Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 89 (2016) 666 – 676

Procedia Computer Science

Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)

Comparison of Speech Enhancement Algorithms

Siddala Vihari, A. Sreenivasa Murthy, Priyanka Soni and D. C. Naik*

U.V.C.E., Bangalore University, Bangalore, India

Abstract

The simplest and very familiar method to take out stationary background noise is spectral subtraction. In this algorithm, a spectral noise bias is calculated from segments of speech inactivity and is subtracted from noisy speech spectral amplitude, retaining the phase as it is. Secondary procedures follow spectral subtraction to reduce the unpleasant auditory effects due to spectral error. The drawback of spectral subtraction is that it is applicable to speech corrupted by stationary noise. The research in this topic aims at studying the spectral subtraction & Wiener filter technique when the speech is degraded by non-stationary noise. We have studied both algorithms assuming stationary noise scenario. In this we want to study these two algorithms in the context of non-stationary noise. Next, decision directed (DD) approach, is used to estimate the time varying noise spectrum which resulted in better performance in terms of intelligibility and reduced musical noise. However, the a priori SNR estimator of the current frame relies on the estimated speech spectrum from the earlier frame. The undesirable consequence is that the gain function doesn't match the current frame, resulting in a bias which causes annoying echoing effect. A method called Two-step noise reduction (TSNR) algorithm was used to solve the problem which tracks instantaneously the non-stationarity of the signal but, not by losing the advantage of the DD approach. The a priori SNR estimation was modified and made better by an additional step for removing the bias, thus eliminating reverberation effect. The output obtained even with TSNR still suffers from harmonic distortions which are inherent to all short time noise suppression techniques, the main reason being the inaccuracy in estimating PSD in single channel systems. To outdo this problem, a concept called, Harmonic Regeneration Noise Reduction (HRNR) is used wherein a non-linearity is made use of for regenerating the distorted/missing harmonics. All the above discussed algorithms have been implemented and their performance evaluated using both subjective and objective criteria. The performance is significantly improved by using HRNR combined with TSNR, as compared to TSNR, DD alone, as HRNR ensures restoration of harmonics. The spectral subtraction performance stands much below the above discussed methods for obvious reasons.

Keywords: Decision Directed Approach; Harmonic Regeneration; Speech Enhancement; Two-step Noise Reduction; Wiener Filtering.

1. Introduction

The processing of noisy speech signals to improve their perception by humans or better decoding by systems is what speech enhancement deals with. A Formulation of speech enhancement algorithms is to improve the performance of a system when its speech input is ruined by noise. It is usually hard to retain speech undistorted while reducing noise and thus, limitation on speech enhancement system's performance- the compromise between speech distortion and noise

*Corresponding author. Tel.: +91 -96114455231
E-mail address: chethan.naik24@gmail.com

doi:10.1016/j.procs.2016.06.032

reduction.For distorted speech with medium to high SNR, objective will be to produce subjectively natural signal by reducing noise level and for those with low SNR, objective could be to decrease the noise level, while preserving the intelligibility.

The most common factor that causes the degradation of speech's quality and intelligibility is background noise which can be stationary or non-stationary and is assumed to be uncorrelated and additive to the speech signal. A broad classification of speech enhancement methods can be given as *spectral processing* and *temporal processing* methods. The degraded speech goes through processing in frequency domain in the spectral processing methods, whereas processing will be in time domain for temporal processing method.

Spectral subtraction[1], known for its minimal complexity and relative ease in implementation is one of the oldest algorithms proposed in the area of background noise reduction. In this technique, the average magnitude of noise spectrum is subtracted from the noisy speech spectrum. The average magnitude of noise spectrum is estimated from the frames of speech absence, usually from initial frames of the signal in case of stationary noise conditions. In case the noise is non-stationary, the noise estimate has to be calculated every time the noise characteristics are changed. So, the spectral subtraction becomes inefficient for speech corrupted with non-stationary noise

Utilizing an MMSE criteria[2], using spectral component distribution models of speech and noise signals, the mean square error between the short time spectral magnitude of the clean speech and enhanced speech may be minimized. Speech enhancement based on noise suppression usually disturbs the spectral balance in speech[3], which results in unpleasant distortions in enhanced speech. LP residual enhancement is a method used for LP residual reconstruction.

Using different SNR parameters[4], we can formulate a short time spectral gain using Wiener filtering with DD approach in which frame delay results in an annoying reverberation effect. The problem is solved by TSNR[5], wherein, a second step is formulated so as to remove the delay. All the classic short time noise reduction algorithms are followed by HRNR algorithm[6] which is used to regenerate harmonics in the reconstructed signal.

Organization of the paper is, in Section 2, we discuss the implementation of Spectral subtraction method. In Section 3, we discuss the Wiener filtering using DD, TSNR, and TSNR followed by HRNR method. In Section 4, obtained results are presented. And in Section 5, summary of our work and conclusion drawn is given.

2. Spectral Subtraction

A noise spectrum estimate, derived from the signal measured while non-speech activity or beginning/ending of a speech signal was subtracted from the noisy speech spectrum to obtain a spectral subtraction estimator. A spectral

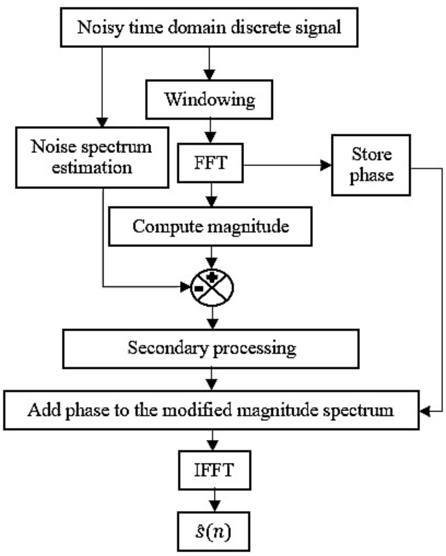

Fig. 1. Flow Chart Showing the Spectral Subtraction Followed by Secondary Processing Needed to Reduce Auditory Effects due to Spectral Error.