Noise Estimation and Noise Removal Techniques for Speech Recognition in Adverse Environment

Urmila Shrawankar, Vilas Thakare

Noise Estimation and Noise Removal Techniques for Speech Recognition in Adverse Environment

Urmila Shrawankar[1], Dr. Vilas Thakare[2]

[1]IEEE Student Member & Research Scholar, (CSE), SGB Amravati University, INDIA
[2]Professor & Head, PG Dept. of Computer Science, SGB Amravati University, INDIA
[1]G H Raisoni College of Engg., Nagpur, INDIA
[1]urmilas@rediffmail.com

Abstract : Noise is ubiquitous in almost all acoustic environments. The speech signal, that is recorded by a microphone is generally infected by noise originating from various sources. Such contamination can change the characteristics of the speech signals and degrade the speech quality and intelligibility, thereby causing significant harm to human-to-machine communication systems.
Noise detection and reduction for speech applications is often formulated as a digital filtering problem, where the clean speech estimation is obtained by passing the noisy speech through a linear filter. With such a formulation, the core issue of noise reduction becomes how to design an optimal filter that can significantly suppress noise without noticeable speech distortion.
This paper focuses on voice activity detection, noise estimation, removal techniques and an optimal filter.

Keywords: Additive Noise, Noise detection, Noise removal, Noise filters, Voice Activity Detector (VAD)

1. Introduction

Noise estimation and reduction [6] is a very challenging problem. In addition, noise characteristics may vary in time. It is therefore very difficult to develop a versatile algorithm that works in diversified environments.
Although many different transforms are available, noise reduction [1] have been focused only on the Fourier, Karhunen–Loeve, cosine, Hadamard transforms. The advantage of the generalized transform domain is the different transforms can be used to replace each other without change the algorithm formulation. The following steps will help to use generalized transform domain; i. Reformulate the noise reduction problem into a more generalized transform domain, where any unitary matrix can be used to serve as a transform and ii. Design different optimal and suboptimal filters in the generalized transform domain.
The points to be considered in signal de-noising applications that are i. Eliminating noise from signal to improve the SNR and ii. Preserving the shape and characteristics of the original signal. An approach is discussed in this paper, to remove the additive noise [2] from corrupted speech signal to make speech front-ends immune to additive noise. We address two problems, i.e., noise estimation and noise removal.

2. Voice Activity Detector (VAD)

VADs are widely evaluated in terms of the ability to discriminate between speech and pause periods at different SNR levels of 20dB, 15dB, 10dB, 5dB, 0dB and -5dB. These noisy signals have been recorded at different places. Detection performance as a function of the SNR [7] was assessed in terms of the non-speech hit-rate (HR0) and the speech hit-rate (HR1). Most of the VAD algorithms [4] fail when the noise level increases and the noise completely mask the speech signal. A VAD module is used in the speech recognition systems within the feature extraction process.

The different approaches of VAD include: Full-band and sub-band energies (Woo 2000), Spectrum divergence measures between speech and background noise (Marzinzik & Kollmeier 2002), Pitch estimation (Tucker 1992), Zero crossing rate (Rabiner 1975), and higher-order statistics (Nemer 2001; Ramirez 2006a; Gorriz., 2006a; Ramirez 2007).

Most of the VAD methods are based on the current observations and do not consider contextual information. However, using long-term speech information (Ramirez2004a; Ramirez 2005a) has shown improvement for detecting speech presence in high noise environment. Some robust VAD algorithms that yield high Speech/non-speech discrimination in noisy environments include i. Long-term spectral divergence; the speech/non-speech detection algorithm (Ramírez 2004a) ii. Multiple observation likelihood ratio tests; An improvement over the LRT (Sohn 1999 and Ramírez 2005b) and iii. Order statistics filters.

3. Noise Estimation Algorithms

A noise-estimation algorithm [14] is proposed for highly non-stationary noise environments. The performance of speech-enhancement algorithms as it is needed to evaluate, i. The Wiener algorithms (Lim & Oppenheim 1978), ii. Estimate the a priori SNR in the MMSE algorithms (Ephraim & Malah 1984) iii. Estimate the noise covariance matrix in the subspace algorithms (Ephraim & Van Trees 1993).

