White Paper
SPRAAJ4–July 2008

Automatic Level Controller for Speech Signals Using PID Controllers

Fitzgerald J. Archibald

An Automatic Level Controller (ALC) for speech signals embedded in additive noise requires Voice Activity Detection (VAD) to avoid noise amplification. The VAD is generally computationally intensive. This paper explores a less computational-intensive ALC, which uses a Proportional Integral (PI) controller for detecting voice activity. The PI controller tracks the energy variation of the signal by means of detecting rapid variations of the signal envelope. The segments that show a low variation of energy are detected as noise. Further, this paper describes gain calculation for amplification of a speech segment using an Integral controller. The speech signal energy is estimated by a second PI controller which follows signal envelope and thus provides the amplitude error that needs to be corrected.

Index terms: AGC, ALC, automatic volume controller, digital gain controller, VAD.

Contents

List of Figures

1 Introduction

In Digital Still Camera (DSC), sound is recorded along with captured video frames for the movie capture application. The sound signal is converted to an electrical signal by a microphone and converted to a digital signal by an Analog to Digital Converter (ADC). Often, the intent of movie capture is to capture the speech associated with the video (either verbal comments of the camera operator or the speech of the human subject under movie capture).

The sound level could vary over time of recording depending on human speaker articulation, multiple speakers in the case of conversation, microphone directional properties, and analog gain levels on the microphone input. This brings in the need for and automatic level controller (ALC) so as to maintain a constant sound level over time to ensure the sound is intelligible, while not amplifying the noise only segments.

Steber (1988) has studied digital signal processing for use in an analog gain controller. The focus is mainly on peak detection for normalization of amplitude. The paper explores AGC for speech and audio without utilizing the knowledge of audio speech signal characteristics. Kang and Lidd (1984) used energy estimation of the voiced and unvoiced segments for voice activity detection. This method uses 0.5 msec for AGC computation which is relatively high on an embedded system like DSC where video needs to be supported along with audio.

This paper describes an ALC using a low computation-intensive Voice Activity Detector (VAD) for detecting speech embedded in stationary noise. Stationary noise is considered since most of the non-stationary noise is stationary in a short time window. The VAD and amplitude error estimation is carried out using the PI controller. The fundamental theory of PID controller design can be understood from publications of Ang et al. and Astrom. In addition, a gain controller used for leveling amplitude error is presented.

This paper is organized into three major sections: design, results and summary. The design is divided into three major parts: amplitude error estimation, VAD and gain controller.

2 Algorithm Design

In the case of stationary noise, the noise power would remain constant with respect to time. Thus, a speech signal *x*[k] embedded in stationary additive noise *n*[k], *xn*[k] in (1), would have varying signal power in the presence of speech, and constant noise power in the absence of speech.

