Suppression of Acoustic Noise In Speech Processing

Miss. Dhanwardhini Shivaji Bhalerao[1], Miss. Rutuja sharad Bhamre[2]
[1, 2] Dept of Electronics And Telecommunication Engineering
[1, 2] IOIT, Pune-33.

Abstract- *Background noise acoustically added to speech can degrade the performance of digital voice processors used for applications such as speech compression, recognition, and authentication [11], [2].Digital voice systems will be used in a variety of environments and their performance must be maintained at a level near that measured using noise-free input speech. To ensure continued reliability, the effects of back-ground noise can be reduced by using noise-cancelling micro-voice phones, internal modification of the processor algorithms to explicitly compensate for signal contamination, or pre-processor noise reduction.The objectives of this effort were to develop a noise suppression technique, implement a computationally efficient algorithm, and test its performance in actual noise environments.The approach used was to estimate the magnitude frequency spectrum of the underlying clean speech by subtracting the noise magnitude spectrum from the noisy speech spectrum.*

I. INTRODUCTION

Suppression of acoustic noise in speech processing is the primary component or subsystem of a human-computer interaction (HCI) system that requires human speech as decoded text input. ASR is comprised of two major components, namely, acoustic model and language model. The acoustic model models the pronunciation of a given word, whereas the language model predicts the likelihood of a given word sequence appearing in a language.

The components of an acoustic model can be the speech signal features and a pattern matching technique for a given word or phone. The term, „phone" represents a basic unit of speech. A word may consist of one or more phones. The most commonly used features of ASR is Mel-Frequency Cepstral Coefficients (MFCC) and Perceptual Linear Predictive (PLP) Coefficients, whereas Hidden Markov Model (HMM) and neural network are the most commonly used pattern matching techniques. HMM understand the sequential nature of speech signal and model the output probability distribution as well as the state transition probability. While recognizing the speech, various words are hypothesized against the acquired signal. HMM performs matching by determining the likelihood of a given word.

The likelihood of a word is estimated based on the combination of likelihood of all the phones associated with the word. Earlier, maximum likelihood (ML) estimation has been widely exploited to train HMM. However, discriminative training has been found as more promising than ML in the later era.

1.1 Basic of Acoustic Noise:

In audio engineering, electronics, physics, and many other fields, the color of noise refers to the power spectrum of a noise signal (a signal produced by a stochastic process). Different colors of noise have significantly different properties: for example, as audio signals they will sound different to human ears, and as images they will have a visibly different texture. Therefore, each application typically requires noise of a specific color. This sense of 'color' for noise signals is similar to the concept of timbre in music (which is also called "tone color; however the latter is almost always used for sound, and may consider very detailed features of the spectrum. The practice of naming kinds of noise after colors started with white noise, a signal whose spectrum has equal power within any equal interval of frequencies. That name was given by analogy with white light, which was (incorrectly) assumed to have such a flat power spectrum over the visible range. Other color names, like pink, red, and blue were then given to noise with other spectral profiles, often (but not always) in reference to the color of light with similar spectra. Some of those names have standard definitions in certain disciplines, while others are very informal and poorly defined. Many of these definitions assume a signal with components at all frequencies, with a power spectral density per unit of bandwidth proportional to $1/f\ \beta$ and hence they are examples of power-law noise. For instance, the spectral density of white noise is flat ($\beta = 0$), while flicker or pink noise has $\beta = 1$, and Brownian noise has $\beta = 2$.Simulated power spectral densities as a function of frequency for various colors of noise (violet, blue, white, pink, brown/red). The power spectral densities are arbitrarily normalized such that the value of the spectra are approximately equivalent near 1 kHz. Note the slope of the power spectral density for each spectrum provides the context for the respective electromagnetic/color analogy.

