Speech Enhancement using Spectral Subtraction

Suma. M. O[1], Madhusudhana Rao. D[2], Rashmi. H. N[3] & Manjunath B. S[4]

[1&3]Dept. ECE, RGIT, Bengaluru, [2]U.G Consultants, Bengaluru
E-mail : suma_mo@rediffmail.com[1], ug.madhu.rao@gmail.com[2],
hn.rashmi87@gamil.com[3], b.smanjunath@yahoo.in[4]

***Abstract* – In modern hands free speech communication environments often occurs the situation that the speech signal is superposed by background noise. 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.**

In this paper, the multi-band spectral subtraction method provides a definite improvement over the conventional power spectral subtraction method and does not suffer from musical noise. The improvement can be attributed to the fact that the multi-band approach takes into account the non-uniform effect of colored noise on the spectrum of speech. The added computational complexity of the algorithm is minimal. Four linearly spaced frequency bands were found to be adequate in obtaining good speech quality.

I. INTRODUCTION

Speech enhancement aims to improve speech quality by using various algorithms. It may sound simple, but what is meant by the word *quality*? It can be at least clarity and intelligibility, pleasantness, or compatibility with some other method in speech processing [1]. Intelligibility and pleasantness are difficult to measure by any mathematical algorithm. Usually listening tests are employed. However, since arranging listening tests may be expensive, it has been widely studied how to predict the results of listening tests. No single philosopher's stone or minimization criterion has been discovered so far and hardly ever will. The central methods for enhancing speech are the removal of background noise, echo suppression [4] and the process of artificially bringing certain frequencies into the speech signal.This focus on the removal of background noise after briefly discussing what the other methods are all about. First of all, every speech measurement performed in a natural environment contains some amount of echo. Echoless speech, measured in a special room, sounds dry and dull to human ear. Echo suppression is needed in big halls to enhance the quality of the speech signal, especially if the distance between the microphone and the speaker is large.

In the current telephone networks speech is band limited between 300–3400 Hz. Sooner or later the markets will be dominated by third generation phones in which the frequency band of the speech is, for instance, 50-7500 Hz. The delight of this wideband speech will be tamed unless the entire conversation is travelling in a wideband network. Artificial bandwidth expansion can be utilized to restore the frequencies that disappear on the route. These methods are also useful in speech compression. When the background noise is suppressed, it is crucial not to harm or garble the speech signal.

This paper is organized as follows. In Section II, we briefly describe about the speech enchancement. Then, the implementation of speech enchancement, spectral subtraction and the flow diagram in Section III. Section IV gives the simulation, synthesis result. Finally, brief summaries are given in Section V to conclude this paper.

II. SPEECH ENHANCEMENT

A. Introduction

Restoring the desired speech signal from the mixture of speech and background noise is amongst the oldest, still elusive goals in speech processing and communication system research. Speech enhancement algorithms attempt to improve the performance of communication systems when their input or output

signals are corrupted by noise. The main objective of speech enhancement or noise reduction is to improve one or more perceptual aspects of speech, such as the speech quality or intelligibility.

This is important in a variety of contexts, such as in environments with interfering background noise (e.g offices, streets and automobiles etc.) and in speech recognition systems, hands free environment for cars, hearing aids etc. Over the year, researchers and engineers have developed a number of methods to address this problem. Yet, due to complexities of the speech signal, this area of research still poses a considerable challenge. It is usually difficult to reduce noise without distorting speech and thus, the performance of speech enhancement systems is limited by the tradeoff between speech distortion and noise reduction. In general, the situation where the noise and speech are in the same channel (single channel systems) is the most common scenario and is one of the most difficult situations to deal with.

The complexity and ease of implementation of any proposed scheme is another important criterion especially since the majority of the speech enhancement and noise reduction algorithms find applications in real-time portable systems like cellular phones, hearing aids, hands free kits etc.

The spectral subtraction method has been one of the most well-known techniques for noise reduction. The spectral subtraction estimates the power spectrum of clean speech by explicitly subtracting the noise power spectrum from the noisy speech power spectrum. Due to its minimal complexity and relative ease in implementation, it has enjoyed a great deal of attention over the past years. This approach generally produces a residual noise commonly called musical *noise*. In this papert, we propose a modified spectral subtraction approach that allows better and more suppression of the noise. In addition, the proposed method also attempts to find the best tradeoff between speech distortion and noise reduction in a *perceptual sense* wherein the criteria for minimizing the residual noise are based on properties closely related to human perception.