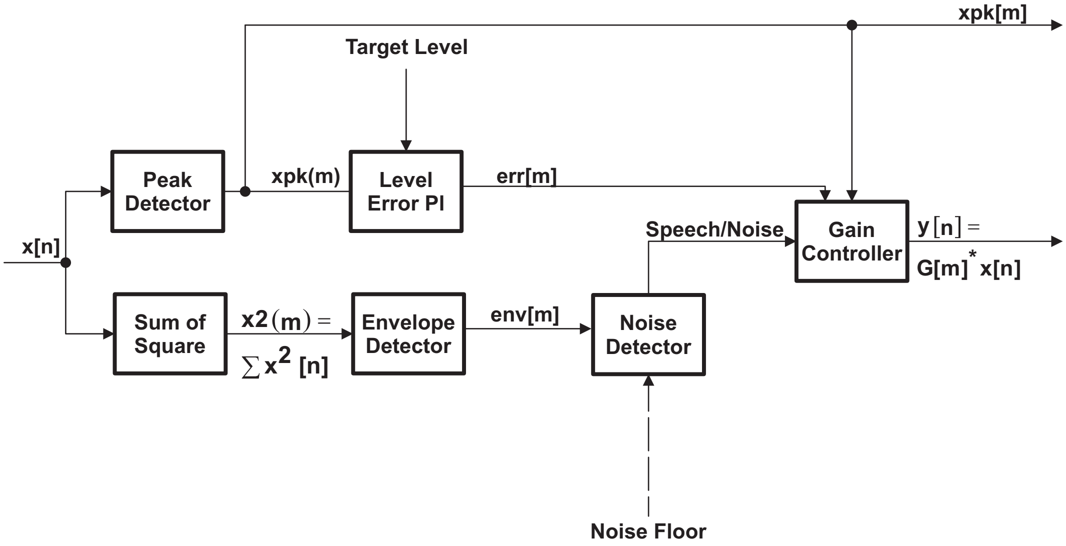
$$x_n[k]=x[k]+n[k] \tag{1}$$

Speech signals consist of voiced/unvoiced and silence segments. The silence segments may contain noise. The speech segment may have additive noise. The noise can be stationary as well as non-stationary.

The ALC algorithm should avoid sudden gain changes to prevent abrupt variation of noise floor, which can cause an audible energy level change in additive stationary noise. At the same time, delay in gain changes or slow gain reduction can cause signal saturation, resulting in audible artifacts. A large step size in gain change will cause zipper noise, which is undesirable.

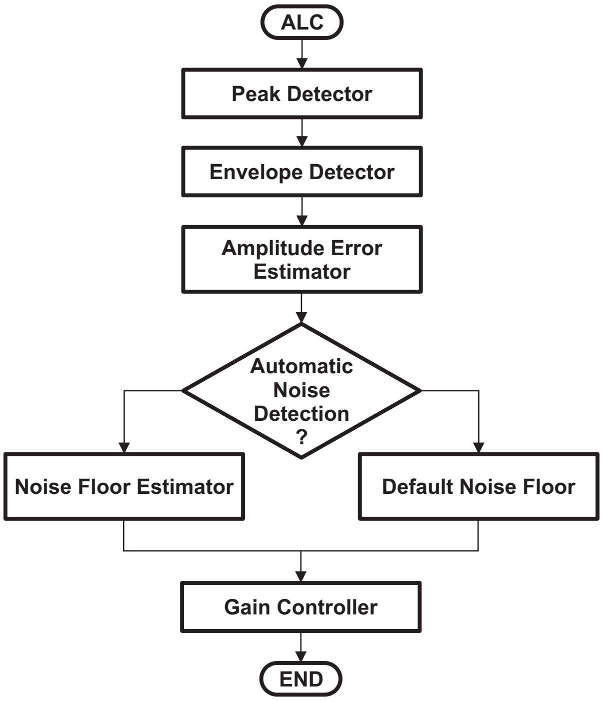
The ALC for a speech signal consists of an error detector, a noise floor estimator and a gain controller. An amplitude error detector consists of a peak detector, and an error detector. A noise floor estimator consists of a signal energy detector, and a noise/speech detector and a noise level estimator. A gain controller consists of a gain estimator and an amplitude level controller (volume controller). Figure 1 illustrates the block diagram of ALC.

Figure 1. ALC Block Diagram



In the case of a static noise floor, noise floor estimation can be bypassed with a user-defined noise floor. A signal below the noise floor level is considered as noise in this case. This is illustrated in Figure 2.

Figure 2. ALC Flowchart



The input for ALC is the input data stream, target level for output data stream, and optionally noise floor level. The input data stream is divided into sub-frames and frames. Sub-frames consist of multiple speech samples. Sub-frames should be preferably large enough to contain at least one phoneme. The frame consists of multiple sub-frames. The frame should preferably be longer than a spoken word (on average).

The output from ALC is the output data stream (gain controlled stream to reach desired target level), and peak signal in current sub-frame. The gain is controlled per sub-frame. The peak signal can be used for monitoring and adjusting clipping resulting from analog Programmable Gain Array (PGA) or pre-amp gain.

