www.ijaret.org Vol. 1, Issue VI, July 2013
ISSN 2320-6802

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY

WINGS TO YOUR THOUGHTS…..

AN OVERVIEW OF SPEECH ENHANCEMENT TECHNIQUES AND EVALUATION

Sukhvinder Kaur[1], Anil Garg[2], Pankaj Bactor[3]

[1] M.Tech Student MMEC,
M.M. University, Mullana-Ambala-Haryana, India
sukhsoni2010@gmail.com

[2] Asst. Prof. in ECE department MMEC,
M.M. University, Mullana-Ambala-Haryana, India
agarg001@yahoo.com

[3] Asst. Prof. in ECE department MMEC,
M.M. University, Mullana-Ambala-Haryana, India

Abstract: *Speech Enhancement Techniques are used to improve the intelligibility and quality of the degraded speech. Different types of speech enhancement algorithms are discussed in this paper. The perceptual aspects of speech degraded by the additive type of noise can be improved and therefore speech algorithms are also known as noise suppression algorithms. Tests like formal subjective listening and Objective measures improve the quality of speech. Dynastat, Inc., designed the ITU-T P.835 methodology for the subjective evaluation. In the subjective evaluation quality of speech is evaluated along three dimensions: signal distortion, noise distortion and overall quality. The objective measures performance depends on the predicted quality of the noisy speech which was enhanced by noise suppression algorithms. There is wide range of distortions introduced in different classes of speech enhancement algorithms: spectral subtractive, subspace, statistical-model based, and Wiener algorithms.*

Keywords: *Speech enhancement; subjective listening tests; ITU-T P.835; Objective measures.*

1. INTRODUCTION

Speech quality can be improved by using different Speech enhancement algorithms. Quality means clarity and intelligibility, pleasantness, or compatibility. It is difficult to measure the intelligibility and pleasantness by using mathematical algorithm. Hence, listening tests are used but may be they are expensive. For the improvement of the performance of modern communication devices in noisy environments different types of speech enhancement algorithms have been proposed. But, it is still remain unclear that which speech enhancement algorithm performs very well in the real-world listening situations where the background noise level and characteristics are constantly changing. The intelligibility and overall perceptual quality of a degraded speech signal can be improved by using audio signal processing techniques. Enhancing of the speech degraded by noise, or noise reduction are the most important field of speech enhancement, and it is used for many of the applications such as mobile phones, VoIP, teleconferencing systems, speech recognition, and hearing aids [1]. The solution of speech enhancement depends largely on the application of the characteristics of noise source or interference, i.e., the relationship of the noise to the clean signal or number of microphone available. The interference could be noise like (e.g., fan noise) or speech like, such as an environment (e.g., restaurant) with competing speakers.

This paper is represented in such a manner that section 2 provides brief review of the different speech enhancement techniques. The subjective and objective measures are described in section 3. Section 4 represents the conclusion part.

2. SPEECH ENHANCEMENT TECHNIQUES

The algorithms of speech enhancement for noise reduction can be categorized into three main categories:
1. Filtering techniques (Spectral Subtraction Method, Wiener Filtering, Signal subspace approaches (SSA).
2. Spectral restoration (Minimum Mean-Square-Error Short-Time Spectral Amplitude Estimator (MMSE-STSA)) 3. Speech-Model-Based [2].

2.1SPECTRAL-SUBTRACTIVE ALGORITHMS

ISSN 2320-6802
INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

Initially Weiss et al [3] proposed the Spectral-subtractive algorithm in the correlation domain and later by Boll [4] in the Fourier transform domain. The spectral subtraction algorithm is historically one of the first algorithms proposed for noise reduction (Boll, 1979; Weiss et al., 1974). It is based on a simple principle of assuming additive noise. Estimation of the clean signal spectrum can be obtained by subtracting an estimate of the noise spectrum from the noisy speech spectrum. The noise spectrum can be estimated during periods when the signal is absent. By computing the inverse discrete Fourier transform of the estimated signal spectrum one can obtain the enhanced signal. This algorithm involves a single forward and inverse Fourier transform. The subtraction process needs to be done carefully to avoid any speech distortion. Too much subtraction results in the removal of speech information while little subtraction results in interfering noise.

