International Journal For Technological Research In Engineering
Volume 1, Issue 9, May-2014 *ISSN (Online): 2347 - 4718*

IMPLEMENTATION OF SPEECH ENHANCEMENT ALGORITHM

Nakul Mane[1], Mahendra Phule[2], Saurabh Shelke[3], Mrs. H. D. Shinde[4] (Assistant Professor)
Deptartment of Electronics and Telecommunication
AISSMS's Institute of Information Technology, University of Pune
Pune, Maharashtra, India.

Abstract: Enhancement of speech deals with the improvement in the value or quality of something. When this is applied to speech, this simply means the improvement in quality and intelligibility of degraded speech signal by using signal processing tools. Quality is subjective and it reflects the individual references of listeners while intelligibility is objective and it gives percentage of words that could be correctly identified by the listeners. Speech enhancement is concerned with improving some perceptual aspects of speech that has been degraded by noise at various stages of speech communication. There are various methods of enhancing the quality of speech signal. But, Ideal binary masking is technique used to improve intelligibility of speech signal. In this project, we will study various speech enhancement methods for quality improvement and intelligibility improvement. Aim of this project is to implement Ideal Binary Masking algorithm to have quality as well as intelligibility improvement.

I. INTRODUCTION

The Enhancement of something is the improvement of it in relation to its value or quality. Enhancement of speech deals with the improvement in the value or quality of something when applied to speech. Speech enhancement is concerned with improving some perceptual aspects of speech that has been degraded by additive noise. In most applications, the aim of speech enhancement is to improve the quality and intelligibility of degraded speech. When we are reducing background noises, it also introduces some speech distortions which harm intelligibility of speech. Hence, the main challenge is to design effective speech enhancement algorithms which will suppress noise without introducing much distortion. Speech quality is highly subjective in nature and can be easily improved, at least to some degree, by suppressing the background noise. In contrast, intelligibility is related to the underlying message or content of the spoken words and can be improved only by suppressing the background noise without distorting the underlying target speech signal. One of the methods in Computational Auditory Scene Analysis (CASA) is Ideal Binary Masking (IdBM) (Gibak Kim, Yang Lu,, Yi Hu and Philipos C. Loizou, 2008, 2009) which is mainly used to improve intelligibility of the speech. This method retains the time-frequency T-F regions of the target signal that are stronger than the interfering noise masker, and removes the regions that are weaker than the interfering noise. This algorithm decomposes the input signal into T-F regions with use of an auditory filter bank and then classifies the signal into target dominated spectro-temporal regions and masker dominated spectro-temporal regions, which is to be removed. Amplitude Modulation Spectrogram (AMS) (Kollmeier and Koch, 1994) was used as a feature in classification.

II. OVERVIEW

The need to enhance speech signals are required in many situations in which the speech signals are occurring or originating from noisy location or it is affected by noise over the communication channel. There are various instances where it is required to enhance the speech signal. In voice communication, for telephone systems suffering from background noises present in car or restaurants at the transmitting end, speech enhancement algorithms can be used to improve the quality of speech at the receiving end i.e. they can be used as preprocessors in speech coding systems. When we are reducing background noises, it also introduces some speech distortions which harm intelligibility of speech. Hence, the main challenge is to design effective speech enhancement algorithms which will suppress noise without introducing much distortion. The presence of background interferences causes the degradation of speech quality and intelligibility. Noisy environment also reduces listener's ability to understand what is said. The purpose of enhancement methods is to reduce background noise, improve speech quality. Speech enhancement is very difficult problem for two reasons. One, the nature and characteristics of the noise signals can change dramatically in time and application to application. It is therefore very difficult to find versatile algorithm that really work in different practical environment. Second, the performance measure can also be defined differently for each application. The algorithm which has been used here is based on CASA (Computational Auditory Scene Analysis) system. CASA is defined as the study auditory scene analysis by computational means. The goal of CASA is to produce separate streams from auditory input, such that each stream represents a single sound source in the acoustic environment (Bregman [13]). Wang [132] proposed that the goal of CASA should be to estimate the ideal time-frequency (T-F) mask. Consider a time frequency representation such as the spectrogram shown in figure 1 in which the frequency axis and time axis are divided into discrete units. An ideal T-F mask is then a binary matrix, whose value is one for a T-F unit in which

the target energy is stronger than the total interference energy, and is zero otherwise.