B. Single channel enhancement system

The problem of enhancing speech degraded by background noise, when noisy speech alone is available, has been a research topic that has received great deal of attention over the past few decades. As mentioned before, there in having only a single microphone available, is one of the most difficult situations in speech enhancement, since no reference signal of the is noise is available, and the clean speech cannot be processed prior to being affected by the noise. The performance of single channel systems is usually limited because they tend to improve the quality of the noisy signal at the expense of some intelligibility loss. Therefore, there is a tradeoff between quality and intelligibility.

Existing single channel enhancement systems can be broadly divided into four categories:

- Suppression of noise using the periodicity of the speech or the noise.
- Model based speech enhancement.
- Short-time Spectral Amplitude estimation based speech enhancement.
- Enhancement based on Perceptual criteria.

C. Model Based speech Enhancement.

Speech enhancement involves estimation of the clean signal from a given sample function of noisy signal, that usually requires the explicit knowledge of the joint statistics of the clean signal and the noise process and it also requires a perceptually meaningful distortion measure for speech signals.

Therefore, if speech signals have been degraded by statistically independent noise, the marginal probability distributions of the clean speech and noise signal must be explicitly known. But in practice, neither the statistics of the signal and the noise nor the most meaningful distortion measures are explicitly known.

Hence the above theoretical approach can be applied as a two-step procedure: The statistics of the signal and noise are first estimated from training data of speech and noise, thereby, providing a sub optimal solution, which capitalize on statistical models, and then used along with currently available distortion measures to address the speech enhancement problem.

Different techniques like autoregressive moving average (ARMA), autoregressive (AR) or moving average (MA) are used to parameterize the speech model.

To estimate the parameters of a speech model, three estimation rules known as maximum likelihood (ML), maximum a posteriori (MAP) and minimum mean-square error (MMSE) are known to have many desirable properties [5] . ML is used often for nonrandom parameters. The MAP and MMSE estimation rules are commonly used for parameters that can be considered as random variables whose *a priori* density function is known.

One such approach for speech enhancement was proposed earlier, where a time-varying AR model is

assumed for speech signal, and both the model and the signal are estimated from the given noisy signal using MAP estimation approach. The maximization of the appropriate likelihood function is iteratively performed, once over the AR model assuming that the clean signal is available and then over the clean signal using the estimated model and an assumed known estimate of the power spectral density of the noise. Many variations and schemes based on this approach have been proposed

III. SPECTRAL SUBTRACTION

Introduction

Spectral subtraction is a method for restoration of the power or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum.

Spectral subtraction algorithm is used for removing only for the white noise and multi band spectral subtraction is used for removal of both white noise and as well as colored noise. A general representation of the technique is given in Fig 2. The input to the system is the noise-corrupted signal $y(n)$. In many methods available for the analysis-synthesis, the Short-term Fourier Transform (STFT) of the signal with Overlap and Add (OLA) is usually the most commonly used method. The spectral amplitude $|Y(w)|$ of the noisy input signal $y(n)$ is modified according to the noise estimation and the noise suppression rule on a frame-by-frame basis.

In Fig 1, the magnitude of speech spectrum is modified according to the estimated noise signal $d(n)$, measured during speech pauses/silences periodic. The weighing coefficients (gain function) with which each spectral component of the noisy input signal is multiplied or the amount of estimated noise subtracted (hence the name *subtractive-type*) from the noisy signal, is based on various subtraction rules. These subtraction rules have been experimentally optimized on the basis of the signal to noise ratio (SNR) of the input noisy signal. When the SNR is low, the speech segment under consideration is most likely to be noise alone, and maximum subtraction/suppression should be applied. Typically, the spectral amplitude of the noise signal is subtracted from that of the noise corrupted signal. Generally, if the noise is assumed to be uncorrelated with the speech signal, then the modified magnitude $|S\hat{}(w)|$ can be considered as an estimate of the clean speech $s(n)$. The time domain estimate $s\hat{}(n)$ of clean speech can be obtained by taking the *inverse discrete fourier transform* (IDFT) of $|S\hat{}(w)|$ along with the phase of the original noisy speech The first version, developed by Boll, is called *magnitude spectral subtraction*. Since then, there have been many variations derived from the original method in an effort to overcome some of the inherent drawbacks of the subtractive type algorithms and in particular Spectral Subtraction.

