ISSN (Online) 2249 - 054 X
International Journal of Computing and Corporate Research
Specialized and Refereed Journal for Research Scholars, Academicians, Engineers and Scientists
http://www.ijccr.com
VOLUME 2 ISSUE 1 JANUARY 2012

ENHANCEMENT OF SUBJECTIVE & OBJECTIVE MEASURES AFFECTED BY THE NOISE

Anil Garg Member IEEE
Asstt. Prof. ECE Deptt.
M.M.Engg. College
Mullana(Ambala)

Dr.O.P. Sahu
Associate prof. & Head ECE Deptt.
NIT Kurukshetra
Haryana

Abstract

In this paper, several techniques based on spectral and temporal processing etc. are analyzed for dereverberation, single channel & multichannel to improve the Quality & intelligibility of the degraded speech. Although intelligibility of the degraded speech can be Enhanced by changing the fundamental frequency & spectral tilt. In this paper, for Single Channel speech Enhancement Techniques, Spectral Subtraction Process, Spectral Subtraction With Over subtraction Model, Non-Linear Spectral Subtraction Process, wiener filtering process & for Multi Channel Speech Enhancement Techniques, Adaptive Noise cancellation Multisensor beamforming have been discussed. This paper also includes the intelligibility comparison of the Lombard & speech produced in the noise

Keywords: *cepstral smoothing, Lombard Speech RASTA processing, Spectral processing, Temporal processing.*

Introduction- Speech enhancement system aims to improve the quality & Intelligibility of degrade speech for various applications[1].On the contrary, Quality improvement is usually associated with the loss of intelligibility relative to that of the degraded signal[2].In many practical situations, Speech is affected by several factors like Additive Noise, Reverberant noise multi speaker Speech etc. Speech produced in noise is called "**Lombard speech**"[17]. So, to

ISSN (Online) 2249 - 054 X
International Journal of Computing and Corporate Research
Specialized and Refereed Journal for Research Scholars, Academicians, Engineers and Scientists
http://www.ijccr.com
VOLUME 2 ISSUE 1 JANUARY 2012

obtain the Quality & intelligibility, Temporal, Spectral processing methods & change in fundamental frequency and spectral tilt for enhancement of speech have been discussed in this paper. In temporal processing method, the degraded speech is processed in time domain & enhance the characteristics of the speech signal in time domain, whereas in spectral processing method, the degraded speech is processed in frequency domain & attempt to suppress the noise. For the enhancement of the multi-speaker speech Spectral methods like speech specific approaches, Computational auditory scene analysis (CASA) methods can be used. In Temporal processing methods, LP residual enhancement, cepstral processing, Blind source separation (BSS) and Independent component Analysis (ICA) methods can be employed.

SECTION-I

Single-channel speech enhancement techniques Single-channel speech enhancement techniques apply to situations in which a unique acquisition channel is available. When the noise process is stationary i.e. does no changes w.r.t. time and speech activity can be detected, Spectral subtraction is performed by subtracting the average magnitude of the noise spectrum from the spectrum of the noisy speech to estimate the magnitude of the enhanced speech spectrum. The noise spectrum is estimated by averaging short-term magnitude spectra of the non speech segments [2]. One of the serious drawbacks of this method is that it produces **musical noise** in the enhanced speech. Although method proposed on cepstral smoothing can effectively prevent spectral peaks of short duration that may be perceived as ***musical noise***. Spectral Subtraction With Over subtraction Model is applied in order to compensate for the "musical noise" effect [3]. Non-Linear Spectral Subtraction combines extended noise model & Non-linear implementation of the subtraction process. **In temporal processing method**, the basic approach for speech enhancement is to identify the high SNR portions in the noisy speech signal, and enhance those portions relative to the low SNR portions, without causing significant distortion in the enhanced speech. The residual signal samples are multiplied with the weight function, and the modified residual is used to excite the time varying all pole filter derived from the given noisy speech to generate the enhanced speech [8]. **In wiener filtering technique**, an optimum filter is first estimated from the noisy speech. The filter is then applied either in time domain or frequency domain to obtain an estimate of the degraded speech. **The Wiener filter using Adaptive Approach** benefits from the varying local statistics of the speech signal.

