Enhancement of Noisy Speech Signals for Hearing Aids

Anuradha R. Fukane, Shashikant L. Sahare
Electronics and Telecommunication department,
Cummins College of Engineering for Women,
Pune, Maharashtra, India
E-mail: anuraj110@rediffmail.com, shashikantsahare@rediffmail.com

***Abstract*—it is very difficult to understand speech signals in presence of background noise for the normal listeners and hearing impaired persons. The human speech and hearing organ is inherently sensitive to interfering noise. Interfering noise decreases speech intelligibility and quality which makes speech communication troublesome and causing human mistakes. Speech signals from the uncontrolled environments may contain degradation components along with the required speech components. Speech enhancement algorithms removes or reduces the noise and improve one or more perceptual aspects of noisy speech most notably quality and intelligibility. This paper reports a performance evaluation of Spectral subtraction algorithm and its modified versions for Hearing aids in different environments such as restaurant, train and Car environments. Clean speech signals are corrupted by background noise respectively multi-talker babble noise, train noise, and car engine noise at four different signal-to-noise ratio levels -2dB, 0dB, 5dB,10dB. Subjective and objective type evaluation of enhanced speech signals were carried out. The evaluation of intelligibility and quality of enhanced speech is reported for hearing Aids. Section I introduction, section II explains basic block diagram of speech enhancement, explains basic Spectral subtraction speech enhancement algorithm in detail and its various modified versions of spectral subtraction algorithm, section III explains performance evaluation of this speech enhancement algorithms in terms of quality and intelligibility for Hearing aid, section- IV conclusion.**

Keywords- speech enhancement; spectral subtraction, noise estimation , DFT, intelligibility.

I. INTRODUCTION

Speech signals from the uncontrolled environment may contain degradation components along with required speech components. The degradation components include background noise, speech from other speakers etc. Speech signal degraded by additive noise, this make the listening task difficult for a direct listener and gives poor performance in automatic speech processing tasks like speech recognition speaker identification, hearing aids, speech coders etc. The degraded speech therefore needs to be processed for the enhancement of speech components. The aim of speech enhancement is to improve the quality and intelligibility of degraded speech signal. Improving quality and intelligibility of speech signals, reduces listener's fatigue. Quality can be measured in terms of signal distortion but intelligibility and pleasantness are difficult to measure by any mathematical algorithm. Perceptual quality and intelligibility are two measures of speech signals. In this study, a speech enhancement algorithm using spectral subtraction and its modified versions of algorithm are proposed for digital hearing aids. IEEE [7] and NOIZOU [7] database are used for testing.

II. SPEECH ENHANCEMENT

Speech enhancement can be achieved by using Speech enhancement Algorithms. Block Diagram is shown in figure.1 Input signal is noisy speech signal from IEEE data base and NOIZOU database Noisy speech signal is segmented for 20-30ms and windowed. Hamming window is used for windowing. Then Discrete Fourier transform (DFT) of segmented and windowed Noisy speech signal is taken. DFT of noisy signal is then given to noise estimation block and speech enhancement block. Noise estimation block estimate the noise during the pauses and find the noise spectrum. In most speech-enhancement algorithms, it is make assumed that an estimate of the noise spectrum is available. The noise estimate can have a major impact on the quality and intelligibility of the enhanced signal. If the noise estimate is too low, unwanted residual noise will be audible, if the noise estimate is too high, speech will be distorted. The simplest approach is to estimate and update the noise spectrum during the silent segments of the signal. Speech enhancement block enhance the noisy speech spectrum to generate clean speech spectrum.

A. Spectral subtractive algorithm

The spectral - subtractive algorithm is historically one of the first algorithms proposed for noise reduction [7]. Simple and easy to implement, based on the principle that one can obtain an estimate of the clean signal spectrum by subtracting an estimate of the noise spectrum from the noisy speech spectrum. The noise spectrum can be estimated, and updated, during the periods when the signal is absent or when only noise is present. Assumption is noise is additive¸ its spectrum does not change with time means noise is stationary or it's slowly time varying signal¸ whose spectrum does not change significantly between the updating periods [3]. Let y(n) be the noise corrupted input speech signal which is composed of the clean speech signal x(n) and the additive noise signal d(n). In mathematical equation form we can write y(n) in time domain and Fourier domain as given in equation one and two respectively