2.1 Signal Level Detection

The peak signal level is required for determining the target gain required to level the signal. The peak level is absolute of positive or negative peak in the input signal.

The input signal is divided into frames. Each frame is sub-divided into sub-frames. The peak signal level in the current sub-frame is detected. The peak signal level of the current frame is detected using available sub-frame peak samples in the current frame. The peak level for determining gain is taken as the largest of the peak sample in the current frame and the previous frame.

Pseudo code of the peak signal level detection logic is given below:

Initialize *xpk*[k] to 0 as given by (2).

$$xpk[k] = 0 \tag{2}$$

For each sample within the sub-frame:

```
If $xpk\_sf[k] > x[n]$
   $xpk\_sf[k] = xpk\_sf[k]$
else if $xpk\_sf[k] \le x[n]$
        $xpk\_sf[k] = x[n]$                                          (3)
   End If
End For
```

In (3) *k* is the index of the sub-frame, range [0...K]; *n* is the index of PCM samples within sub-frame, range [0...K]; and *xpk_sf*[*k*] is the peak PCM sample in the current frame.

For each sub-frame within frame:

$$xpk\_f = MAX\left(xpk\_sf[k]\right) \tag{4}$$

End For

In (4) *k* = 0, 1, 2, …, K and *xpk_f* is the max amplitude PCM sample in current frame.

For successive frames:

$$xpk = MAX\left(xpk\_f, xpk\_f\_l\right) \tag{5}$$

End For

In (5) *xpk_f_1* is the previous frame amplitude peak and *xpk* is the amplitude peak for the past two frames.

Update the previous frame peak with the current frame peak for the next iteration as given by (6).

$$xpk\_f\_1 = xpk\_f \tag{6}$$

The peak signal *xpk*[k] is discontinuous. The error signal is computed by subtracting the peak signal from the maximum allowed signal level. This error is passed through the integral controller to produce the smoothed error signal *err*[k].

2.2 Noise Level Estimation

VAD consists of a signal envelope tracker followed by an envelope variation computation. It can be seen that for a stationary noise signal, the envelope remains flat (due to constant signal energy). In the case of a speech segment, the envelope will show a larger variation due to changes in signal energy. Thus, by determining the envelope level variation within a short span in time, it is possible to identify speech and noise segments accurately.

2.2.1 Envelope Detection

Envelope detection involves computing the mean level of the signal in the current sub-frame and using a Proportional Integral (PI) controller to highlight the mean signal envelope changes.

Energy as given by (Equation 7) is proportional to the mean signal as given by (8). The amount of variation in energy levels helps to identify speech and noise segments.

$$E = \int |\, x(t)\,|^{2}\, dt \tag{7}$$

Pseudo code for a mean signal computation is shown in (8) . For each sample within a sub-frame:

$$xmean[k] = \sqrt{\frac{1}{N}\sum_{n=0}^{n=N} x^{2}[n]} \tag{8}$$

End For

In (8) *xmean*[k] is the mean amplitude in the current sub-frame, and the range of *k* is [0...K]