SPECTRAL SUBTRACTION

Spectral subtraction is a method for restoration of the power spectrum or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum. The noise spectrum is usually estimated, and updated, from the periods when the signal is absent and only the noise is present. The assumption is that the noise is a stationary or a slowly varying process, and that the noise spectrum does not change significantly in-between the update periods. For restoration of time-domain signals, an estimate of the instantaneous magnitude spectrum is combined with the phase of the noisy signal, and then transformed via an inverse discrete Fourier transform to the time domain. In terms of computational complexity, spectral subtraction is relatively inexpensive. However, owing to random variations of noise, spectral subtraction can result in negative estimates of the short-time magnitude or power spectrum. The magnitude and power spectrum are non-negative variables, and any negative estimates of these variables should be mapped into non-negative values. This non-linear rectification process distorts the distribution of the restored signal. The processing distortion becomes more noticeable as the signal-to-noise ratio decreases. In applications where, in addition to the noisy signal, the noise is accessible on a separate channel, it may be possible to retrieve the signal by subtracting an estimate of the noise from the noisy signal. However, in many applications, such as at the receiver of a noisy communication channel, the only signal that is available is the noisy signal. In these situations, it is not possible to cancel out the random noise, but it may be possible to reduce the average effects of the noise on the signal spectrum. The effect of additive noise on the magnitude spectrum of a signal is to increase the mean and the variance of the spectrum as illustrated in Figure Theincrease in the variance of the signal spectrum results from the random fluctuations of the noise, and cannot be cancelled out. Theincrease in the mean of the signal spectrum can be removed by subtraction of an estimate of the mean of the noise spectrum from the noisy signal spectrum

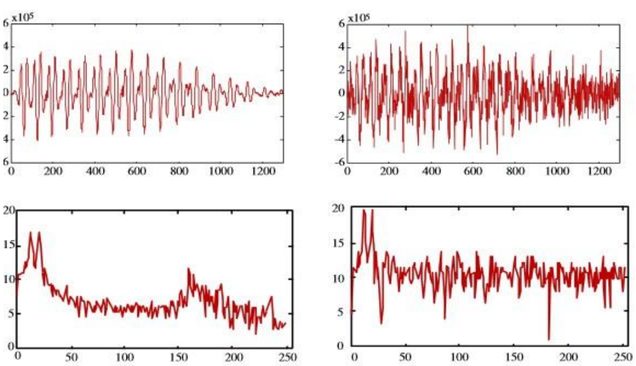


Figure 1: Illustrations of the effect of noise on a signal in the time and the frequency domain

Y (m) =x (m) + n (m)

Where y(m),x(m) and n(m) are the signal, the additive noise and the noisy signal respectively, and is the discrete time index. In the frequency domain, the noisy signal model of Equation is expressed as

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

Where Y (f), (f) and N (f) are the Fourier transforms of the noisy signal y(m), the original signalX (m) and the noise n (m) respectively, and f is the frequency variable. In spectral subtraction, the incoming signal x(m) is buffered and divided into segments of N samples length. Each segment is windowed, using a Hanning or a Hamming window, and then transformed via discrete Fourier transform (DFT) to N spectral samples. The windows alleviate the effects of the discontinuities at the endpoints of each segment.

The windowed signal is given by

yw(m)=w(m)y(m)
=w(m)[x(m)+n(m)]
=xw(m)+nw(m)

The windowing operation can be expressed in the frequency domain as

Yw(f)=W(f)*Y(f)
=Xm(f)+Nw(f)

Where the operator * denotes convolution. Figure 14 illustrates a block diagramconfiguration of the spectral subtraction method. A more detailed implementation is described in Section The equation describing spectral subtraction may be expressed as in this Equation controls the amount of noise subtracted from the noisy

$$\left|\hat{X}(f)\right|^{b} = \left|Y(f)\right|^{b} - \alpha\overline{\left|N(f)\right|^{b}}$$

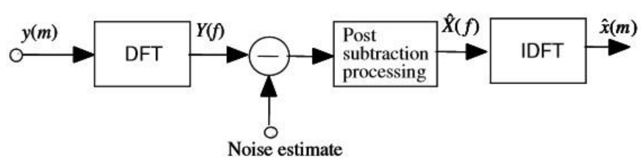


Figure 13: A block diagram illustration of spectral subtraction

a discrete Fourier transformer (DFT) for transforming the time domain signal to the frequency domain; the DFT is followed by a magnitude operator; a post-processor for removing the processing distortions introduced by spectralsubtraction. an inverse discrete Fourier transform (IDFT) for transforming the processed signal to the time domain.

II. LITERATURE SURVEY

1. STEVEN F BOLL Suppression of acoustic noise in speech using spectral subtraction.

This paper describes the A stand-alone noise suppression algorithm is presented for reducing the spectral effects of acoustically added noise in speech. Effective performance of digital speech processors operating in practical Environments may require suppression of noise from the digital waveform. Spectral subtraction offers a computationally efficient, processor independent approach to effective digital speech analysis. The method, requiring about the same computation as high-speed convolution, suppressesStationary noise from speech by subtracting the spectral noise bias calculated during non-speech activity. Secondary procedures are then applied to attenuate the residual noise left after subtraction. Since the algorithm re-synthesizes a speech waveform, it can be used as apre-processor to narrow-band voice communications systems, speech recognition systems, or speaker authentication systems.