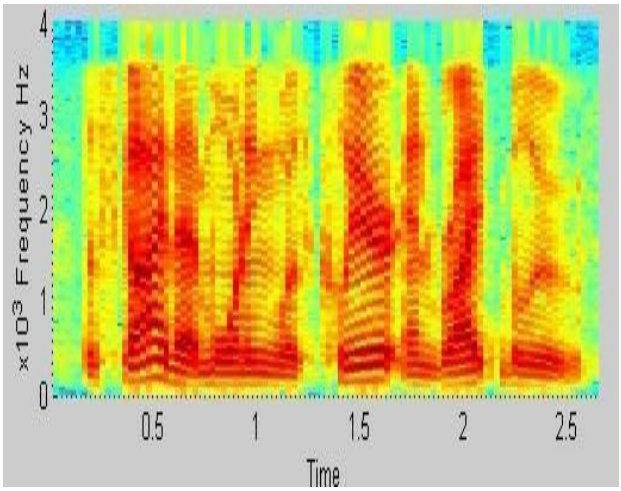
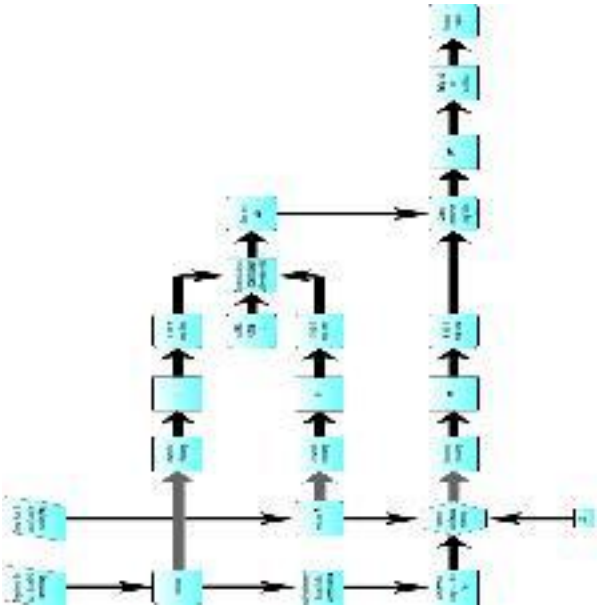


Fig.1: Spectrogram

CASA system emphasize mainly on improvement of intelligibility. It consists of three methods pitch tracking, ideal binary mask and Onset-Offset detection. Out of these methods here the ideal binary mask algorithm is implemented which is mainly concerned about improvement in ineligibility of the speech signal. Figure2 is the block diagram of the Ideal binary mask algorithm. An algorithm is proposed that decomposes the input signal into time-frequency (T-F) units and makes binary decisions, based on a Bayesian classifier, as to whether each T-F unit is dominated by the target or the masker. Speech corrupted at low signal-to-noise ratio (SNR) levels (−5 and 0 dB) using different types of maskers is synthesized by this algorithm and presented to