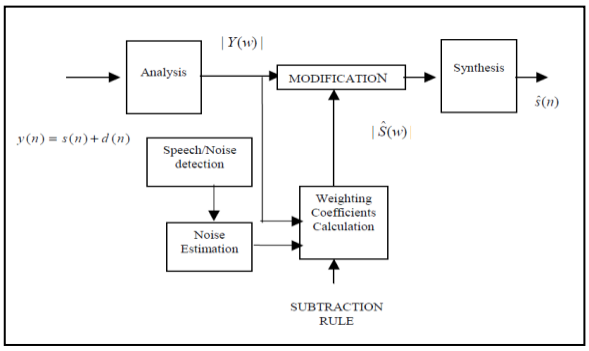


Fig 1: Diagrammatic representation of spectral subtraction

The weighing coefficients (gain function) with which each spectral component of the noisy input signal is multiplied or the amount of estimated noise subtracted (hence the name *subtractive-type*) from the noisy signal, is based on various subtraction rules.

These subtraction rules have been experimentally optimized on the basis of the signal to noise ratio (SNR) of the input noisy signal. When the SNR is low, the speech segment under consideration is most likely to be noise alone, and maximum subtraction/suppression should be applied. Typically, the spectral amplitude of the noise signal is subtracted from that of the noise corrupted signal.

Generally, if the noise is assumed to be uncorrelated with the speech signal, then the modified magnitude $|S\hat{}(w)|$ can be considered as an estimate of the clean speech $s(n)$. The time domain estimate $s\hat{}(n)$ of clean speech can be obtained by taking the *inverse discrete fourier transform* (IDFT) of $|S\hat{}(w)|$ along with the phase of the original noisy speech The first version, developed by Boll, is called *magnitude spectral subtraction*.

Since then, there have been many variations derived from the original method in an effort to overcome some of the inherent drawbacks of the subtractive type algorithms and in particular Spectral Subtraction.

Thus, assume that $y(n)$, the discrete noise corrupted input signal, iscomposed of the clean speech signal $s(n)$ and $d(n)$ the uncorrelated additive noise signal, then it the noisy signal can be represented as:

$$y(n)=s(n)+d(n) \qquad (1)$$

This assumption is based on the fact that $s(n)$, is stationary, but speech is not a stationary signal. The processing, is carried out on a short-time basis (frame-

by frame), therefore, a time-limited window $w(n)$ multiplies the original speech, noisy speech signal as well as the noise.

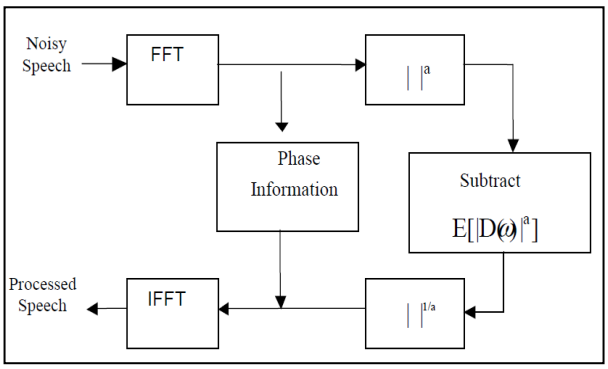


Fig. 2: Genereal representation of spectral subtraction

B. *Multi band spectral subtraction*

Assuming the additive noise to be stationary and uncorrelated with the clean speech signal, the resulting corrupted speech can be expressed as:

$$y(n)=s(n)+d(n) \quad (3)$$

Where y(n), s(n) and d(n) are the corrupted speech signal, clean speech signal and the noise respectively. The power spectrum of the corrupted speech can be approximately estimated as:

$$|Y(k)|^2 \approx |S(k)|^2 + |D(k)|^2 \quad (4)$$

where S(k) andD(k) are the magnitude spectra of the clean speech and the noise respectively.