2.Priyanka Wani1,U.G.Patil2, Dr.D.S.Bormane3. Dr.S.D.Shirbahadurkar4 Automatic Speech recognition of isolated words in Hindi language.

In this paper, speech recognition is a broad subject as speech natural way of communication.they acoustic and language model for this system are available but mostly in English language .in India there are so many people who can‟t understand or speak English so the speech recognition system in English language is of no use for these people. Automatic speech recognition is also called as computer speechrecognition. the main goal ASR system understand a voice by computer or microphone and converts it into the text to perform required task.

3.Karthikeyan Umapathy Audio, Speech, and Language Processing

The objective of automatic speaker recognition is to extract, characterize and recognize the information about speaker identity. A direct analysis and synthesizing the complex voice signal is due to too much information contained in the signal. Therefore the digital signal processes such as Feature Extraction and Feature Matching are introduced to represent the voice signal. Several methods such as Liner Predictive Coding (LPC), Hidden Markov Model (HMM).

III. METHODOLOGIES

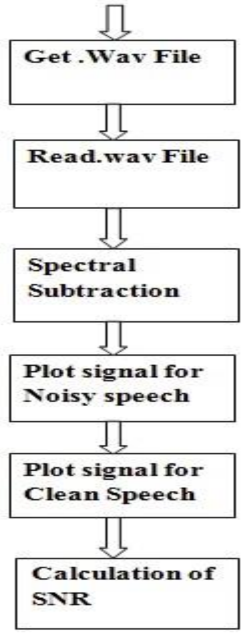
3.1 AIM:

Suppression of Acoustic Noise in Speech Processing

3.2 Objective:

- To develop a noise suppression technique.
- To implement a computationally efficient algorithm.
- Test its performance in actual noise environments.

3.3 Block diagram



3.3.1 .Wav File:

Wav is a file extension for an audio file format created by Microsoft. The WAV file hasbecome a standard PC audio file format for everything from system and game sounds to CD-quality audio.

Also referred to as pulse code modulation (PCM) or waveform audio, a WAV file isuncompressed audio. A Wave file also stores information about the file's number of tracks, sample rate, bit depth, and whether it's mono or stereo.

3.3.2 Spectral Subtraction:

Spectral amplitude estimation method to restore the signal degraded by additive noise .phase distortion can be ignored since human ear is insensitive to phase .restoring the signal by subtraction and estimate of the noise spectrum. from the noisy signal spectrum.

3.3.3 Noisy Speech signal:

In signal processing, noise is a general term for unwanted (and, in general, unknown) modifications that a signal may suffer during capture, storage, transmission, processing, or conversion.

Sometimes the word is also used to mean signals that are random (unpredictable) and carry no useful information; even if they are not interfering with other signals or may have been introduced intentionally, as in comfort noise.

3.3.4 Clean Speech Signal:

Clean speech signal or Speech enhancement aims to improve speech quality by using various algorithms. The objective of enhancement is improvement in intelligibility and/or overall perceptual quality of degraded speech signal using audio signal processing techniques.

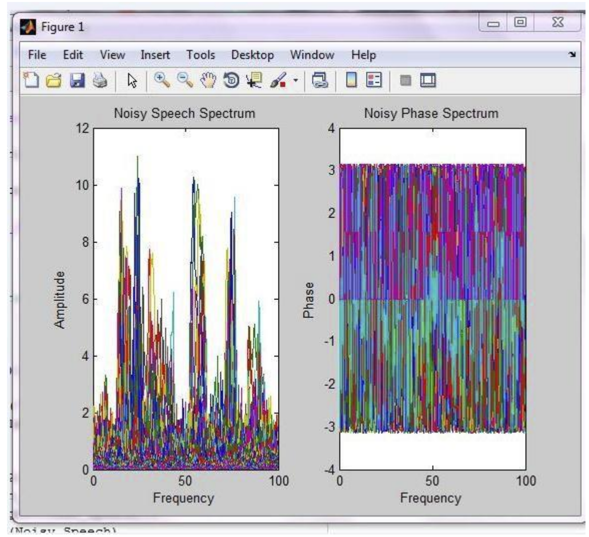
Enhancing of speech degraded by noise, or noise reduction, is the most important field of speech enhancement, and used for many applications such as mobile phones, VoIP, teleconferencing systems, speech recognition, and hearing aid