The noise spectrum can be estimated, and updated, during the periods when the signal is absent or when only noise is present.

$$X_e(\omega) = [|Y(\omega)| - |D(\omega)|]e^{j\theta y} \quad (1)$$

Here $Y[\omega]$ can be expressed in terms of Magnitude and phase as

$$Y[\omega] = |Y(\omega)|e^{j\theta y} \quad (2)$$

Noise spectrum in terms of magnitude and phase spectra as

$$D[\omega] = |D(\omega)|e^{j\theta y} \quad (3)$$

The enhanced speech signal is finally obtained by computing the inverse Fourier Transform of the estimated clean speech | Se[w]| for magnitude. General version of the spectral subtraction algorithms is

$$X_e[\omega]^P = |Y[\omega]|^P - |D_e[\omega]|^P \quad (4)$$

Where P is the power exponent¸ when P=1 yielding the magnitude spectral subtraction algorithm and P=2 yielding the power subtraction algorithm. Further spectral subtraction can be divided into two categories: spectral subtraction with over subtraction model and non-linear spectral subtraction.

2.1.1. SPECTRAL SUBTRACTION WITH OVER SUBTRACTION MODEL

The alternative model is the classical spectral subtraction **(SS)** procedure which was first introduced in order to compensate the "musical noise" effect [5]. The general expression of the **SS** with over subtraction model is given by:

$$\left|\tilde{X}_i(\varpi)\right|^2 = \begin{cases} Y_i(\varpi)^2 - \alpha \cdot \left|\overline{R_i(\varpi)}\right|^2, if\, |Y_i(\varpi)|^2 - \left|\overline{R_i(\varpi}\right. \\ \qquad > \beta \cdot \left|\overline{R_i(\varpi)}\right|^2 \\ \beta \cdot \left|\overline{R_i(\varpi)}\right|^2, otherwise \end{cases} \quad (5)$$

Where α>l minimizes the appearance of negative values that generate spectral spikes, and 0<β<<1sets an spectral flooring which reduces the perception of musical noise. The optimal value for α can be set as a function of the *SNR*, as high *SNR* frames need less compensation that low *SNR* frames

2.1.2. NON LINEAR SPECTRAL SUBTRACTION

(NSS) approach [6] is based on combining two different ideas: i) The use of an extended noise model and an over subtraction model, and ii) Non-linear implementation of the subtraction process, therefore it is clear that the subtraction process must depend on the SNR of the frame, in order to apply less subtraction with high SNRs and vice versa. In the NSS technique, an estimate of both noise and speech with an estimator of the noisy can be derived from the following expressions

$$\left|\overline{R_\iota(\varpi)}\right| = \lambda_R\left|\overline{R_{\iota-1}(\varpi)}\right| + (1-\lambda_R)|R_i(\varpi)| \quad (6)$$

$$\left|\overline{Y_\iota(\varpi)}\right| = \lambda_y\left|\overline{Y_{\iota-1}(\varpi)}\right| + (1-\lambda_y)|Y_i(\varpi)| \quad (7)$$

The extended model of noise use the generic function $\phi[\rho_i(\varpi), \alpha_i(\varpi), |R_i(\varpi)|]$ which depends on noise estimator, on the spectral-dependent over subtraction factor, $\alpha_i(\varpi)$ and the SNR of each spectral component of the analysis frame, $\rho_i(\varpi)$ can be calculated as

$$\rho_i(\varpi) = \frac{\left|\overline{Y_{SNR,\iota}(\varpi)}\right|}{\left|\overline{R_\iota(\varpi)}\right|} \quad (8)$$

Being..

$$\left|\overline{Y_{SNR,\iota}(\varpi)}\right| = \lambda_{SNR}\left|\overline{Y_{\iota-1}(\varpi)}\right| + (1-\lambda_{SNR})|Y_i(\varpi)| \quad (9)$$

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

Where Φ is an arbitrary non linear function of subtraction process that includes upper and lower boundaries spectral components?

$$|\overline{R_{\iota}(\varpi)}| \leq \Phi[\rho_{i}(\omega), \alpha_{i}(\square), |\overline{R_{\iota}(\square)}|] \leq 3 \cdot |\overline{R_{\iota}(\square)}| \quad (10)$$