$$y(n) = x(n) + d(n) \quad (1)$$

$$Y[w] = x[w] + D[w] \quad (2)$$

978-0-7695-4437-3/11 $26.00 © 2011 IEEE
DOI 10.1109/CSNT.2011.105

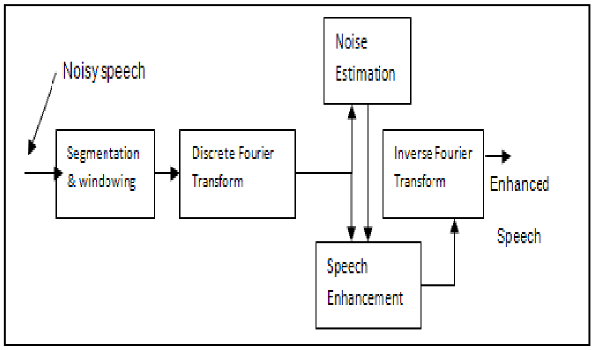

Figure-1 Block Diagram of speech enhancement

Y[w] can be expressed in terms of Magnitude and phase as

$$Y[w] = Y |(w)| e^{j Ø y}$$

Where |Y(w)| is the magnitude spectrum and Ø is the phase spectra of the corrupted noisy speech signal. Noise spectrum in terms of magnitude and phase spectra as

$$D[w] =| D[w] | e^{j Ø y}$$

The magnitude of noise spectrum |D(w)| is unknown but can be replaced by its average value computed during non speech activity i.e. during speech pauses. The noise phase is replaced by the noisy speech phase $Ø_y$ that does not affect speech ineligibility [7]. We can estimate the clean speech signal simply by subtracting noise spectrum from noisy speech spectrum, in equation form

$$Xe (w) = [|Y(w)|-|D(w)| |] e^{j Ø y} \quad (3)$$

Where Xe(w) is estimated clean speech signal. Many spectral subtractive algorithms are there depending on the parameters to be subtracted such as magnitude (Amplitude) spectral subtraction, power spectral subtraction, autocorrelation subtraction. The estimation of clean speech magnitude signal spectrum is

$$Xe[w] = |Y[w]| - |D[w]|$$

Similarly for Power' spectrum subtraction is

$$Xe[w]^2 = |Y[w]|^2 - |D[w]|^2 \quad (4)$$

The enhanced speech signal is finally obtained by computing the inverse Fourier Transform of the estimated clean speech | Xe[w]| for magnitude. Spectrum subtractions and |Xe[w]|2 for power spectrum substation subtraction, using the phase of the noisy speech signal. The more general version of the spectral subtraction algorithms is

$$Xe[\omega]^p = |Y[\omega] |^p - |De[\omega] |^p \quad (5)$$

Where P is the power exponent, when P=1 yielding the magnitude spectral subtraction algorithm. The general form

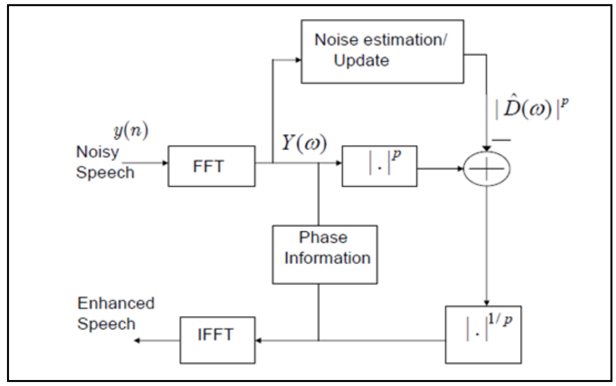

Figure2-The general form of the spectral subtraction algorithm [7]

of the spectral subtraction and P=2 yielding the power subtraction algorithm. The general form of the spectral subtraction algorithm is shown in figure 2. The spectral subtraction algorithm is computationally simple as it only involves a forward and inverse Fourier Transform.