3.3.5 Calculation of SNR:

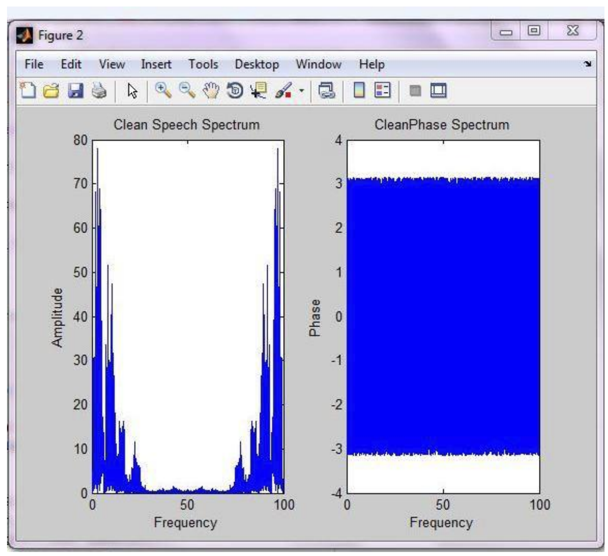
Signal-to-noise ratio (abbreviated SNR or S/N) is a measure used in science and engineering that compares the level of a desired signal to the level of background noise. It is defined as the ratio of signal power to the noise power, often expressed in decibels. A ratio higher than 1:1 (greater than 0 dB) indicates more signal than noise. While SNR is commonly quoted for electrical signals, it can be applied to any form of signal (such as isotope levels in an ice core or biochemical signaling between cells).The signal-to-noise ratio, the bandwidth, and thechannel capacity of a communication channel are connected by the Shannon-Hartleytheorem. In signal to noise ratio if the SNR is maximum then noise less and if the SNR is minimum then noise is more in speech signal.

IV. TESTING AND RESULT

4.1 Noisy Speech Signal:



4.2 Clean Speech Signal:



V. FUTURE SCOPE

Proposed work is one of the application of automatic speech recognition . Speech is natural vocalized and primary means of communication. Speech is easy hand free, fast and do not require any technical knowledge. Communicating with computer using speech in simple and comfortable way for human being rather than using the other medium such as keyboard and mouse , as it is requires certain skill and good coordination in hand eye .Physically challenged people or blind people find it difficult to use computer speech recognition solves all these issues.

VI. CONCLUSION

Overall, spectral subtraction seems to let us achieve better acoustic noise reduction, and does provide more parameters we can tune to achieve a better performance. To some Extend, this explains why spectral subtraction remained so popular .spectral subtraction technique seems to allow us to remove a fair amount of noise from speech signal.

REFERENCES

[1] Anuj Mohamed, K. N. Ramachandran Nair-Continuous Malayalam speech recognition using Hidden Markov Models. Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India, September 16-17, 2010, Tamilnadu, India; 01/2010.

[2] N. Rajput M. Kumar and A. Verma. A large-vocabulary continuous speech recognition system for Hindi. IBM Journal for Research and Development 2004.

[3] Kumar, K. and Aggarwal, R. K., "H`indi Speech Recognition System Using HTK", International Journal of Computing and Business Research, ISSN (Online): 2229-6166, Volume 2 Issue 2 May 2011.

[4] Aggarwal, R.K. and Dave, M., -Using Gaussian Mixtures for Hindi Speech Recognition System‖, International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 2, No. 4, December 2011.

[5] Mishra, A. N., Biswas, A., Chandra, M., Sharan, S. N., "Robust Hindi connected digits recognition", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 4, No. 2, June, 2011.

[6] Sivaraman. G.; Samudravijaya, K., Hindi Speech Recognition and Online Speaker Adaptation, International Conference on Technology Systems and Management: ICTSM-
2011, IJCA.

[7] Kumar, R., Comparison of HMM and DTW for Isolated Word Recognition System for Punjabi Language, International Journal of Soft Computing 5(3):88-92, 2010.

[8] R. Thangarajan, A.M. Natarajan, M. Selvam, "Word and Triphone Based Approaches in Continuous Speech Recognition for Tamil Language", Wseas Transactions on Signal Processing, Issue 3, Volume 4, March 2008.

[9] V.S.Dharun, M.Karnan, 2012, Voice and Speech Recognition for Tamil Words and Numerals, IJMER, Vol. 2, 5, pp 3406-3414.

[10] C.Vimala, M.Krishnaveni , 2012, Continuous Speech Recognition system for Tamil language using monophone-based Hidden Markov Model , Proceedings of the Second International Conference on Computational Science, Engineering and Information Technology CCSEIT '12 , pp 227-231.