# *Implementation of Advanced Speech Enhancement System on TMS320C6713 DSK system*

Nikitha S.R.
M. Tech, 4th Semester,
Department of IT
Dayananda Sagar College of Engineering,
India, Bangalore
snikitha33@yahoo.com

Gopalaiah
Assistant Professor
Department of IT
Dayananda Sagar College of Engineering
India, Bangalore
gopaliah@gmail.com

***Abstract*– The performance of a speech communication system can degrade in the presence of acoustic background noise and quantization noise. Due to their different nature, these two problems have been addressed independently. The perceived quantization noise is typically reduced by means of a post-filter. Background noise is attenuated by noise suppression systems.**

**In this project, TIA127-B compliant (Narrow Band Speech Enhancement System) Noise Suppression systems are to be simulated using MATLAB and is implemented on TMS320C6713 DSK to demonstrated in real time.**

## I. INTRODUCTION

In modern hands free speech communication environments often occurs the situation that the speech signal is superposed by background noise (see Fig.1). This is particular the case if the speaker is not located as close as possible to the microphone. The speech signal intensity decreases with growing distance to the microphone. It is even possible that background noise sources are captured at a higher level than the speech signal. The noise distorts the speech and words are hardly intelligible. In order to improve the intelligibility and reduce the listeners (FES) stress by increasing the signal to noise ratio a noise reduction procedure also called speech enhancement algorithm is applied.

Historically, pre-processor single-channel speech enhancement algorithms have been considered in the context of robust speech coding, (see Fig. 2). These algorithms are designed to operate in an environment where only the noisy signal is available, and both facilitate the operation of the speech codec (coding and decoding) and improve the perceived sound quality at the end user.

Acoustic background noise in mobile speech communication systems, while largely inevitable, can have a severely detrimental effect on speech intelligibility. Noise suppression is highly desirable in these systems. However, the process of reducing noise in a speech signal is associated with distortion of the processed signal, the severity of which is generally proportional to the amount of noise suppression applied.

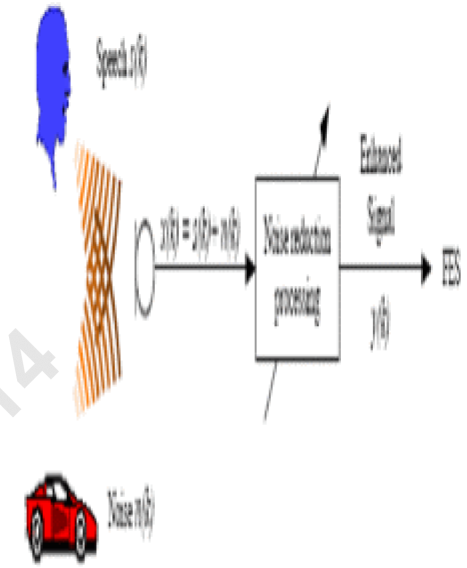



Fig.1. Speech signal superposed by background noise

In a single-channel application, the noise suppression algorithm requires an additional module for the estimation of the noise and clean speech statistics. The underlying idea in all these algorithms is that the noise statistics can be estimated from the signal segments, either in the time or in the frequency domain, where the speech energy is either low, or the speech signal is not present at all.

The classical noise suppression scheme is based on the idea of spectral subtraction. It is widely used nowadays, mainly because of its simplicity. Spectral subtraction schemes are based on direct estimation of the short time spectral magnitude of clean speech. A drawback of this algorithm is the musical noise. Musical noise consists of tones with the same duration as the window length of algorithm and with a different set of frequencies for each frame. Musical noise is a result of variability in the power spectrum.