2.2 WIENER FILTERING

Wiener filtering approach derives the enhanced signal by optimizing a mathematically tractable error criterion, the mean square error. Wiener filter that are based on post filtering are used in microphone array speech enhancement systems. In the statistical linear filtering the input signal goes through a linear and time- invariant system to produce an output signal. And the system is designed in such a way that the output signal is as close as to the desired signal. This can be done by computing the estimation error and making it as small as possible. The optimal filter that minimizes the estimation error is called the wiener filter, named after the mathematician Norbert Wiener [7], who first formulated and solved the filtering problem in the continuous domain. The constraints placed on the linear filter makes the analysis easier to handle. In principle, the filters could be finite impulse response (FIR) or the infinite impulse response (IIR), but FIR filters are used because: they are inherently stable, and the resulting solution is linear and computationally easy to evaluate.
Assuming a FIR system i.e., having

$$\tilde{d}(n) = \sum h_k\, y(n-k) \quad n = 0, 1, 2, 3\ldots\ldots \quad (11)$$

Where $[h_k]$ are the FIR filter coefficients, and M is the number of coefficients. The filter coefficients $[h_k]$ is computed to minimized the estimation error i.e., d(n) - $\tilde{d}(n)$. The mean square error is commonly used as a criterion for minimization, and the optimal filter coefficients ($\dot{h}$) can be derived in the time or frequency domain. (eq., 12and 13)

$$R_{yy}\dot{h} = \dot{r}_{yd} \quad (12)$$

$$H(\omega_k) = \frac{P_{dy}(\omega_k)}{P_{yy}(\omega_k)} \quad (13)$$

The Wiener filtering algorithm can be implemented either iteratively or non-iteratively. The iterative algorithm generally assumed a model of the clean spectrum and attempted to estimate the parameters of the model iteratively. The AR speech production model was used successfully in [8] for estimating the wiener filter.

Wiener filters were derived by minimizing the speech distortion subject to the noise distortion falling below a given threshold level (e.g., masking threshold). The Wiener filters are considered to be linear estimators of the clean signal spectrum, and they are optimal in the mean square sense. In these types of filters enhanced time-domain signal is obtain by convolving the noisy speech signal with a linear (Wiener) filter. While in frequency domain the enhanced spectrum is obtained by multiplying the input (noisy) spectrum by the Wiener filter. Non Linear estimators of clean signal spectrum yields better performance.

2.2.1. WAVELET TRANSFORM BASED WIENER FILTER

Wiener filter in the wavelet transform domain *H* is provided in the Fourier domain. Under the assumption of the diagonality of the wavelet covariance matrices of the desired signal and noise, the scalar Wiener filter is expressed as:

$$\mathcal{H}_{\varpi} = \frac{E\{\mathcal{S}^2\}}{E\{\mathcal{X}^2\}} \quad (14)$$

Where the quantities $E\{S^2\}$ and $E\{X^2\}$ are the wavelet power spectra of the desired and the noisy speech signals, respectively. The wavelet transform Wiener filter *H*, applied in sub-band can be described by the following expression:

$$H_{\varpi_\phi}(n) = \frac{\frac{2}{M.M-1}\sum_{i=1}^{M-1}\sum_{J=i+1}^{M} X_{i,\phi}(n)X_{j,\phi}(n)}{\frac{1}{M}\sum_{i-1}^{M} X_{i,\phi}(n)^2} \quad (15)$$

2.3 STATISTICAL MODEL BASED METHODS

The speech enhancement problem is posed in a statistical estimation frame work. The Fourier transform coefficients of the noisy signal are the given corresponding asset of measurements. The minimum mean square error algorithms fall in this category. These are used to find a linear (or non linear) estimator of the parameter of interest, the transform coefficients of the clean signal. Work in this area is initiated by Mac Aulay and Malpass [9], who proposed a maximum likelihood approach for estimating Fourier coefficients (spectrum) of the clean signal, and was followed by the work by Ephraim and Malah [10], who proposed the MMSE estimator of the magnitude spectrum.

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

Statistical estimators of the magnitude-squared spectrum are derived based on the assumption that the magnitude- squared spectrum of the noisy speech signal can be computed as the sum of the (clean) signal and noise magnitude-squared spectra. Maximum *a posterior* (MAP) and minimum mean square error (MMSE) estimators are derived based on a Gaussian statistical model. The gain function of the MAP estimator was found to be identical to the gain function used in the ideal binary mask (IdBM) that is widely used in computational auditory scene analysis (CASA).