SECTION-II

Multi Channel Speech Enhancement Techniques Multi-channel speech enhancement techniques take advantage of the availability of multiple signal input to system, making possible the use of noise references in an adaptive noise cancellation device. **In This paper**, two different systems adaptive noise cancellation & speech beamforming through array processing have been consider [3]. In Dual channel Enhancement Techniques, a reference signal for the

ISSN (Online) 2249 - 054 X
International Journal of Computing and Corporate Research
Specialized and Refereed Journal for Research Scholars, Academicians, Engineers and Scientists
http://www.ijccr.com
VOLUME 2 ISSUE 1 JANUARY 2012

noise is available & hence adaptive noise cancellation Technique can be applied. **Adaptive noise cancellation** based in the availability of an auxiliar channel, known as reference path, where a correlated sample or reference of the contaminating noise is present. This reference input will be filtered following an adaptive algorithm, in order to subtract the output of this filtering process from the main path, where noisy speech is present [3]. **Multisensor beamforming** through microphone arrays being delay-and-sum beamforming is based on the assumption that the contribution of the reflections is small & direction of arrival of desired signal is pre-determined. However, important distortions are introduced because of the inaccurate speech and noise power spectral densities estimates. On the other hand, poor performance is noticed at low frequencies which exhibit a high spatial coherence between noises in the microphone output signals. This is due to the small spacing required between the microphones [14]. The method that uses the **spectral characteristics** rely on the estimation of the pitch of the individual speaker and using this information, the desired speaker is enhanced by retaining only pitch and harmonic components and ignoring the remaining spectral components [9]. **In temporal processing method** the basis for the multi-speaker speech enhancement is that the relative positions of the instants of significant excitation in the direct component of the speech signal remain unchanged at each of the microphones for a given speaker. These sequences differ only by a fixed delay corresponding to the relative distances of the microphones from the speaker. By estimating time delays and using the knowledge of excitation source characteristics a weight function is derived for each speaker to identify the speech components of desired speaker relative to the other speaker [10].

In Speech specific Approaches, first the spectral peaks are identified from the windowed mixed speech spectrum. The peaks are accumulated in a table that was used to construct a histogram. The fundamental frequency (F_0) of a first speaker is determined from the histogram and the F_0 of the second speaker was obtained by removing the harmonics belonging to the first speaker from the peak table and repeating the histogram calculation for remaining peaks. The speech of each speaker is then resynthesized by taking IDFT of separated pitch and harmonics[9].

Computational Auditory Scene Analysis (CASA), A approach to speech enhancement i.e. speech stream segregation. Speech *stream* segregation works as the frontend system for automatic speech recognition just a**s** hearing aids for hearing impaired people. In CASA it is expected to enhance a sound, not restricted to a human voice, by reducing background noises, echoes and the sounds of competing talkers, and thus improve the performance **of** hearing aids [11]. A main advantage of CASA is that it does not make strong assumptions about interference. A typical CASA system contains four stages as shown in Fig-1 : peripheral analysis, feature extraction, segmentation and grouping

ISSN (Online) 2249 - 054 X
International Journal of Computing and Corporate Research
Specialized and Refereed Journal for Research Scholars, Academicians, Engineers and Scientists
http://www.ijccr.com
VOLUME 2 ISSUE 1 JANUARY 2012

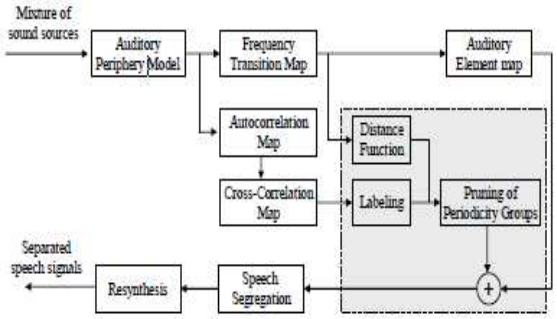

Fig. 1. Block diagram of the proposed speech separation system.

Blind Source Separation Method (BSS) Blind source separation (BSS) is an approach to estimating original source signals using only mixed signal observed in each input channel [12]. Typically, mixed signals are acquired by a number of sensors, where each sensor receives a different combination of the source signals. The term blind refers to the fact that only the recorded mixtures are known[13].

Independent Component Analysis (ICA)- Independent component analysis(ICA) is a novel statistical technique in signal processing and machine
learning that aims at finding linear projections of the data that maximize their mutual independence. Its main applications are blind source separation (BSS)] and feature extraction. When applied to speech frames, ICA provides a linear representation that maximizes the statistical independence of its coefficients, and therefore finds the directions with respect to which the coefficients are as sparsely distributes as possible.

SECTION-III

Reverberation, is described by the concept of reflections. The desired source produces wavefronts, which propagate outward from the source. The wavefronts reflect off the walls of the room and superimpose at the microphone. These distortions result in an audible difference between the anechoic and the reverberant speech, and degrades speech intelligibility and fidelity [4]. In the usual formulation of the deconvolution problem, it is assumed that the system input s(t) and system output x(t) are both known. The spectral subtraction based enhancement methods aims at the suppression of late reverberation to improve speech intelligibility [5]. In excitation source information based reverberant based speech enhancement algorithms which primarily aim to emphasize the high signal to reverberant ratio(SRR) regions relative to low SRR regions of the reverberant speech signal in the temporal domain[6,7].

