A}_Y(\tau,k))} = \frac{P(\lambda_0^0)P(\mathbf{A}_Y(\tau,k)|\lambda_0^0) + P(\lambda_0^1)P(\mathbf{A}_Y(\tau,k)|\lambda_0^1)}{P(\mathbf{A}_Y(\tau,k))}. \quad - (1)$$

Then the $P(\lambda_1|A_y(\tau,k))$ is calculated in a similar way. These probabilistic features are reconstructed using overlap and add method.

IV. Overlap and Add Method of Reconstruction

Overlap and add method of reconstruction of the speech signal is a common method which is also used for reconstruction. It has many applications in reconstruction of various signal. It mainly involves the time scale modification of the signal. It is a common time domain technique. This method first collects the similar segments of the speech signal which is devoid of noise i.e., it collects the DC components of speech signal where the signal dominated components are taken. These frames are taken as overlapping segment which has 0.4 millisecond duration. These overlapping frames are added together to form the reconstructed noise free speech signal.

V. SNR Estimation Between Clean Speech Segment and Noisy Speech Segment

Signal-to-noise ratio is defined as the power ratio between a signal and the background noise. The signal to noise ratio is an important feature in determining the quality of the audio data. This is very important in speech recognition process since the SNR values strongly influence the speech recognition performance. In most of the applications the SNR values are not easily determined since the noise estimate is not known. This introduces the short term analysis of speech signal given the prior knowledge of speech data to characterize the noise and speech data.

$SNR = 10\log_{10}$(Signal Energy/ Noise Energy) – (2)

Here the signal to noise ratio is calculated between the noise signal and reconstructed signal. This is calculated using the formula,

$$SNR = \log_{10}(F_o - F)^2 / \log_{10} F_o^2 \quad - (3)$$

Where F_o – reconstructed speech segment and F – noise dominated speech segment. Here each speech signal is of 2 seconds duration. Five types of speech segments are used. They are bigtips, butter, draw, peaches and scholar speech segments. Each speech segment is corrupted with six different types of noises. The noises used for corrupting the speech segments are, burst_r1, f16_r1, factory_r1, pink_r1, Volvo_r1 and white_r1.

ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology
Volume 1, Issue 5, July 2012

Here the signal to noise ratio is calculated between the clean speech segment and reconstructed speech segment. This is calculated using the formula,

$$SNR = \log_{10} (F_0 - F_1)^2 / \log_{10} F_0^2 \qquad -(4)$$

Where F_o – reconstructed speech segment and F_1 – original clean speech segment.

VI. Results and Discussions

The tables given below portray the two sets of signal to noise ratio estimates between the clear speech segment and reconstructed speech and between noisy speech and reconstructed speech segment for the corresponding speech segments.

BIGTIPS- SPEECH SEGMENT		
Noisy speech signal	SNR values between the original and reconstructed signal	SNR values between the original and noisy signal
Burst_r1	0.8790	20.7209
F16_r1	1.9107	5.3630
Factory_r1	2.0050	5.9266
Pink_r1	1.1534	5.0261
Volvo_r1	2.4397	8.8841
White_r1	1.1603	5.0470

Table – 1 Signal to noise ratio for Bigtips type speech

BUTTER – SPEECH SEGMENT		
Noisy speech signal	SNR values between the original and reconstructed signal	SNR values between the original and noisy signal
Burst-r1	3.1422	26.0417
F16_r1	3.2954	6.5271
Factory_r 1	3.2829	7.0432
Pink_r1	3.5784	5.8888
Volvo_r1	3.2512	10.3689
White_r1	2.2031	6.4620

Table – 2 Signal to noise ratio for Butter type speech segment

PEACHES – SPEECH SEGMENT		
Noisy speech signal	SNR values between the original and reconstructed signal	SNR values between the original and noisy signal
Burst-r1	1.4240	21.4579
F16_r1	1.4107	8.8095
Factory_r 1	1.7055	9.7530
Pink_r1	1.1729	8.4101
Volvo_r1	1.2274	13.0101
White_r1	1.2735	7.8040

Table – 3 Signal to noise ratio for Peaches type speech segment

DRAW – SPEECH SEGMENT		
Noisy speech signal	SNR values between the original and reconstructed signal	SNR values between the original and noisy signal
Burst-r1	0.3638	21.6511
F16_r1	0.4792	9.0468
Factory_r1	0.4295	10.0269
Pink_r1	0.4656	8.4262
Volvo_r1	0.3452	13.2181
White_r1	0.2285	8.7704

Table – 4 Signal to noise ratio for Draw type speech segment

VII. Future Scope

From these signal to noise estimates it is evident that noise ratio has reduced when comparing the SNRs between the original clean segment and noisy speech segment and SNRs between the original clean speech segment and reconstructed speech segment. Using this method the speech segments are segregated into similar patterns. These patterns are classified according to the degree of similarity. This method is also used in various speech recognition techniques where the noise estimates are finalized with the degree of the signal to noise ratio. Drawback of this project is continuous training with various noisy speech signal and their corresponding speech segments are needed.

REFERENCES

[1] Gibak Kim, Yang Lu, Yi Hu, and Philipos C. Loizou 'An algorithm that improves the speech intelligibility for normal human listeners' 2009 Acoustical Society of America. DOI: 10.1121/1.3184603 PACS numbers: 43.72.Ar, 43.72.Dv MSS Pages: 1486–1494

[2] Jürgen Tchorz and Birger Kollmeier 'SNR estimation based on amplitude modulation analysis with applications to noise suppression' IEEE transactions on speech and audio processing, vol. 11, no. 3, may 2003.

[3] Jorn Anemuller, Denny Schmidt, Jorg-Hendrik Bach 'Detection of Speech Embedded in Real Acoustic Background Based on Amplitude Modulation Spectrogram Features' Accepted after peer review of full paper Copyright © 2008 ISCA pages 2582-2585.

[4] Jorg-Hendrik Bach, Birger Kollmeier, Jorn Anemuller 'modulation-based detection of speech in real background noise: Generalization to novel background classes' 978-1-4244-4296-6/10/$25.00 ©2010 IEEE pages(41 – 44).

[5] Neri merhav, member IEEE and Yariv . E. Phraim, Senior member IEEE ' A Bayesian classification approach with application to speech recognition' IEEE Transaction on Signal Processing Vol. 39, No. 10, Oct 1991 Pages(2157 – 2166).

[6] Peter H. W. Wong, Oscar C. Au, Justy W.C. Wong, William H. B. Lau 'On improving the intelligibility of synchronized overlap and add at low TSM factor'.

[7] Xin Ma, Weidong Zhou 'AMS based spectrum subtraction algorithm with confidence interval test' Yi Peng, Xiaohong Weng (Eds.): APCMBE 2008, IFMBE Proceedings 19, pp. 389–391, 2008.