Fig.2: Block Diagram



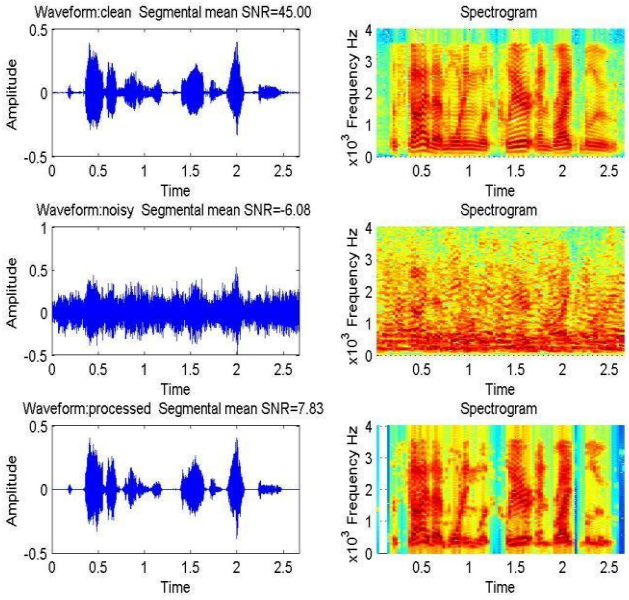


Fig.3: Clean signal, Noisy speech signal and processed signal waveforms and spectrogram

Results indicated substantial improvements in intelligibility (over 60% points in −5 dB babble) over that attained by human listeners with unprocessed stimuli. The findings from this study suggest that algorithms that can estimate reliably the SNR in each T-F unit can improve speech intelligibility.

III. IDEAL BINARY MASK

The goal of this study is to evaluate the intelligibility of speech synthesized via an algorithm that decomposes the input signal into T-F regions, with the use of crude auditory like filter bank, and uses a simple binary Bayesian classifier to retain target-dominated spectro-temporal regions while removing masker-dominated spectro-temporal regions. Amplitude modulation spectrograms (AMSs), (Kollmeier and Koch, 1994) were used as features for training Gaussian mixture models (GMMs) to be used as classifiers. The present work tests the hypothesis that algorithms that make use of knowledge of when the target is stronger than the masker (at each T-F unit) can improve speech intelligibility in noisy conditions. In the training stage, features are extracted, typically from a large speech corpus, and then used to train two GMMs representing two feature classes: target speech dominating the masker and masker dominating target speech. AMS are used in this work as features, as they are neuro-physiologically and psycho-acoustically motivated (Kollmeier and Koch, 1994; Langner and Schreiner, 1988). In the enhancement stage, a Bayesian classifier is used to classify the T-F units of the noise-masked signal into two classes: target-dominated and masker dominated. Individual T-F units of the noise-

masked signal are retained if classified as target-dominated or eliminated if classified as masker-dominated, and subsequently used to reconstruct the enhanced speech waveform. Following is the example of the IBM algorithm as shown in figure 3. The figure shows the clean speech signal and its spectrogram, noisy signal and its spectrogram and finally processed signal after applying mask and its spectrogram. The effect of LC on mask is studied as shown in table.

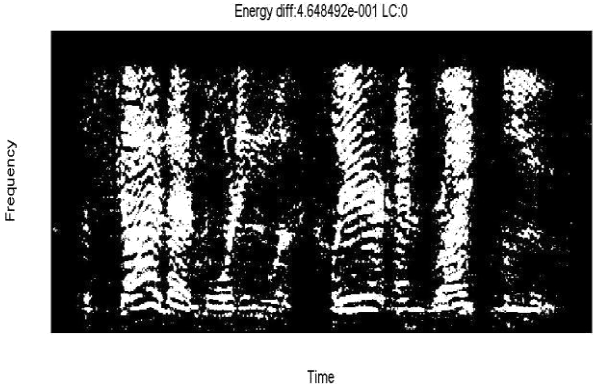


Fig. 4: Ideal binary mask (SNR=0, LC=0) Figure shows ideal binary mask. Also we have compared masks with different SNR criteria (LC)