1) Issues Related to Spectral subtractive algorithm

The simple subtraction processing comes with a price. The subtraction process needs to be done carefully to avoid any speech distortion. If too much is subtracted, then some speech information might be removed as well, while if too little is subtracted then much of the interfering noise remains [1]. It is clear from equation three that spectral subtraction method can lead to negative values, resulting from differences among the estimated noise and actual noise frame. Simple solution is set the negative values to zero, to ensure a non negative magnitude spectrum. This non linear processing of the negative values called negative rectification or half-wave rectification [4]. This ensure a non-negative magnitude spectrum given by equation

$$|Xe(\omega)| = |Y(\omega)| - | De(\omega)|, \text{ if } |Y(\omega)| > |De(\omega)| \text{ else}$$

$$= 0 \quad (6)$$

This non-linear processing of the negative values creates small isolated peaks in the spectrum occurring at random frequency locations in each frame as shown in figure3. When converted in the time-domain, these peaks sound like tones with frequencies that change randomly from frame to frame. That is, tones that are turned on and off at the analysis frame rate (every 20 to 30 ms). This new type of noise introduced by the half-wave rectification process has been described as warbling and of tonal quality, and is commonly referred to in the literature as musical noise. Minor shortcoming of the spectral subtraction algorithm is the use of noisy phase that produces a roughness in the quality of the synthesized speech [4]. The phases of the noise - corrupted signal are not enhanced, because the presence of noise in the phase information does not contribute much to the degradation of speech quality [3]. Combating musical noise is much more critical than finding methods to preserve the original phase.

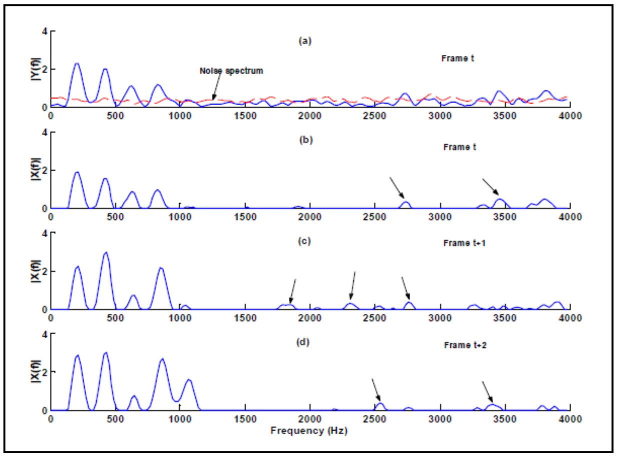

Figure 3 effects of half -wave rectification [12]

For that reason, much effort has been focused on finding methods to reduce musical noise.

B. Spectral Subtraction with over subtraction

This Method consists of subtracting an overestimate of the noise power spectrum and presenting the resultant spectral components from going below a preset minimum spectral floor value. This algorithm is given in equation (7), where $|Xe_j(\omega)|$ denotes the enhanced spectrum estimated in frame j and $|De(\omega)|$ is the spectrum of the noise obtained during non-speech activity.

$$|Xe_j(\omega)|^2 = |Y_j(\omega)|^2 - |De(\omega)|^2, \text{ if } |Y_j(\omega)|^2 > (\alpha + \beta)|De(\omega)|^2 \text{ else}$$
$$= \beta|De(\omega)|^2 \quad (7)$$

With $\alpha \geqslant 1$ and $D < \beta \leqslant 1$. Where α is over subtraction factor and β is the spectral floor parameter. Parameter β controls the amount of residual noise and the amount of perceived Musical noise. If β is too small, the musical noise will became audible but the residual noise will be reduced. If β is too large, then the residual noise will be audible but the musical issues related to spectral subtraction reduces. Parameter α affects the amount of speech spectral distortion. If α is too large then resulting signal will be severely distorted and intelligibility may suffer. If α is too small noise remains in enhanced speech signal. When $\alpha > 1$, the subtraction can remove all of the broadband noise by eliminating most of wide peaks. But the deep valleys surrounding the peaks still remain in the spectrum [1]. The valleys between peaks are no longer deep when $\beta > 0$ compared to when $\beta = 0$ [4]. Berouti found that speech processed by equation (7) had less musical noise. Experimental results showed that for best noise reduction with the least amount of musical noise, α should be smaller for high SNR frames and large for low SNR frames.