The noise estimation can have a major impact on the quality of the enhanced signal i.e. i. If the noise estimate is too low, annoying residual noise will be audible and ii. If the noise estimate is too high, speech will be distorted resulting possibly in eligibility loss. The simplest approach is to estimate and update the noise spectrum during the silent (pauses) segments of the signal using a voice-activity detection (VAD) [4]. An approach might work satisfactorily in stationary noise, it will not work well in more realistic environments where the spectral characteristics of the noise might be changing constantly. Hence there is a need to update the noise spectrum continuously over time and this can be done using noise-estimation algorithms.

Several noise-estimation algorithms are available like, Doblinger 1995; Hirsch & Ehrlicher 1995; Kim 1998; Malah 1999; Stahl 2000; Martin 2001; Ris & Dupont 2001 Afify & Sioham 2001; Cohen 2002; Yao & Nakamura 2002; Cohen 2003; Lin 2003; Deng 2003; Rangachari, 2004;

Noise estimation algorithms consider the following aspects: i. Update of the noise estimate without explicit voice activity decision, and ii. Estimate of speech-presence

probability exploiting the correlation of power spectral components in neighboring frames.

Noise-Estimation algorithm follows four steps; i. Tracking the minimum of noisy speech methods, ii. Checking speech-presence probability iii. Computing frequency-dependent smoothing constants and iv. Update of noise spectrum estimate

4. Noise Reduction Techniques

The noise is classify into following category like, adaptive, additive, additive random, airport, background, car, Cross-Noise, exhibition hall, factory, multi-talker babble, musical, Natural, non-stationary babble, office, quantile-based, restaurant, street, suburban train, ambient, random, train-station, white Gaussian etc. Noise is mainly dividing into four categories: Additive noise, Interference, Reverberation and Echo. These four types of noise has led to the developments of four broad classes of acoustic signal processing techniques include, Noise reduction/Speech enhancement, Source separation, speech dereverberation and Echo cancellation/Suppression. The scope of this paper limited to noise reduction techniques only. Noise reduction techniques depending on the domain of analyses like Time, Frequency or Time-Frequency/Time-Scale.

4.1 Noise Reduction Algorithms

The Noise reduction methods [13, 16] are classified into four classes of algorithms: Spectral Subtractive, Subspace, Statistical-model based and Wiener-type. Some popular Noise reduction algorithms are, The log minimum mean square error logMMSE (Ephraim & Malah 1985), The traditional Wiener (Scalart & Filho 1996), The spectral subtraction based on reduced-delay convolution (Gustafsson 2001), The exception of the logMMSE-SPU (Cohen & Berdugo 2002), The logMMSE with speech-presence uncertainty (Cohen & Berdugo 2002), The multiband spectral-subtractive (Kamath & Loizou 2002), The generalized subspace approach (Hu & Loizou 2003), The perceptually-based subspace approach (Jabloun & Champagne 2003), The Wiener filtering based on wavelet-thresholded multitaper spectra (Hu & Loizou 2004), Least-Mean-Square (LMS), Adaptive noise cancellation (ANC) [3], Normalized(N) LMS, Modified(M)-NLMS, Error nonlinearity (EN)-LMS, Normalized data nonlinearity (NDN)-LMS adaptation etc.

4.2 Fusion Techniques for Noise Reduction

4.2.1 The Fusion of Independent Component Analysis (ICA) and Wiener filter

The fusion uses following steps: i. ICA [10] is applied to a large ensemble of clean speech training frames to reveal their underlying statistically independent basis ii. The distribution of the ICA transformed data is also estimated in the training part. It is required for computing the covariance matrix of the ICA transformed speech data

used in the Wiener filter iii. Then a Wiener filter is applied to estimate the clean speech from the received noisy speech iv. The Wiener filter minimizes the mean-square error between the estimated signal and the clean speech signal in ICA domain v. An inverse transformation from ICA domain back to time domain reconstructs the enhanced signal. vi. The evaluation is performed with respect to four objective quality measure criteria. The properties of the two techniques will yield higher noise suppression capability and lower distortion by combining them.