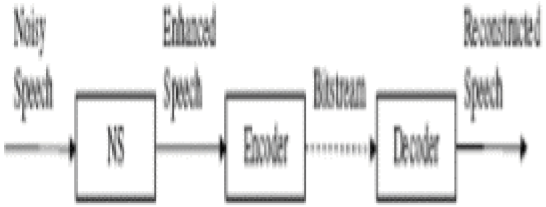



Fig. 2. Configuration of Noise Suppression (NS) as a speech enhancement Pre-processor for speech codec

## II. NOISE REDUCTION PRINCIPLES

The requirements of a noise reduction system for speech enhancement are:

- Intelligibility and naturalness of the enhanced signal
- Improvement of signal-to-noise ratio
- Short signal delay
- Computational simplicity

The quality of the enhanced signal is a diverse issue, it may be characterised by the terms intelligibility and naturalness. There are several methods for performing noise reduction, but all can be regarded as a kind of filtering. In our application speech and noise are mixed to one signal channel. They reside in the same frequency band and may have similar correlation properties. Consequently the filtering will inevitably have an effect on both the speech and the noise. Therefore it is a very challenging task to distinguish between them. Sometimes speech components can be detected as noise and thus will be suppressed as well. Especially fricatives and plusives are attenuated due to their noise-like properties.

Furthermore the residual noise characteristics should preserve the characteristics of the background noise in the recording environment. Typical single channel noise reduction algorithms add a synthetic noise, also called Musical Noise., which sounds artificial and has a disturbing effect on the listener.

Single channel noise reduction algorithms are based on the fact that the statistical properties of speech are only stationary over short periods of time whereas the noise often can be assumed to be stationary over much longer periods. Another aim for the algorithm design is the limitation of the signal delay because of its annoying effect in dialog situations.

The noise reduction algorithms can be split into two groups: time domain algorithms and those utilising some kind of transform, e.g. Fourier Transform. Whereas the filter calculation for time domain solutions generally relies on the usage of correlation estimates, there is a large variety of algorithms operating in the frequency domain.

## III. NOISE SUPPRESSION SYSTEM

The Noise suppression algorithm used by the EVRC is based on the Spectral Subtraction technique, in which the main emphasis is given on the Spectral Weighting .The Fig.3 shows the general principle of such a system.

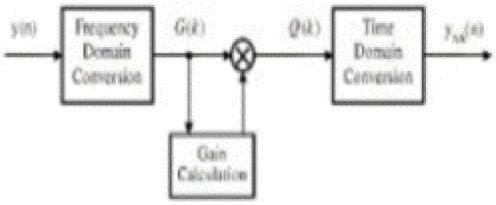



Fig.3. General Principle of the EVRC NS system

Firstly, the input signal, y(n), is block wise transformed from the time domain, to the frequency domain. Secondly, a set of gain factors, q(k), are calculated. The actual spectral subtraction takes the form of a multiplication of G(k) with the gain factors from the gain calculation, resulting in the enhanced spectrum Q(k). Lastly, this spectrum is transformed from the frequency domain, to the time domain, and the signal is block wise reassembled to form the enhanced output $y_{NS}(n)$ .

### *A. TIA 127-B (Narrow Band) Speech Enhancement System*

The fundamental concept of a frequency domain solution is spectral weighting and block processing. The architecture of such a system is presented in Figure 4. Since in a single/multi-channel approach the estimation of the noise and the weighting function can only be derived in frequency domain, the time domain input signal has to be transformed. The transformations are performed by means of standard analysis and synthesis systems operating on a frame-by-frame basis. It consists of three major components:

- the analysis/synthesis framework for time domain / frequency domain transformation
- the noise estimation
- the weighting function.

If the noise estimation equals the disturbing noise spectrum the output signal spectrum Y(n, Ωi), will be very similar to the noiseless speech spectrum S(n, Ωi).

Estimating the noise spectrum *Nest(n, Ωi )* is one of the major tasks of a noise cancelling system. Based on the above mentioned assumption that the noise part of the signal is stationary over longer periods of time than the speech part, an estimate of the noise is obtained by extracting slowly changing portions of the signal spectrum. The output frame is obtained by applying the inverse frequency transformation to the weighted enhanced spectrum *Y(n, Ωi )* and the noisy phase $\Phi_x(n, \Omega i)$.