The optimal complex discrete Fourier transform coefficients of the clean signal are derived in this approach in mean square sense. This method focuses on the non linear estimators of magnitude rather than the complex spectrum of the signal by using different statistical models and optimization criteria. The non linear estimators take the probability density function (PDF) of the noise and the speech DFT coefficients into account and use it in some cases of non Gaussian prior distributions. To find a non linear estimator parameter in a speech enhancement problem i.e., posed in statistical estimation, and then given set of measurements depends on other unknown parameter. Measurements corresponding to the set of DFT coefficients of the noisy signal and parameters of interest are the set of DFT coefficients of the clean signal. Various techniques that are used for deriving non linear estimators are: Maximum likelihood estimators and the Bayesian estimators (e.g., MMSE and maximum *a posteriori estimators).*

y is the n-point data set that depends on an unknown parameter θ In speech enhancement, y is the noisy speech magnitude spectrum and the parameter of interest θ might be the clean speech magnitude spectrum.

The pdf of y is denoted by p(y;)

$$\tilde{\theta}_{\mathrm{ML}} = \arg \underbrace{max}_{\theta}\, p(y\,;\,) \tag{16}$$

$\tilde{\theta}_{\mathrm{ML}}$ is called the maximum-likelihood estimate of θ. The pdf of y i.e., p(y;) is called likelihood function as it is function of unknown parameter.

$$\tilde{X}(\omega_k) = \left[\frac{1}{2} + \frac{1}{2}\sqrt{\frac{\gamma_k - 1}{\gamma_k}}\right] Y(\omega_k) \tag{17}$$

$$= G_{ML}\gamma_k Y(\omega_k) \tag{18}$$

$G_{\mathrm{ML}}(\gamma_k)$is the gain function of maximum-likelihood estimator. The maximum-likelihood suppression rule provides considerably smaller attenuation as compared to the power subtraction and wiener suppression rules.

3. TESTING METHODOLOY

It is very difficult to find reliable and fair comparison among different algorithms. This is due to lack of common speech database used in algorithms evaluation, different types of noise and differences in the testing methodology the most accurate method for evaluating speech quality is subjective listening tests. But sometimes objective measures would also be used for predicting speech quality with high correlation.

3.1 SUBJECTIVE LISTENING

Subjective evaluation of speech enhancement algorithms is concerned with a new ITU-T standard (P.835) that was designed to lead the listeners to integrate the effects of both signal and background distortion in making their ratings of overall quality. Although subjective evaluation of speech enhancement algorithms is accurate and reliable when it is performed under stringiest conditions (e.g., sizeable listener panel, inclusion of anchor conditions, etc. but it is costly and time consuming.

The noisy speech corpus (NOIZEUS) is subsequently used for the subjective evaluation of 13 speech enhancement methods encompassing four classes of algorithms: spectral subtractive, subspace, statistical-model based and Wiener-type algorithms. Dynastat, Inc., performed The subjective evaluation using the ITU-T P.835 methodology that was designed to evaluate the speech quality along three dimensions: signal distortion, noise distortion and overall quality. The P.835 methodology was designed to reduce the listener's uncertainty of a noisy speech signal in a subjective listening test i.e., the speech signal, the background noise, or both, should be form on the basis of their ratings of overall quality. This method instructs the listener to successively attend to and rate the enhanced speech signal on:

(1) The speech signal alone using a five-point scale of signal Distortion (SIG);
(2) The background noise alone using a five-point scale of Background intrusiveness (BAK);
(3) The overall quality using the scale of the mean opinion Score (OVRL) - 1 bad 2 poor 3=fair 4=good 5=Excellent.

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

Table I and Table II describes the SIG and BAK scale. A total of 32 listeners were recruited for the listening tests. The results of the subjective listening tests were reported in and In this paper, we make use of the subjective ratings along the three quality scales (SIG, BAK, OVRL) are used to evaluate conventional and new objective measures [11], [12].