The mean signal, *xmean*[k], change between sub-frames is limited before passing through the PI controller. Dither is added, when necessary, to prevent the PI controller from settling down to a steady state. the PI controller is used to highlight the envelope for speech and noise segment detection, for preventing unwanted signal transition between speech and noise.

```
//!< if noise floor to be raised
     if( xmean[k] > estNoiseLt[k-1] )
    {
        //!< AmplifyFactor1dB = 1.0 / 0.8913
        estNoiseLt[k] = estNoiseLt[k-1] * AmplifyFactor1dB;
        //!< Saturate the estimated noise floor limit with mean signal level
        if(estNoiseLt[k] >  xmean[k]) {
            estNoiseLt[k] = xmean[k];
        //!< if estNoiseLt[k] is too small or zero and xmean[k] is non-zero
        } else if( (estNoiseLt[k] == 0) && (xmean[k] != 0) ) {
            //!< use optimal noise floor    (0.0008)
            estNoiseLt[k] = optNoiseFloorConst;
        }

        // PI aided increase of noise floor (P=+0.1, I=+0.1)
        env[k] = env[k-1]*I  + estNoiseLt[k] * P;
    }
    //!< if noise floor to be lowered

    else if(xmean[k] < estNoiseLt[k-1])
    {
        // AttenuateFactor1dB = 0.8913
        estNoiseLt[k] = estNoiseLt[k-1] * AttenuateFactor1dB;
        //!< Don’t allow the noise floor limit to go below mean signal level
        if(estNoiseLt[k] < xmean[k]) {
            estNoiseLt[k] = xmean[k];
        //!< if estNoiseLt[k] is too small or zero and xmean[k] is non-zero
        } else if( (estNoiseLt[k] == 0) &&  (xmean[k] != 0) ) {
            //!< use optimal noise floor (0.0008)
            estNoiseLt[k] = optNoiseFloorConst;
        }

        // PI aided decrease of noise floor (P=-0.1, I=+0.2)
        env[k] = env[k-1]*I + estNoiseLt[k]*P;
        if( env[k] < 0) {
            //!< prevent negative
            env[k] = 0;
        }
    }
    //!< if noise floor level to be maintained
    else
    {
        estNoiseLt[k] = estNoiseLt[k-1];
    }

    //!< update state
    env[k] = env[k-1];
    //!< update state estNoiseLt[k-1]
    if( estNoiseLt[k-1] != estNoiseLt[k] ) {
        estNoiseLt[k-1]    = estNoiseLt[k];
    } else {
        //!< Try to bias towards lowering of noise floor by 1 dB
        //!< Introduce error to keep the PI loop running
        estNoiseLt[k-1]    = 1.121*estNoiseLt[k];
    }
```

A proportional integrator is given by (9).

$$env[k] = I \times env[k-1] + P * estNoiseLt[k] \tag{9}$$

In (9) P and I are proportional and integral coefficients, *env*[k] is the pseudo-envelope of signal energy (energy variations are amplified) computed for the present sub-frame k, and *estNoiseLt* [k] is the envelope (filtered noise floor value) computed for the current sub-frame (*k*).

Two sets of P & I can be used for controlling the increasing and decreasing of the envelope, respectively.

2.2.2 Noise Detection

Signal *env*[k] variation is observed over multiple sub-frames for voice activity detection. If the signal variation is within the limit (flat), the sub-frame is classified as stationary noise. Also, if the signal level is below the threshold (noise floor setting), the signal is classified as noise. To classify a signal as noise, the signal level variation is observed for more than one sub-frame. This is required to account for constant signal energy within short segments of speech. Otherwise, noise segments may be spuriously detected within a short speech segment.

If the signal variation is greater than limit (changing energy across sub-frames), the signal segment is classified as speech. The first occurrence of a large variation of env[k] is used to detect speech in order to start leveling the speech signal sooner.

If adaptive noise detection is disabled, then all sub-frames with a peak signal level below the specified noise floor are detected as noise segments.

2.3 Gain Control

If the sub-frame is classified as noise, amplification is reduced gradually to reach 0 dB (pass-through). If continuous frames are observed to be noise, the signal is attenuated gradually to reach the desired level of attenuation. On speech signal detection, any attenuation present is removed gradually to reach the desired attenuation. In the case of a rapid change of signal energy in the speech segment of the signal, the gain is reduced gradually to avoid signal clipping (saturation). The rate of change of gain can be controlled to be faster or slower by means of gain scaling factors, depending on the cause and nature of the gain change. If the rate of gain change is too fast, audible zipper noise and noise energy fluctuation can result in deteriorated sound quality. If the gain change is too slow, noise amplification and clipping can cause audible noise. Thus, the optimal rate of gain change should be selected based on the cause for gain change.