ISSN (Online) 2249 - 054 X
International Journal of Computing and Corporate Research
Specialized and Refereed Journal for Research Scholars, Academicians, Engineers and Scientists
http://www.ijccr.com
VOLUME 2 ISSUE 1 JANUARY 2012

SECTION-IV

RASTA Processing of Speech

This Method can perform well in white noise condition but failed in real colored noise environments with different SNRs. This leads to the use of RelAtive SpecTrAl (RASTA) algorithm for speech enhancement which was originally designed to alleviate effects of convolutional an additive noise in automatic speech recognition (ASR) [16]. In automatic recognition of speech (ASR), the task is to decode the linguistic message in speech. This linguistic message is coded into movements of the vocal tract. The speech signal reflects these movements. The rate of change of nonlinguistic components in speech often lies outside the typical rate of change of the vocal tract shape. It suppresses the spectral components that change more slowly or quickly than the typical range of change of speech. The maximum modulation frequency of the modulation spectrum is half of the sampling frequency of RASTA filter.

SECTION-V

Lombard speech Speech intelligibility degrades in the presence of moderate and intense noise. While factors such as an increase in speech output level can, to some extent, boost intelligibility by raising signal- to-noise ratio (SNR), level increases alone are undesirable due to their unpleasant and fatiguing effect on the listener. Speech produced in noise is called "**Lombard speech**" has been found to be more intelligible than speech produced in quiet when both are mixed with noise at the same SNR [18]. Lombard speech demonstrates an overall increase in duration, and increase in F0 and a flattening of spectral tilt (more energy at higher frequencies). The intelligibility gain of Lombard speech over speech produced in quiet was thus attributed to durational increases (i.e., slow speaking rate) and more spectral energy in higher frequencies: an increase in duration provides more opportunities to glimpse acoustic information useful for phonetic distinctions and more spectral energy in higher frequencies leads to more glimpses in the presence of a speech-shaped masker. To evaluate the contribution of increases in F0, each quiet sentence was artificially manipulated using a high-quality source-filter vocoder to add a constant amount to the F0 across the utterance to obtain a signal having the same mean F0 as that of the corresponding Lombard sentence. Similarly, to examine the effect of spectral tilt flattening, each quiet sentence was passed through an infinite impulse response filter of order 100 whose magnitude response was designed in such a way that the overall spectrum of the filtered signal was the same as that of the corresponding Lombard sentence. The pattern of an increased amount of the time– frequency plane glimpsed, as a result of spectral energy shift to higher frequencies, together with durational lengthening, is illustrated in Fig.2

ISSN (Online) 2249 - 054 X
International Journal of Computing and Corporate Research
Specialized and Refereed Journal for Research Scholars, Academicians, Engineers and Scientists
http://www.ijccr.com
VOLUME 2 ISSUE 1 JANUARY 2012

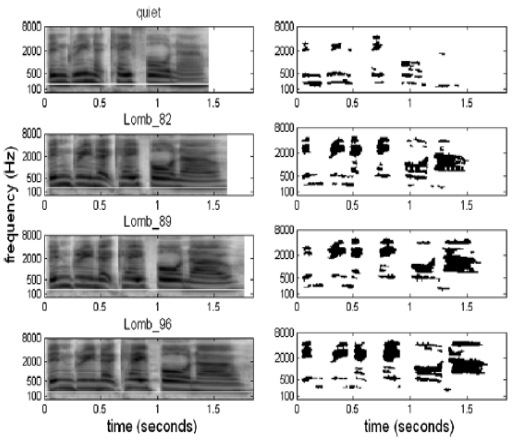

Fig-3 Spectro-temporal excitation patterns (left column) and glimpses (right column) for the sentence. . Horizontal lines in the excitation patterns indicate a frequency of 200 Hz.

The two findings that F0-shifted speech was no more intelligible than the baseline ‘‘quiet” speech and shifting F0 of spectrum-manipulated speech did not further improve intelligibility suggest that increases in F0 make little contribution. However, it was found that there were significant intelligibility gains of spectrum-manipulated speech over quiet speech and the gain tended to increase with manipulation scale. These findings support that a flattening of spectral tilt helps to improve intelligibility in the presence of speech shaped noise [19].