the clean speech spectrum is obtained as:

$$|\hat{}\, S(k)|2 = |Y(k)|2 - \alpha|\hat{}D(k)|2 \quad (5)$$

Using the SNRi value calculated in Eq. (5), αi can be determined as:

$$\alpha_i = \begin{cases} 5 & SNR_i < -5 \\ 4 - \frac{3}{20}(SNR_i) & -5 \le SNR_i \le 20 \\ 1 & SNR_i > 20 \end{cases} \quad (6)$$

While the use of the over-subtraction factor αi provides a degree of control over the noise subtraction level in each band, the use of multiple frequency bands and the use of the δi weights provide an additional degree of control within each band.

The negative values in the enhanced spectrum were floored to the noisy spectrum as:

$$|\hat{S}_i(k)|^2 = \begin{cases} |\hat{S}_i(k)|^2 & |\hat{S}_i(k)|^2 > 0 \\ \beta |Y_i(k)|^2 & else \end{cases} \quad (7)$$

Where the spectral floor parameter was set to β = 0.002.

This section describes the proposed method for speech enhancement with reduced residual noise. A block diagram of the proposed method is shown in Figure 3. It consists of 4 stages. In the first stage, the signal is windowed and the magnitude spectrum is estimated using the FFT. In the second stage, we split the noise and speech spectra into different frequency bands and calculate the over-subtraction factor for each band. The third stage includes processing the individual frequency bands by subtracting the corresponding noise spectrum from the noisy speech spectrum. Lastly, the modified frequency bands are recombined and the time signal is obtained by using the noisy phase information and taking the IFFT in the fourth stage. The effect of pre-processing operations is to neutralize the distortion in the spectral content of the input data due to the analysis window and to precondition the input data to surmount the distortion due to errors in the subtraction process.

C. *Implementation and Flow diagram*

Spectral subtraction is very popular method for noise suppression in speech signals due its robustness and the simplicity for implementation. The method described in this paper is based on the combination of spectral subtraction with iterative Weiner filtering.

The main advantage of this method is estimation of a background noise continually without VAD. Modified spectral subtraction was implemented on fixed-point DSP.

Evaluation of a speech enhancement algorithm is not simple. While objective quality assessment methods can indicate an improvement or degradation in speech quality based on mathematical measures.

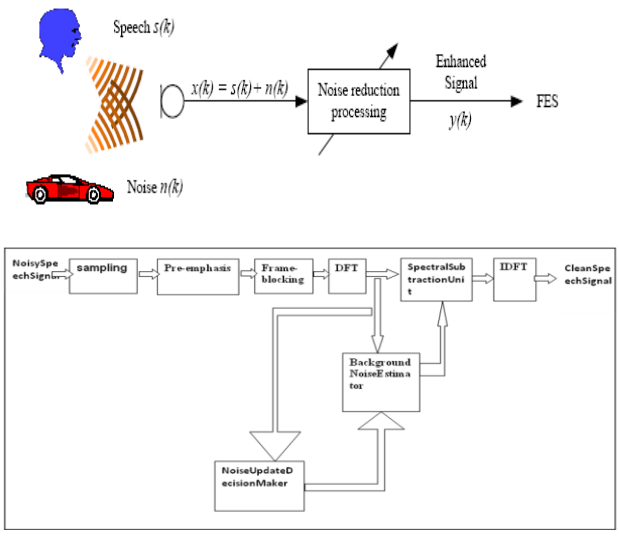


Fig 4: Functional block diagram

Sampling: As the speech rate varies from 300-3.4kHz the sampling rate is taken above the Nyquist rate to avoid aliasing effect and other distortions. The sampling rate in our process is 8kHz.

Pre-emphasis: The high frequency components in any section of the spectrum will have less energy compared to the low frequency components. Certain applications like speech recognition and others we need the features of high frequency components.

Hence to extract the features at high frequency region we need to energise the high frequency components, by pre-emphasising. The process of pre-emphasising also reduces the noise in the high frequency region especially when the noise is stationary. The process of pre-emphasising is given by

$$y(n)=x(n)+\alpha x(n1) \qquad (8)$$

Where x(n) is the present input sample to the pre-emphasiser y(n) is the pre-emphasised sample ,α is the pre-emphasising factor.