Table I. Description of SIG scale

SIG scale	
Rating	Description
5	Very natural, no degradation
4	Fairly natural, little degradation
3	Somewhat natural, somewhat degraded
2	Fairly unnatural, fairly degraded
1	Very unnatural, very degraded

Table II. Description of BAK scale

BAK scale	
Rating	Description
5	Not noticeable
4	Somewhat noticeable
3	Noticeable but not intrusive
2	Fairly conspicuous, somewhat intrusive
1	Very conspicuous, very intrusive

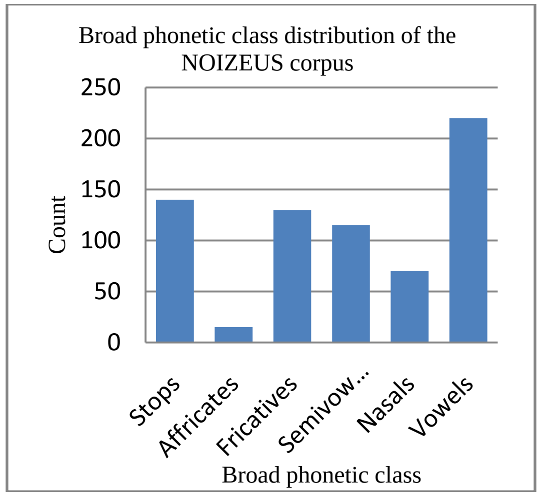

Figure 3: Broad phonetic class distribution of the NOIZEUS corpus

NOIZEUS is a noisy speech corpus recorded in the lab to facilitate comparison of speech enhancement algorithms among research groups. The noisy database contains 30 IEEE sentences (IEEE Subcommittee, 1969) produced by three male and three female speakers, and was corrupted by eight different real-world noises at different SNRs. The noise includes suburban train noise, multi-talker babble, car, exhibition hall, restaurant, street, and airport and train-station noise. The broad phonetic class distribution used in NOIZEUS is shown in Fig.3.

3.2 OBJECTIVE MEASURES

Evaluation of several objective speech quality measures can be carried out in the terms of: segmental SNR (seg SNR), weighted-slope spectral distance (WSS), PESQ, LPC-based objective measures including the log-likelihood ratio (LLR), Itakura-Saito distance measure (IS), and cepstrum distance measures (CEP) and frequency-weighted segmental SNR (fwsegSNR).

3.2.1 PESQ

The PESQ measure is the most complex objective measures to compute. ITU-T recommended it for the speech quality assessment of 3.2 kHz (narrow-band) handset telephony and narrow-band speech codec [13]. The PESQ measure can be computed by equalizing as the original clean and degraded signals at first level of standard listening and then filtered it by a filter having similar response to that of a standard telephone handset. For correction of time

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

delays the signals are time aligned through an auditory transform for obtaining the loudness Spectra. The difference in loudness between the original and degraded signals is computed and averaged over time and frequency to produce the prediction of subjective quality rating.

3.2.2 LPC-BASED OBJECTIVE MEASURES

The spectral envelope difference between the input (clean) signal and the processed (or corrupted) signal is assessed by the LPC-based measures. Three different LPC based objective measures were: the log likelihood ratio (LLR), the Itakura–Saito (IS), and the cepstrum (CEP) distance measures. All these three measures assess the difference between the spectral envelopes, as computed by the LPC model, of the noise-free and processed signals. The LLR measure is defined as (Quackenbush *et al.*, 1988)
The LLR measure is defined as

$$d_{LLR}(\vec{a}_p, \quad \vec{a}_c) = log\left(\frac{\vec{a}_p \boldsymbol{R}_c \vec{a}_p{}^T}{\vec{a}_c \boldsymbol{R}_c \vec{a}_c{}^T}\right) \tag{19}$$

The IS measure is defined as

$$d_{IS}(\vec{a}_{p,}\vec{a}_c) = \frac{\sigma_c{}^2}{\sigma_p{}^2}\left(\frac{\vec{a}_p \boldsymbol{R}_c \vec{a}_c{}^T}{\vec{a}_c \boldsymbol{R}_c \vec{a}_c{}^T}\right) + log\left(\frac{\sigma_c{}^2}{\sigma_p{}^2}\right) - 1 \tag{20}$$

The CEP distance provides an estimate of the log spectral distance between two spectra and is computed as follows

$$d_{CEP}(\vec{c}_c, \vec{c}_p) = \frac{10}{\log 10}\sqrt{2\sum_{k=1}^{p}\left[c_c(k) - c_p(k)\right]^2} \tag{21}$$