For a speech signal, the goal of gain change is to make *err*[k] → 0. The difference between the maximum signal level and *err*[k] would provide the peak signal level for gain computation. The automatic gain for leveling the input signal, in simplest form, is given by (10).

$$G = \frac{DesiredAmplitude}{PeakAmplitude} \tag{10}$$

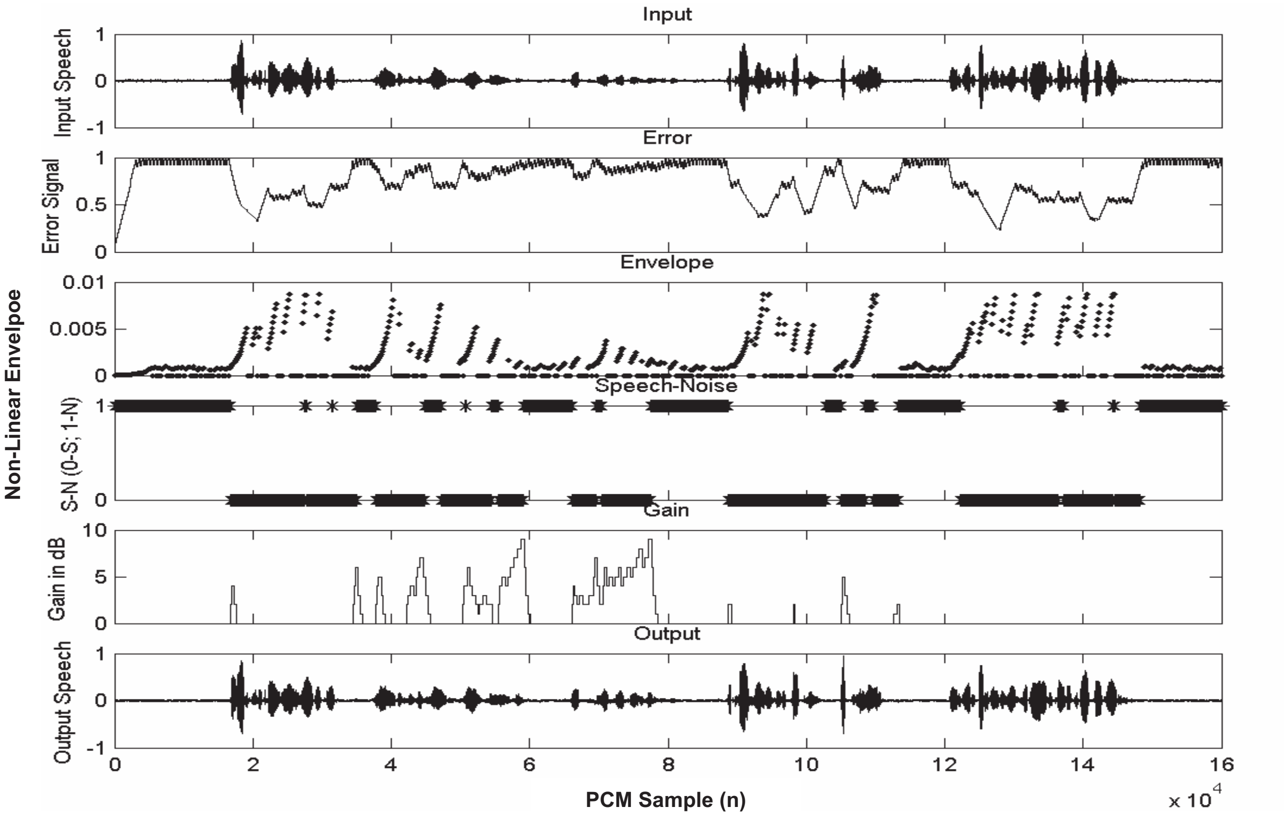
The gain can be applied either on digital data or analog signal.