SECTION-VI

Spectral subtraction for single-channel speech enhancement techniques produces **musical noise** in the enhanced speech. cepstral smoothing can effectively prevent spectral peaks of short duration that may be perceived as ***musical noise***. In multi-speaker case spectral processing based methods depends on the fundamental frequency characteristics & in temporal processing based methods speech of each speaker is enhanced by deriving speaker-specific weight function. In case of reverberant speech enhancement, The spectral subtraction based enhancement methods aims at the suppression of late reverberation to improve speech intelligibility. In temporal processing based approach the speech-specific features used for identifying the gross & the fine weight function. RASTA processing, improves the performance of a recognizer in presence of convolutional and additive noise. This method does not require a long term average, which may be difficult to obtain in real time implementations. It has been observed that there is a significant contribution to Lombard speech intelligibility of spectrum flattening and failed to find a perceptual influence of an increase in F0. The possibility that a lengthened duration helps to improve the intelligibility of Lombard speech in noise

ISSN (Online) 2249 - 054 X
International Journal of Computing and Corporate Research
Specialized and Refereed Journal for Research Scholars, Academicians, Engineers and Scientists
http://www.ijccr.com
VOLUME 2 ISSUE 1 JANUARY 2012

REFERENCES

[1] Y. Hu and P. C. Loizou, "Subjective comparison of speech enhancement algorithms," in *Proc. IEEE Int.Conf. Acoust., Speech, Signal Processing*, 2006, pp.

[2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," in *Proc. IEEE ASSP-27,* Apr. 1979, pp. 113-120

[3] Javier Onega-Garcia and Joaquin Gonza'lez-"overview of speech enhancement techniques for automatic speech recognition" supported by CICYT under project TIC94-0030

[4] E.Habets,"single-and multimicrophone speech dereverberation using spectral enhancement," Ph.D. dissertation, technische universiteit Eindhoven, the Netherlands, june. 2007.[online]. Avilable: http://alexandria.tue.nl/extra2/200710970.pdf

[5] K.Lebart and J. Boucher. " A new method based on spectral subtraction for speech dereverberation." Acta Acoustica, vol.87,pp.359-366,2001.

[6] B. Yegnanarayana and P. satyanarayana Murthy, " Enhancement of reverberant speech using LP residual signal," IEEE. Trans. Speech, Audio Process, vol.8, pp. 267-281, May 2000.

[7] B. Yegnanarayana, S.R.M. Prasanna, R. Duraiswami, and D. Zotkin, " processing of reverberant speech for time delay estimation," IEEE. Trans. Speech, Audio Process, vol.13, pp. 1110-1118, Nov. 2005.

[8] B. Yegnanarayana, C. avendano H. hermansky and P. satyanarayana Murthy, " Speech Enhancement using LP residual ," speech communication , vol.28, pp. 25-42, May 1999.

[9] T.Pearson, "separation of speech from interfering speech by means of harmonic selection," J. Aco. Am., vol. 60, pp. 911-918, oct. 1976.

[10] B. Yegnanarayana, S.R.M. Prasanna and M. Mathews," Enhancement of speech in multispeaker enviorment," in Proc. European conf. speech process. technology, Geneva, Switzerland, 2003, pp. 581-584.

[11] Hiroshi G. Okuno, Tomohiro Nakatani, and Takeshi Kawabata, "A new speech enhancement :speech stream segregation."in proc. ICSPL 96, 1996, pp.2356-2359.

ISSN (Online) 2249 - 054 X
International Journal of Computing and Corporate Research
Specialized and Refereed Journal for Research Scholars, Academicians, Engineers and Scientists
http://www.ijccr.com
VOLUME 2 ISSUE 1 JANUARY 2012

[12] Yu Takahashi, Tomoya Takatani, Keiichi Osako, Hiroshi Saruwatari, Kiyohiro Shikano, "Blind Spatial Subtraction Array for Speech Enhancement in Noisy Environment," IEEE Trans. Audio, speech, and language process., vol 17, No. 4, pp. 650-664, May 2009.

[13] J. Cardoso, "Blind Signal Separation : Statistical Principles," Proc. IEEE, vol. 86, pp. 2009-2025,1998.

[14] Djamila Muhmoudi and Andrzej Drygujlo, "combined wiener and coherence filtering in wavelet Domain for microphone array speech enhancement" in proc. Acoustic, speech, pp 385-388, 1998.

[15] Hynek Hermansky, Nelson Morgan "RASTA Processing of Speech"in IEEE transaction on speech and audio processing,vol.2.no.4 dct.1994.

[16] S.K.Shah1, J.H.Shah2, N.N.Parmar3, " Evaluation of RASTA approach with modified parameters for speech enhancement in communication systems"in IEEE symposium on computers and informatics, pp. 159-162

[17] Youyi Lu, Martin Crooke " The contribution of changes in F0 and spectral tilt to increase intelligibility of speech produced in noise" Speech Communication, pp. 1253-62, july 2009.

[18] Pittman, A.L., Wiley, T.L., " Recognition of speech produced in noise" J.speech Lang. Hear. Pp. 487-496. 2001.

[19] Lu, Y., Cooke, M.P., 2008. "Speech production modifications produced by competing talkers, babble and stationary noise". J. Acoust. Soc. Amer. Pp. 3261–3275. 2008

.