C. Non – linear Spectral Subtraction (NSS)

The NSS proposed by [5] Lockwood and Boudy. NSS is basically a modification of the method suggested in [4] by making the over subtraction factor frequency dependent and the subtraction process non-linear. In case of NSS assumption is that noise does not affects all spectral components equally. Certain types of noise may affect the low frequency region of the spectrum more than high frequency region. This suggests the use of a frequency dependent subtraction factor for different types of noise. Due to frequency dependent subtraction factor, subtraction process becomes nonlinear. Larger values are subtracted at frequencies with low SNR levels and smaller values are subtracted at frequencies with high SNR levels. The subtraction rule used in the NSS algorithm has the following form.

$$|Xe(\omega)| = |Y(\omega)| - \alpha(\omega) N(\omega) \quad \text{if}$$
$$|Y(\omega)| > \alpha(\omega)N(\omega) + \beta|\check{D}e(\omega)| \text{ else}$$
$$= \beta|Y(\omega)| \quad (8)$$

Where β is the spectral floor set to 0.1 in [8] $|Y(\omega)|$ and $|\check{D}e(\omega)|$ are the smoothed estimates of noisy speech and Noise respectively, $\alpha(\omega)$ is a frequency dependent subtraction factor and $N(\omega)$ is a non-linear function of the noise spectrum where

$$N(\omega) = \mathrm{Max}(|De(\omega)|) \quad (9)$$

The N (ω) term of obtained by computing the maximum of the noise magnitude spectra $|De(\omega)|$ over the part 40 frames. [4] The $\alpha(\omega)$ given as [8].

$$\alpha(\omega) = 1/r + p(\omega) \quad (10)$$

Where γ_r is a scaling factor and $P(\omega)$ is the square root of the posteriori SNR estimate given as

$$P(\omega) = |Y(\omega)| / |\check{D}e(\omega)| \quad (11)$$

The NSS algorithm was successfully used in [8] as a pre-processor to enhance the performance of speech recognition systems in noise.

D. Multiband Spectral Subtraction (MBSS)

In MBSS approach [6,4] the speech spectrum is divided into N overlapping bands and spectral subtraction is performed independently in each band. The processes of splitting the speech signal into different bands can be performed either in the time domain by using band pass filters or in the frequency domain by using appropriate windows. The estimate of the clean speech spectrum in the i^{th} band is obtained by [6].

$$|Xe_i(\omega_k)|^2 = |Yi(\omega_k)|^2 - \alpha_i \delta_i |Di(\omega_k)|^2 \quad (12)$$
$$b_i < \omega_k < e_i$$

Where $\omega_k = 2pi\ k/N$, $k = 0, 1, \ldots, N-1$ are the discrete frequencies $|De_i(\omega k)|^2$ is the estimated noise power spectrum obtained during speech absent segment, α_i is the over subtraction factor of the i^{th} band and δ_i is an additional band. Subtraction factor can be individually set for each frequency band to customize the noise removal processor b_i and e_i are the beginning and ending frequency bins of the i^{th} frequency band. The band specific over subtraction factor is a function of the segmented SNR_i of the i^{th} frequency band. The main difference between the MB and the NSS algorithm is in the estimation of the over subtraction factors. The MB approach estimates one subtraction factor for each frequency band, whereas the NSS algorithm estimates one subtraction factor for each frequency bin [4]. The performance of the multiband spectral subtraction algorithm [9] was evaluated by Hu Y. and Loizou [12, 13] using formal subjective listening tests conducted according to ITU–T P.835 [14].

III. PERFORMANCE OF SPECTRAL SUBTRACTION ALGORITHMS

The spectral subtraction algorithm was evaluated primarily using objective quality measures [13] such as SNR, segmental SNR, Log likehood ratio (LLR), Itakura-Satio(IS), Ceptrum distances(CEP) and then subjective listening tests were carried out. The intelligibility and speech quality measures reflect the true performance of speech enhancement [4] algorithms in realistic scenarios. Results are mentioned as follows when SNR of -2dB no algorithm performed well.Figure-4 shows spectrogram of clean speech signal, noisy speech signal corrupted by car noise having SNR 0dB and enhanced speech signal using nonlinear Spectral subtraction. Figure-5 shows spectrogram of noisy speech signal corrupted by Train noise having SNR 0dB and enhanced speech signal using Spectral subtraction with over subtraction factor. Figure-6 shows spectrogram of noisy speech signal corrupted by Babble noise having SNR 0dB and enhanced speech signal using Multiband Spectral Subtraction. We performed intelligibility and quality measurement tests using the Diagnostic Rhyme Test (DRT). Result indicated that SS did not decrease speech intelligibility but improved speech quality particularly in the area of pleasantness and inconspicuousness of the background noise.