2.3.1 Analog PGA

Analog PGA on ADC can be used for gain control, provided headroom is available on the digital domain for the converted digital PCM samples. Headroom is needed to detect clipping caused by analog gain and the naturally peaking input signal. In the case of clipping caused by ALC, the gain needs to be lowered.

If analog gain were to be used, delay in cause and effect needs to be addressed. The computed gain does not correspond to the sub-frame on which the gain is applied. The problem gets alleviated depending on the amount of buffering between ADC and ALC.

Figure 3. ALC Internal and External Signal Plot



2.3.2 Digital Gain

Gain can be applied on digitized PCM samples with low computational load by fixed point multiplication. Digital gain eliminates the cause-effect delay, as the gain can be applied on the sub-frame for which the gain is computed. However, digital gain does not increase resolution or precision when the gain is increased.

3 Results

The CPU cycle-intensive parts of the ALC algorithm are peak detection and energy computation, since computation is done using all the samples in the current sub-frame. the rest of the computation is based on sub-frame and frame interval. The ALC on the ARM9EJ processor consumes less than 1 MHz.

Figure 3 shows internal waveforms of the ALC algorithm for a speech signal embedded in colored noise, with varying SNR. The error signal along with the S-N signal when provided to the gain controller results in the gain curve shown in the graphs. The non-linear envelope is used to estimate the S-N signal. The speech segment starting at around 65,000 samples and ending at 80,000 samples illustrate the effect of ALC very clearly. The gain curve shows the amplification that is applied for signal starting from 35K samples and ending at 80K samples.

Band Pass Filter (BPF) is used to band-limit the input signal to the speech spectrum. The BPF filter takes about 1 MHz on ARM9EJ. BPF is not required if the input signal is already band-limited.

4 Summary

By use of an integral controller for computing amplitude error and a gain controller with programmable gain step size, the ALC achieved an enhanced listening experience by eliminating zipper noise and noise amplification. The detection of energy level variations across sub-frames and energy level in current sub-frame is effective in detecting speech and noise segments in a variety of noisy environments. The PI controller used in VAD is effective in improving the voice activity detection accuracy. Due to the fast tracking of signal and noise segments, amplification up to 18 dB is achieved on noisy speech signals with 20 dB SNR. The overall computational load of ALC including VAD is negligible in embedded systems using a general purpose or a digital signal processor intended for handheld devices.

5 References

1. *PID control system analysis, design, and technology* by Ang K. H et al., 2005. Control Systems Technology, IEEE Transactions, Volume: 13, Issue: 4, pp. 559- 576.
2. *The Control Handbook* by Astrom, K. J. and Hagglund, T., 1996. PID control. In Levine, Ed., CRC Press and IEEE Press, pp. 198-209.
3. *Automatic gain control. Acoustics, Speech, and Signal Processing* by Kang, G.S. and Lidd, M.L., 1984. IEEE International Conference on ICASSP. Volume 9, pp. 120 – 123.
4. *Digital Signal Processing In Automatic Gain Control Systems* by Steber, G. R., 1988. Industrial Electronics Society, 14 Annual Conference. Volume 2, pp. 381 – 384, 24-28.

Fitzgerald Archibald earned a B.E (Electronics and Communication Engineering) from PSG College of Technology, Coimbatore, Tamil Nadu, India in 1996. He worked on control systems software development for geo-synchronous satellites from 1996 to 1999 in ISRO Satellite Centre, Bangalore, India. In 2001-2002, he worked on speech decoder, real-time kernel, and audio algorithms for DVD audio team in Sony Electronics, San Jose, USA. While in Philips Semiconductors (Bangalore, India, and Sunnyvale, USA) in 1999-2001 and 2002-2004, he worked on audio algorithms and systems for STB, DTV, and internet audio. He is part of the Personal Audio Video and Digital Imaging groups in Texas Instruments Inc, Bangalore, India from 2004-till date working on audio, video, and imaging systems and algorithm development. Interests include multimedia and control algorithms and systems.

Mr. Archibald is member of AES.