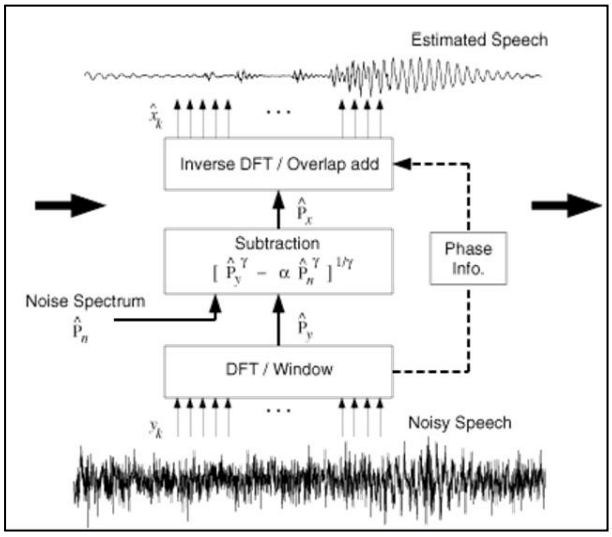


Fig. 5: Flow Diagram Of Spetral Subtraction

The popularity of spectral subtraction is largely due to its relative simplicity and ease of implementation. As shown in Fig 4 the short-term power spectrum $\hat{\mathbf{P}}_y$ (magnitude squared of the short-term Fourier transform) of the noisy signal is computed, and an *estimate* of the short-term noise spectrum $\hat{\mathbf{P}}_n$ is subtracted out to produce the estimated spectrum $\hat{\mathbf{P}}_x$ of the clean speech. Explicitly,

$$\left[\hat{\mathbf{P}}_y - \alpha.\hat{\mathbf{P}}_n^{\gamma}\right]^{1/\gamma} \qquad (9)$$

Where the scaling factor α allows for emphasis or de-emphasis of the noise estimate, and γ allows for several variants, including power subtraction (γ = 1) and magnitude subtraction (γ = 0.5).The estimate $\hat{\mathbf{P}}_x$ is combined with the phase from the original noisy signal to produce an estimate of the Fourier transform of x. Finally, the inverse Fourier transform is applied with the overlap-and-add method to construct a time-domain estimate of the speech waveform $\hat{x}$.

IV. RESULTS

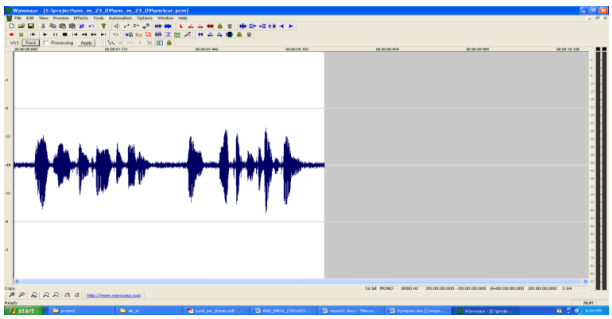
For the noisy speech signal, sampling is performed at the rate of 10 ms and in each segment it contains 80 samples. The output of noisy sampled signal is pre-emphasized so that only the low frequency signal will pass through it. The output of the pre-emphasis is divided into blocks of frames containing 128 samples in each frame. The output of frame blocking is given to overlap and add method and applying windowing (trapezoidal) technique and the windowed output is given to DFT/FFT.

After applying DFT/FFT, it calculates for the magnitude and phase for each frame. Then it checks all the condition and calculates the noise estimation by using the formula. After performing all the operations next step is to perform IDFT/IFFT to get back the original signal

The working of the spectral subtraction algorithm was first tested in Microsoft VC++. Once the simulation was achieved the code was reconstructed to make it compatible with the TMS320C5416 DSK by converting it into a fixed point code and by introducing the necessary APIs to make the real time implementation possible.