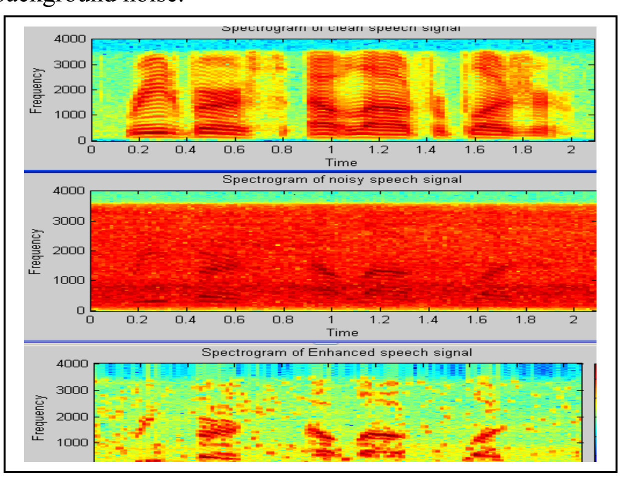

Figure-4 spectrogram of clean speech signal, noisy speech signal having SNR 0dB car noise and Enhanced speech signal using non linear Spectral Subtraction.

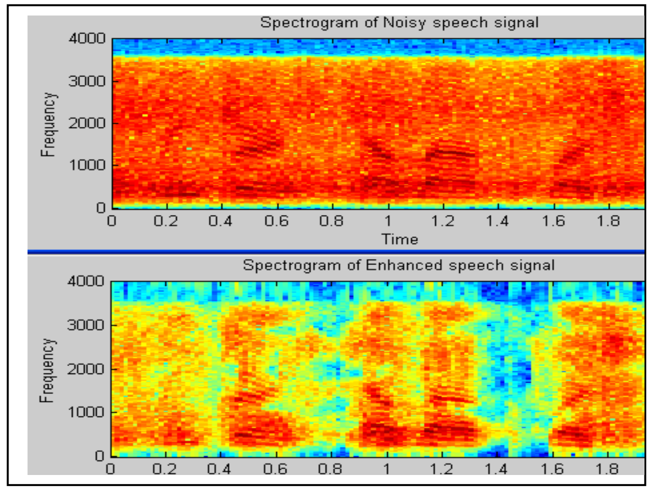

Figure-5 spectrogram of noisy speech having SNR 0dB Train noise and Enhanced speech signal using Spectral Subtraction with over subtraction factor

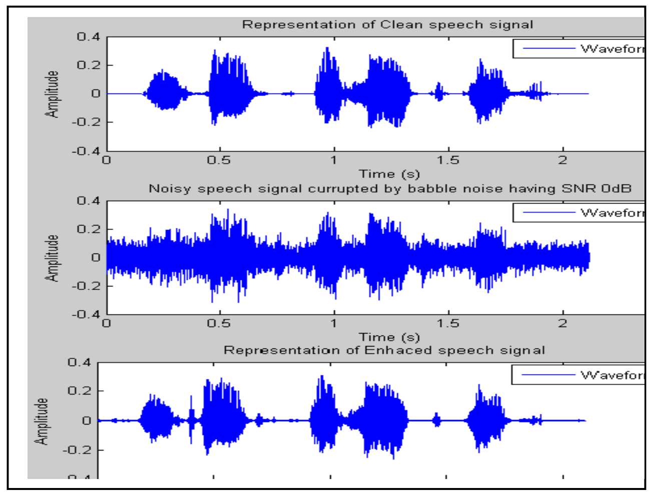

Figure-6 representation of noisy speech having SNR 0dB, Babble noise and Enhanced speech signal using Multiband Spectral Subtraction

The largest improvement in speech quality was noted for relatively stationary noise sources [4, 2] for train and car noise and less improvement in quality for non stationary noise that is for multi-talker babble noise. The performance of the multiband spectral subtraction algorithm [9] was evaluated by using formal subjective listening tests conducted according to ITU–T P.835 [14]. Results indicated that the MBSS algorithm performed the best consistently across all noise conditions, [4] in terms of overall quality. In terms of noise distortion the MBSS algorithms performed well, except in 5dB train conditions