3.2.3 TIME-DOMAIN AND FREQUENCY-WEIGHTED SNR MEASURES

The time-domain segmental SNR (segSNR) measure was computed as per [14]. Only frames with segmental SNR in the range of 10 to 35 dB were considered in the average. The frequency-weighted segmental SNR (fwSNRseg) was computed using the following equation:

$$fwSNR_{seg} = \frac{10}{M} \times \sum_{m=0}^{M-1} \frac{\sum_{j=1}^{k} W(j,m) log_{10} \frac{|X(j,m)|^2}{\left(|X(j,m)| - |\hat{X}(j,m)|\right)^2}}{\sum_{j=1}^{k} W(j,m)} \tag{22}$$

4. CONCLUSION

The subjective evaluation was done by Dynastat using the ITU-T P.830 methodology. The performance of the Statistical-model based algorithm is consistently best across all conditions and yields highest quality. The intelligibility evaluation of speech enhancement algorithms indicates that there is not any algorithm which provides significant improvement in intelligibility as reference to the noisy speech.
Objective measures evaluation is presented in terms of correlation with quality. The objective measure should be able to assess the quality of the processed speech without needing access to the original speech signal. Current objective measures are limited to the original speech signal, and some can only model the low-level processing (e.g., masking effects) of the auditory system. But, despite there are some limitations of these objective measures have been found to correlate very well with subjective listening tests (e.g., MOS scores).

1. The spectra obtained from the subtractive rules may contain some negative values due to errors in estimating the noise spectrum. In the time-domain, these peaks sound like tones with frequencies that change randomly from frame to frame and introduce a new type of ''noise", often called musical noise (Berouti et al., 1979).
2. The main drawback of the iterative Wiener filtering approach was that as additional iterations were performed, the speech formants shifted in location and decreased in formant bandwidth [15].
3. The maximum-likelihood suppression rule provides considerably smaller attenuation as compared to the power subtraction and wiener suppression rules.

REFERENCES

[1] J.Benesty, S. Makino, J. Chen (Ed). Speech Enhancement. Pp.1-8. Springer, 2005. ISBN 978-3-540-24039-6.
[2] J. Benesty, M. M. Sondhi, Y. Huang (Ed). Springer Handbook of Speech Processing. Pp.843-869. Springer, 2007.ISBN978-3-540-49125-5.

[3] Weiss, M., Aschkenasy, E., and Parsons, T. (1974), Study and development of the INTEL techniques for improving speech intelligibility, Technical Report NSC-FR/4023, Nicolet Scientific Corporation.

[4] Boll, S.F.(1979), Suppression of acoustic noise in speech using spectral subtraction, IEEE Trans, Acoust. Speech Signal Process. , 27(2), 113-120

[5] M. Beroua et al., "Enhancement of Speech Corrupted by Acoustic Noise", Proc. ICASSP, pp.208-211, 1979

[6] P.Lockwood et al., "Experiments with a Non linear Spectral Subtractor ...", Speech &"communication 11: 215-228, 1992

[7] Wiener, N. (1949), Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications, Cambridge, MA: MIT Press.

[8] Lim, J. and Oppenheim, A.V. (1978), All-pole modeling of degraded speech, IEEE Trans, Acoust. Speech Signal Process., 26(3), 197-210

[9] MacAulay, R.J, and Malpass, M.L. (1980), speech enhancement using soft-decision noise suppression filter, IEEE Trans, Acoust. Speech Signal Process., 28, 137-145.

[10] Ephraim and Malah, D. (1984), Speech enhancement using a minimum mean square error short time spectral amplitude estimator, IEEE Trans, Acoust. Speech Signal Process. 32(6), 1109-1121.

[11] Y. Hu and P. Loizou, “Subjective comparison of speech enhancement algorithms,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2006, vol. 1, pp. 153–156.

[12] Y. Hu and P. Loizou, “Subjective comparison and evaluation of speech enhancement algorithms,” Speech Commun., vol. 49, pp. 588–601, 2007.

[13] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, “Perceptual evaluation of speech quality (PESQ)-A new method for speech quality assessment of telephone networks and codecs,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. 2001, vol. 2, pp. 749–752.

[14] J. Hansen and B. Pellom, “An effective quality evaluation protocol for speech enhancement algorithms,” in Proc. Int. Conf. Spoken Lang. Process., 1998, vol. 7, pp. 2819–2822

[15] Hansen, J. and Clements, M. (1991), Constrained iterative speech enhancement with application to speech recognition, IEEE Trans Signal Process., 39(4), 795-805.