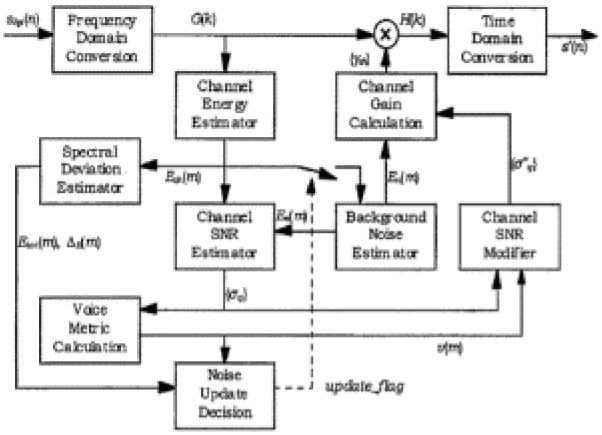



Fig . 4. TIA 127-B (Narrow Band) Speech Enhancement System.

A TIA/EIA/IS127-B Compliant Speech Enhancement System is pre-processing block in Enhanced Variable Rate Codec (EVRC) used to enhance the speech signal before encoding the speech signal. The main components of the TIA/EIA/IS127-B Compliant Speech Enhancement System are:

1. High Pass System
2. Adaptive Noise Suppression System.

High Pass System comprises 6th order Butterworth filter implemented using 3 sections of Biquad Filter. Adaptive noise suppression system consists of subsystems as shown in Fig.4.

The input to the noise suppressor are the noisy speech samples s(n) which have been previously high-pass filtered. These are passed through a pre-emphasis filter and transformed into the frequency-domain values G(k). In the frequency domain, a filtering operation is performed by multiplying G(k) by the scalar gain values Y(k) to yield H(k). The filtered spectral values H(k) are transformed back into the time domain and passed through a de-emphasis filter to provide the noise suppressed speech samples s'(n) to the speech coder.

The channel energy estimator divides this spectrum into $N_c$ channels and calculates an estimate of the signal energy in each one. The spectral deviation estimator calculates the difference between the current channel energies and an average long-term estimate. An estimated signal-to noise ratio is calculated by the SNR estimator, using the channel energy and background noise estimates. The SNR estimate is used to calculate the voice metric, which is a weighted sum which provides an estimate of the signal "quality". It is used mainly as an indication as to whether or not the current frame contains speech. When the input signal is deemed to contain no speech, the background noise estimator is updated. Under some conditions the SNR estimates are changed by the SNR modifier. Based on the (modified) SNR estimates and the background noise the gains for each channel are calculated by the channel gain calculator. These gains are then used to perform the filtering of the input signal.

The overall gain factor for the current frame, γn, is calculated according to

$$\gamma_n = \max\left\{\gamma_{min} - 10\log_{10}\left(\frac{1}{Efloor}\sum_{i=0}^{NC-1} En(m,i)\right)\right\} \quad (1)$$

Where γmin=-13 is the minimum overall gain, Efloor = 1 is the noise floor energy and En(m; i) is the estimated noise spectrum calculated during the previous frame. The dB-scale channel gains are calculated as

$$\gamma_{dB}(i) = \mu_g(\sigma''(i) - \sigma_{th}) + \gamma_n; \quad 0 \le i < N_c \quad (2)$$

Where μg=0.39 is the gain slope and $\sigma_{th}$ the SNR threshold, both constants. In Fig 4 the gain curve for a single channel resulting from following equation is plotted in comparison with the gain curve resulting from the spectral subtraction rule.

$$\sigma'' = \max\{\sigma_{th}, \sigma\} \quad (3)$$

To simulate the single channel behaviour of EVRC-NS (3) was used. As is evident, the gain curve for EVRC-NS is quite different from that of spectral subtraction. The channel gains are converted to linear scale according to

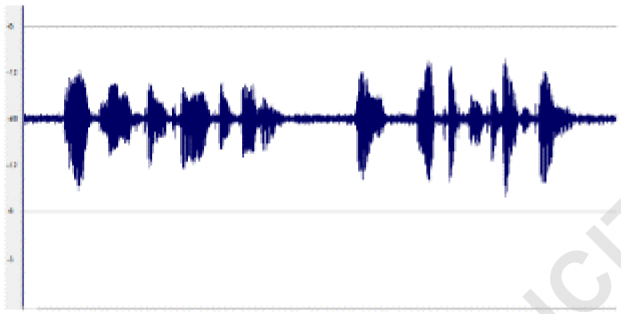
$$\gamma_{ch}(i) = \min\left\{1, 10^{\frac{\gamma_{db}(i)}{20}}\right\} \quad 0 \le i < N_c \quad (4)$$