IMPORTANT NOTICE

Texas Instruments Incorporated and its subsidiaries (TI) reserve the right to make corrections, modifications, enhancements, improvements, and other changes to its products and services at any time and to discontinue any product or service without notice. Customers should obtain the latest relevant information before placing orders and should verify that such information is current and complete. All products are sold subject to TI's terms and conditions of sale supplied at the time of order acknowledgment.

TI warrants performance of its hardware products to the specifications applicable at the time of sale in accordance with TI's standard warranty. Testing and other quality control techniques are used to the extent TI deems necessary to support this warranty. Except where mandated by government requirements, testing of all parameters of each product is not necessarily performed.

TI assumes no liability for applications assistance or customer product design. Customers are responsible for their products and applications using TI components. To minimize the risks associated with customer products and applications, customers should provide adequate design and operating safeguards.

TI does not warrant or represent that any license, either express or implied, is granted under any TI patent right, copyright, mask work right, or other TI intellectual property right relating to any combination, machine, or process in which TI products or services are used. Information published by TI regarding third-party products or services does not constitute a license from TI to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property of the third party, or a license from TI under the patents or other intellectual property of TI.

Reproduction of TI information in TI data books or data sheets is permissible only if reproduction is without alteration and is accompanied by all associated warranties, conditions, limitations, and notices. Reproduction of this information with alteration is an unfair and deceptive business practice. TI is not responsible or liable for such altered documentation. Information of third parties may be subject to additional restrictions.

Resale of TI products or services with statements different from or beyond the parameters stated by TI for that product or service voids all express and any implied warranties for the associated TI product or service and is an unfair and deceptive business practice. TI is not responsible or liable for any such statements.

TI products are not authorized for use in safety-critical applications (such as life support) where a failure of the TI product would reasonably be expected to cause severe personal injury or death, unless officers of the parties have executed an agreement specifically governing such use. Buyers represent that they have all necessary expertise in the safety and regulatory ramifications of their applications, and acknowledge and agree that they are solely responsible for all legal, regulatory and safety-related requirements concerning their products and any use of TI products in such safety-critical applications, notwithstanding any applications-related information or support that may be provided by TI. Further, Buyers must fully indemnify TI and its representatives against any damages arising out of the use of TI products in such safety-critical applications.

TI products are neither designed nor intended for use in military/aerospace applications or environments unless the TI products are specifically designated by TI as military-grade or "enhanced plastic." Only products designated by TI as military-grade meet military specifications. Buyers acknowledge and agree that any such use of TI products which TI has not designated as military-grade is solely at the Buyer's risk, and that they are solely responsible for compliance with all legal and regulatory requirements in connection with such use.

TI products are neither designed nor intended for use in automotive applications or environments unless the specific TI products are designated by TI as compliant with ISO/TS 16949 requirements. Buyers acknowledge and agree that, if they use any non-designated products in automotive applications, TI will not be responsible for any failure to meet such requirements.

Following are URLs where you can obtain information on other Texas Instruments products and application solutions:

Products		**Applications**	
Amplifiers	amplifier.ti.com	Audio	www.ti.com/audio
Data Converters	dataconverter.ti.com	Automotive	www.ti.com/automotive
DSP	dsp.ti.com	Broadband	www.ti.com/broadband
Clocks and Timers	www.ti.com/clocks	Digital Control	www.ti.com/digitalcontrol
Interface	interface.ti.com	Medical	www.ti.com/medical
Logic	logic.ti.com	Military	www.ti.com/military
Power Mgmt	power.ti.com	Optical Networking	www.ti.com/opticalnetwork
Microcontrollers	microcontroller.ti.com	Security	www.ti.com/security
RFID	www.ti-rfid.com	Telephony	www.ti.com/telephony
RF/IF and ZigBee® Solutions	www.ti.com/lprf	Video & Imaging	www.ti.com/video
		Wireless	www.ti.com/wireless

Mailing Address: Texas Instruments, Post Office Box 655303, Dallas, Texas 75265