4.2.2 Recursive Least Squares (RLS) algorithm : Fusion of DTW and HMM

Recursive Least Squares (RLS) algorithm is used to improve the presence of speech in a background noise [11]. Fusion pattern recognition is used such as with Dynamic Time Warping (DTW) and Hidden Markov Model (HMM). There are a few types of fusion in speech recognition amongst them are HMM and Artificial Neural Network (ANN) [10] and HMM and Bayesian Network (BN) [11]. The fusion technique can be used to fuse the pattern recognition outputs of DTW and HMM.

5. Experimental Steps for Implementing RLS Algorithm

- Recording speech, WAV file was recorded from different speakers
- RLS : The RLS [8] was used in preprocessing for noise cancellation
- End point detecting: two basic parameters are used: Zero Crossing Rate (ZCR) and short time energy [11].
- Framing, Normalization, Filtering
- MFCC : Mel Frequency Cepstral Coefficient (MFCC) is chosen as the feature extraction method.
- Weighting signal, Time normalization, Vector Quantization (VQ) and labeling.
- Then HMM is used to calculate the reference patterns and DTW is used to normalize the training data with the reference patterns
- Fusion HMM and DTW**:**
  - DTW measures the distance between recorded speech and a template.
  - Distance of the signals is computed at each instant along the warping function.
  - HMM trains cluster and iteratively moves between clusters based on their likelihoods given by the various models.

As a result, this algorithm performs almost perfect segmentation for recoded voice, recoding is done at noisy places, segmentation problem happens because in some cases the algorithm produces different values caused by background noise. This causes the cut off for silence to be raised as it may not be quite zero due to noise being interpreted as speech. On the other hand for clean speech both zero crossing rate and short term energy should be zero for silent regions.

6. Comparative Study of Various Speech Enhancement Algorithms

Total thirteen methods encompassing four classes of algorithms [17], that are, three spectral subtractive, Two subspace, Three Wiener-type and Five statistical-model based. The noise, consider at two levels of SNR (0 dB, 5 dB, 10 dB and 15 dB)

6.1. Intelligibility comparison among algorithms [16]

At 5 dB SNR : KLT and Wiener-as algorithms performed equally well in all conditions, followed by the logMMSE and MB algorithms. pKLT, RDC, logMMSE-SPU and WavThr algorithms performed poorly.

At 0 dB SNR : Wiener-as and logMMSE algorithms performed equally well in most conditions, followed by the MB and WavThr algorithms. The KLT algorithm performed poorly except in the babble condition in which it performed the best among all algorithms. Considering all conditions, the Wiener-as algorithm performed consistently well for all conditions, followed by the logMMSE algorithms which performed well in six of the eight noise conditions, followed by the KLT and MB algorithms which performed well in five conditions.

6.2. Intelligibility comparison against noisy speech

The Wiener-as algorithm maintained speech intelligibility in six of the eight noise conditions tested, and improved intelligibility in 5 dB car noise. Good performance was followed by the KLT, logMMSE and MB algorithms which maintained speech intelligibility in six conditions. All algorithms produced a decrement in intelligibility in train noise at 0 dB SNR. The pKLT and RDC algorithms significantly reduced the intelligibility of speech in most conditions.