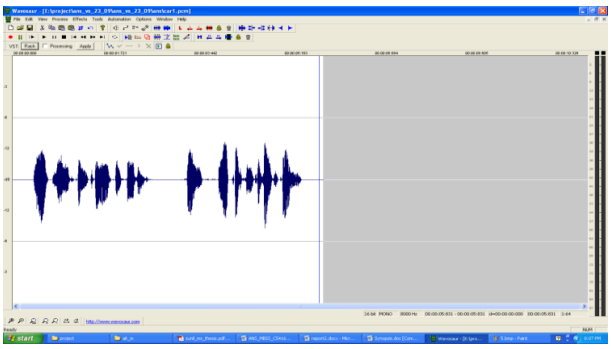
A. Spectral Subtraction In Vc++

Noisy speech signal:-This is the waveform of the generated the noisy speech signal of the original signal. This signal is generated using a single delay buffer logic wherein the original signal is delayed in time axis by one unit and is attenuated .



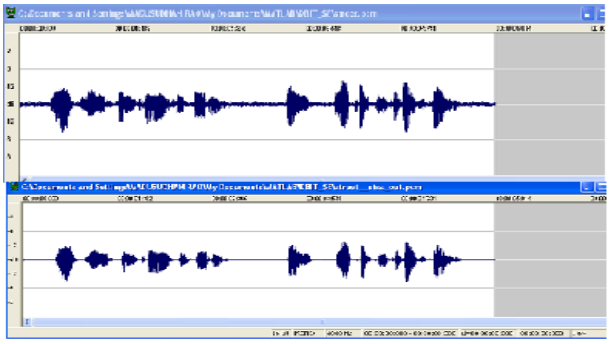
Residual noisy signal: This is the waveform of the error signal the error signal is removed by using a spectral subtraction algorithm. The error signal is the difference between the noisy signal and the signal generated by poly phase filter.

Clear speech signal:-This is the waveform of the original clean speech signal. Removing the noisy signal by using multi band spectral subtraction algorithm. Here by maintaining the sanctity of speech information as it is, i.e. without changing magnitude and phase information of signal. And it has been degraded.



B. *Matlab results*

First code is written in matlab 7.0.1 for referenceand for writing the code in VC++ it is easy to write the code in C language. The results of both Matlab and VC++ will be same.



V. CONCLUSION

The multi-band spectral subtraction method provides a definite improvement over the conventional power spectral subtraction method and does not suffer from musical noise. The improvement can be attributed to the fact that the multi-band approach takes into account the non-uniform effect of colored noise on the spectrum of speech. The added computational complexity of the algorithm is minimal. Four linearly spaced frequency bands were found to be adequate in obtaining good speech quality.

The algorithm can be implemented in real-time on a fixed point Digital Signal Processor (DSP) (e.g., the Texas Instruments TMS320C54x/55x) platform for evaluation in real-world conditions. This would require a detailed quantization analysis of the algorithm. Fixed-point DSPs are becoming increasingly popular in applications such as cellular-phones, personal entertainment devices, digital hearing aids and headsets due to their low-power consumption and high processing rates. Speech enhancement algorithms are a major component of these applications for operation in adverse environments. The proposed method can eventually be incorporated into such systems. However, these applications also demand low MIPS (Million Instructions Per Second), i.e., low number of operations, to conserve battery life.

VI. REFERENCES

[1] Speech Enhancement using a non causal A Priori SNR Estimator, IEEE Signal Processing Letters. VOL.11.No.9.September 2004.

[2] Rabiner Lawrence, Juang Bing-Hwang. Discrete Time Speech Signal Processing, Prentice Hall , New Jersey, 2004, ISBN 0-13-015157-2.

[3] Sunil D. Kamath and Philipos C. Loizou A Multi-Band Spectral Subtraction Method For Enhancing Speech Corrupted By Colored Noise IEEE Signal Processing Letters. VOL.23.No.7.September 2009.

[4] S. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. Acoust., Speech, Signal Process., vol.27, pp. 113-120, Apr. 2010.

[5] S. Kamath and P. Loizou, “A multi-band spectral subtraction method for enhancing speech corrupted by colored noise,” submitted to ICASSP 2002.

[6] B. Picinbono, Random Signals and Systems. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[7] TMS320C54x DSP: CPU and Peripherals: Reference Set, Volume #1, SPRU131; TI web page: http://www.ti.com.

◈◈◈