In our implementation, the input speech is presented to the noise suppressor in frames of 80 samples (10 ms frames at 8 kHz sampling). These samples along with 24 samples of the previous frame are multiplied by a smoothed trapezoidal window and transformed into the frequency domain by a 128-point FFT In the frequency domain, the spectral values are grouped together to form 16 unequal frequency bands (similar to critical bands) referred to as channels.

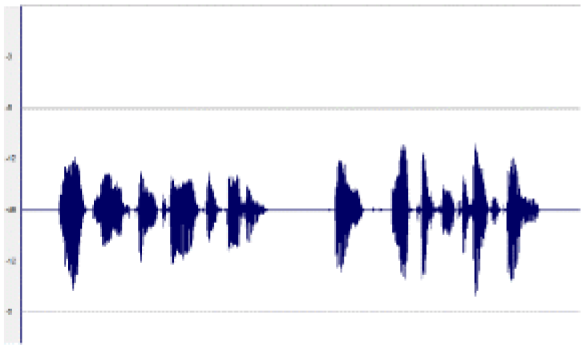
A scalar gain value is computed for each channel and applied to all the spectral values corresponding to that channel including both positive and negative frequencies. The filtered values Y (k) are transformed back into time domain using a 128-point IFFT and overlap-added with the last 48 noise-suppressed samples of the previous frame. The first 80 samples are then released to the speech coder. It is seen that the noise suppressor essentially operates as a time-adaptive filter

## VI. Simulation Results

### A. Original Speech Sample corrupted by Noise



### B. Recovered speech sample obtained from EVRC



Algorithm

| Noise Type | Noise Level | Correlation Coefficient | Segmental SNR | Log Spectral Distance | VQ based Minimum Mean Euclidean Distance |
|---|---|---|---|---|---|
| AIRPORT | 5db | 0.91452 | 26.7644 | 3.0013 | 0.0075871 |
| | 10db | 0.96744 | 32.199 | 2.8148 | 0.0054113 |
| | 15db | 0.97391 | 34.7675 | 2.5615 | 0.00469 |
| CAR | 5db | 0.9036 | 27.3531 | 3.0350 | 0.0096192 |
| | 10db | 0.96356 | 36.6884 | 2.8895 | 0.0041665 |
| | 15db | .99071 | 6.9255 | 2.8929 | 0.0099597 |

TABLE I: Object Measures For Various Noise Types

Table I presents the results of Correlation Coefficient, Segmental SNR, Log Spectral Distance, Vector Quantization based Minimum Mean Euclidean Distance values for various noise types and levels obtained by using the EVRC TIA-127-B speech enhancement system.

## VII. Conclusion

A noise suppression algorithm based on EVRC TIA/EIA/IS127-B has been proposed. The proposed algorithm continuously updates the noise estimate by noisy speech in accordance with an estimated SNR. The spectral gain is modified with the SNR so that it better fits the new noise estimate for higher speech quality.

## VIII. References

[1] J. S. Lim and A. V. Oppenheim, "All-pole modelling of degraded speech," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP- 26, no. 3, pp. 197-210, Jun. 1978.

[2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short time spectral amplitude estimator," IEEE Trans Acoust., Speech, Signal Processing, vol. ASSP-32, no. 6, pp. 1109-1121, Dec. 1984.

[3] P. Scalart and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," Proc. ICASSP'96, pp.629-632, May1996.

[4] TIA/EIA/IS-127, "Enhanced Variable Rate Codec," Jan 1997.

[5] R. Martin, "Spectral subtraction based on minimum statistics,"EUSIPCO '94, pp.1182-1185, Sep.1994.

[6] M. Kato, A. Sugiyama and M. Serizawa, "Noise suppression with high speech quality based on weighted noise estimation and MMSE STSA,"Technical Report of IEICE, DSP/IE/MI2001-8,pp.53-60, Apr.2001.