6.3. Consonant intelligibility comparison among algorithms

pKLT and RDC, most algorithms performed equally well. A similar pattern was also observed at 0 dB SNR. The KLT, logMMSE, MB and Wiener-as algorithms performed equally well in most conditions. The logMMSESPU performed well in most conditions except in car noise. Overall, the Wiener-type algorithms Wiener-as and WavThr and the KLT algorithm performed consistently well in all conditions, followed by the logMMSE and MB algorithms. The RDC and pKLT algorithms performed poorly relative to the other algorithms

6.4. The following algorithms performed equally well across all conditions:

MMSE-SPU, logMMSE, logMMSE-ne, pMMSE and MB. The Wiener-as method also performed well in five of the eight conditions

6.5. The following algorithms performed the best, in terms of yielding the lowest speech distortion, across all conditions:

MMSE-SPU, logMMSE, logMMSE-ne, pMMSE, MB and Wiener-as. The KLT, RDC and WT algorithms also performed well in a few isolated conditions. The pKLT method also performed well in five of the eight conditions. The KLT, RDC, RDC-ne, Wiener-as and AudSup algorithms performed well in a few isolated conditions

6.6. Comparisons in reference to noisy speech :

The algorithms MMSE-SPU, log-MMSE, logMMSE-ne, and pMMSE improved significantly the overall speech quality but only in a few isolated conditions. The algorithms MMSE-SPU, log-MMSE, logMMSE-ne, pMMSE, MB and Wiener-as performed the best in all conditions. The algorithms WT, RDC and KLT also performed well in a few isolated conditions. The algorithms MMSE-SPU, log-MMSE, logMMSE-ne, log-MMSE-SPU and pMMSE lowered significantly noise distortion for most conditions. The MB, pKLT and Aud-Sup also lowered noise distortion in a few conditions.

6.7. In terms of overall quality and speech distortion, the following algorithms performed the best:
MMSESPU, logMMSE, logMMSE-ne, pMMSE and MB. The Wiener-as method also performed well in some conditions. The subspace algorithms performed poorly.

9. Conclusion

The optimal filters can be designed either in the time or in a transform domain. The advantage of working in a transform space is that, if the transform is selected properly, the speech and noise signals may be better separated in that space, thereby enabling better filter estimation and noise reduction performance. The suppress noise from the speech signals without speech distortion it is an art of the noise removal approach. All filters do not give equal performance in every condition. Fusion techniques give better performance in noise reduction than the single noise removal approach. The discussion given in this paper will help for developing improved speech recognition system for noisy environment.

References:

1. A. Zehtabian and H. Hassanpour, A Non-destructive Approach for Noise Reduction in Time Domain, World Applied Sciences Journal 6 (1): 53-63, 2009
2. J. Chen, Subtraction of additive noise from corrupted speech for robust speech recognition, 1998
3. J. M. Górriz, A Novel LMS Algorithm Applied to Adaptive Noise Cancellation, 2009
4. J. Ramírez„ Voice Activity Detection. Fundamentals and Speech Recognition System Robustness, 2007
5. J.H. Husoy, Unified approach to adaptive filters and their performance, 2008
6. Jacob Benesty, Noise Reduction Algorithms in a Generalized Transform Domain, 2009
7. Jasha Droppo and Alex Acero, Noise robust speech recognition with a switching linear dynamic model, 2004
8. Li Deng, Large-Vocabulary Speech Recognition Under Adverse Acoustic Environments, 2000
9. Li Deng„ High-Performance Robust Speech Recognition Using Stereo Training Data, 2001
10. Liang Hong, Independent Component Analysis Based Single Channel Speech Enhancement Using Wiener Filter, 2003
11. Syed Abdul Rahman, Robust Speech Recognition Using Fusion Techniques and Adaptive Filtering, 2009
12. Sharath Rao K, Improved Iterative Wiener Filtering For Non-Stationary Noise Speech Enhancement, 2004
13. Taufiq Hasan, Suppression of Residual Noise From Speech Signals Using Empirical Mode Decomposition, 2009
14. Sundarrajan, A noise-estimation algorithm for highly non-stationary environments, 2005
15. Weizhong Zhu, Using Noise Reduction And Spectral Emphasis Techniques To Improve Asr Performance In Noisy Conditions, 2003
16. Yi Hu and Philipos C. Loizou, A comparative intelligibility study of single-microphone noise reduction algorithms, 2007

17. Yi Hu, Subjective comparison and evaluation of speech enhancement algorithms, 2007