IV. CONCLUSION

Various spectral subtraction algorithms proposed for speech enhancement were described in above sections. These algorithms are computationally simple to implement as they involve a forward and an inverse Fourier transform. The simple subtraction processing comes at a price. The subtraction of the noise spectra from the noisy spectrum introduces a distortion in the signal known as Musical noise [4]. We studied different techniques that mitigated the musical noise distortion. Different methods were proposed for computing the over subtraction factor based on different criteria that included linear [5] and nonlinear functions [8] of the spectral SNR of individual frequency bins or bands [6]. Use of spectral floor parameter prevents the resultant spectral components from going below a preset minimum value. The spectral floor value controlled the amount of remaining residual noise and the amount of musical noise [4]. Evaluation of spectral subtractive algorithms revealed that these algorithms [4] improve speech quality and not affect much more on intelligibility of speech signal. Speech signals corrupted by Car engine noise and Train noise get less affected than Multi-talker Babble noise, hence easily get enhanced. Noisy Speech signals SNR plays an important role in speech Enhancement, if SNR is less than 0dB no algorithm performed well. Effect on formant frequencies is also observed. F1 is the most affected by train and car noise and F5 is less affected. Female speech signal is less affected by train and car noise than male speech signal. Train noise has more impact on Formant SNR than car noise. Result indicated that Spectral Subtractive Algorithms does not decrease speech intelligibility but improved speech quality particularly in the area of pleasantness in presence of the background noise. so spectral subtractive algorithms are suitable for Hearing Aids in different noisy Environments with their modified versions.

ACKNOWLEDGMENT

Anuradha R. Fukane thanks to Dr. S. D. Bhide and Dr. Madhuri Khambete for their valuable guidance and support.

REFERENCES

[1] Yi Hu and Philipos C. Loizou, “Subjective comparison and evaluation of speech enhancement algorithms” *IEEE Trans. Speech Audio Proc.*2007:49(7): 588–601.

[2] Berouti,M. Schwartz,R. and Makhoul,J.,"*Enhancement of Speech Corrupted by Acoustic Noise*", *Proc ICASSP 1979, pp208-211*,.

[3] Paliwal K. and Alsteris L.(2005), “On usefulness of STFT phase spectrum in human listening tests ” *Speech Commmun.45(2),153-170*

[4] Boll,S.F.,"Suppression of Acoustic Noise in Speech using Spectral Subtraction", *IEEE Trans ASSP 2*7(2):113-120, April 1979

[5] Lockwoord, P. and Boudy,J.,"Experiments with a Nonlinear Spectral Subtractor (NSS), Hidden Markov Models and the projection, for robust speech recognition in cars", *Speech Communication, 11, pp215-228, Elsevier 1992.*

[6] Kamath S. and Loizou P.(2002) “A multiband spectral subtraction methode for enhancing speech currupted by colored noise” *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing*

[7] Hu Y., Bhatnager M. Loizou P.(2001) “A crosscorellation technique for enhancing speech currupted with correlated noise” . *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing1.pp 673-676*

[8] Sim B, Tong Y, chang J., Tan C.(1998) ‘A parametric formulation of the generalized spectral subtraction method ’ *IEEE. Trans. Speech Audio Process,6(4), 328-337*

[9] He C, and Zweig G. (1999) “Adaptive two band spectral subtraction with multiwindow spectral estmation” *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing ,2, pp 793-796*

[10] Lebart K, Boucher J M,(2001)” A New method based on spectral subtraction for speech enhancement” *Acta acustica, Acustica vol. 87 pp359-366..*

[11] R. Martin, “Spectral Subtraction Based on Minimum Statistics,” in *Proc. Euro. Signal Processing Conf. (EUSIPCO), pp. 1182–1185, 1994*

[12] Phillips C Loizou “*Speech enhancement theory and practice*” 1st ed. Boca Raton, FL.: CRC, 2007. Releases Taylor & Francis

[13] Yi Hu and Philipos C. Loizou, Senior Member, IEEE “Evaluation of Objective Quality Measures for Speech Enhancement” *IEEE, Trans. on Audio, Speech, and Language pross.vol 16, (2008)*

[14] ITU-T(2003) “subjective test Methodalogy for evaluating speech communication system that include noise supression slgorithm.” *ITU-T recommendation p.835*