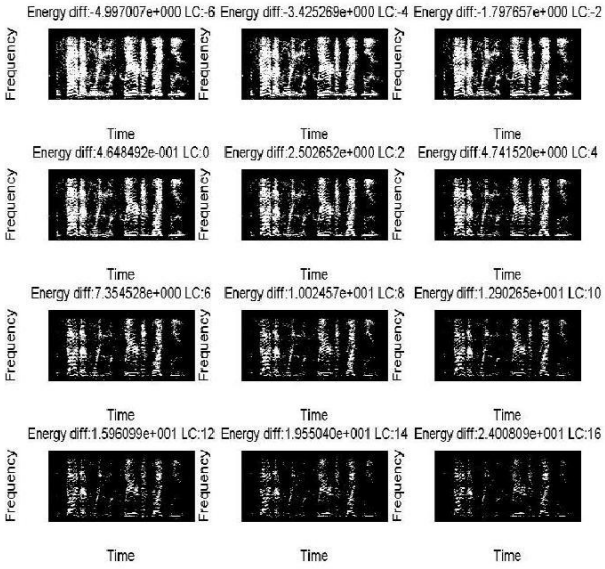


Fig. 5: Compassion of ideal masks with different SNR criteria (LC). Figure shows one of that compassion in which we can see the changes in ideal mask as we change the SNR criteria.

Table 1 Mask difference in % Table shows the mask difference in % as compared with reference mask

SNR	**LC** -6	-4	-2	0	2	4	6	8	10	12	14	16
-6	REF	-2.66	-5.25	-7.55	-9.68	-11.51	-13.11	-14.47	-15.77	-16.78	-17.59	-18.22
-4	2.93	REF	-2.66	-5.25	-7.55	-9.68	-11.51	-13.11	-14.47	-15.77	-16.78	-17.59
-2	5.97	2.93	REF	-2.66	-5.25	-7.55	-9.68	-11.51	-13.11	-14.47	-15.77	-16.78
0	9.22	5.97	2.93	REF	-2.66	-5.25	-7.55	-9.68	-11.51	-13.11	-14.47	-15.77
2	12.51	9.22	5.97	2.93	REF	-2.66	-5.25	-7.55	-9.68	-11.51	-13.11	-14.47
4	15.85	12.51	9.22	5.97	2.93	REF	-2.66	-5.25	-7.55	-9.68	-11.51	-13.11
6	19.28	15.85	12.51	9.22	5.97	2.93	REF	-2.66	-5.25	-7.55	-9.68	-11.51
8	22.72	19.28	15.85	12.51	9.22	5.97	2.93	REF	-2.66	-5.25	-7.55	-9.68
10	26.22	22.72	19.28	15.85	12.51	9.22	5.97	2.93	REF	-2.66	-5.25	-7.55
12	29.66	26.22	22.72	19.28	15.85	12.51	9.22	5.97	2.93	REF	-2.66	-5.25
14	33.10	29.66	26.22	22.72	19.28	15.85	12.51	9.22	5.97	2.93	REF	-2.66
16	36.42	33.10	29.66	26.22	22.72	19.28	15.85	12.51	9.22	5.97	2.93	REF

IV. CONCLUSION

Ideal Binary Masking algorithm can improve quality as well as intelligibility. But its performance is dependent on many factors as segmentation method used, length and type of window used in time domain, but the most important factor is the correct estimation LC value more it is near to SNR value at which noise is added better are the results of the enhancements. This can be observed from the above table.

REFERENCE

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

[2] Boll, S.F. (1979), Suppression of acoustic noise in speech using spectral subtraction, IEEE Trans. Acoust. Speech Signal Process. 28, 137-145.

[3] Dendrinos, M., Bakamides, S., and Caravans, G. (1991), Speech Enhancement from noise: a regenerative approach, Speech Commun., 10, 45-57.

[4] Ephraim, Y. and Van Trees, H.L. (1993), A signal subspace approach for speech enhancement, Proc. IEEE Int. Conf. Acoust. Speech Signal Process. II, pp. 355-358.

[5] Mc Aulay, R.J. and Malpass, M.L> (1980), Speech enhancement using a soft decision noise suppression

filtr, IEEE Trans. Acoust. Speech Signal Process. 28, 137-145.

[6] Ephraim. Y. 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.

[7] Lim, J. and Oppenheim, A.V. (1979), Enhancement and bandwidth compression of noisy speech, Proc. IEEE, 67(12), pp. 1586-1604.

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

