Subband-Based Acoustic Shock Limiting Algorithm On A Low-Resource DSP System

G. Choy, D. Hermann, R. L. Brennan, T. Schneider, H. Sheikhzadeh, E. Cornu

Dspfactory Ltd., 611 Kumpf Drive, Unit 200, Waterloo, Ontario, Canada N2V 1K8
gary.choy@dspfactory.com

Abstract

Acoustic Shock describes a condition where sudden loud acoustic signals in communication equipment causes hearing damage and discomfort to the users. To combat this problem, a subband-based acoustic shock limiting (ASL) algorithm is proposed and implemented on an ultra low-power DSP system with an input-output latency of 6.5 msec. This algorithm processes the input signal in both the time and frequency domains. This approach allows the algorithm to detect sudden increases in sound level (time-domain), as well as frequency-selectively suppressing shock disturbances in frequency domain. The unaffected portion of the sound spectrum is thus preserved as much as possible. A simple ASL algorithm calibration procedure is proposed to satisfy different sound pressure level (SPL) limit requirements for various communication equipment. Acoustic test results show that the ASL algorithm limits acoustic shock signals to below specified SPL limits while preserving speech quality.

1. Introduction

Users of communication equipment (e.g. telephones and headsets) are often exposed to loud sounds or sounds containing rapid increases in level. Examples of these unwanted sounds include feedback oscillation and loud signaling tones (e.g. fax tone). Without protection against these sounds, users may experience a phenomenon known as acoustic shock. Exposure to acoustic shock leads to undesirable outcomes including headaches, nausea and hearing loss. This problem is particularly acute for people using headsets, such as call-center operators, since sound pressure level (SPL) at the ear-drum is typically higher when headsets are used.

Several existing methods can be used to prevent users from experiencing acoustic shock. They include: 1) high-level limiting using automatic gain control [1] and 2) clipping of high-level signals using diodes or similar devices [2]. While these methods provide some protection against acoustic shock, their output signals have reduced fidelity compared to the input signal. These methods fail to identify and attenuate the portions of the spectrum affected by the shock disturbance, thus causing audio dropouts, inter-modulation and harmonic distortion. A subband-based compression scheme proposed in [3] performs frequency-dependent dynamic range control, but is designed mainly for speech and does not account for shock disturbances. More complex systems may also suffer from excessive input-output latency (i.e. group delay), which adversely impacts network, acoustic and line echo-cancellers.

This paper proposes a subband-based acoustic shock limiting (ASL) algorithm that suppresses shock disturbances while preserving speech quality. Our research has several novel contributions: **(1)** The ASL algorithm processes the input signal in both time and frequency domains. This allows the algorithm to detect sudden increases in sound level, as well as frequency-selectively suppressing shock disturbances while preserving the rest of the sound spectrum. **(2)** The ASL algorithm is implemented on an ultra-low power DSP system using an oversampled weighted overlap-add [4] (WOLA) filterbank. This oversampled filterbank efficiently separates the input signal into subband signals while achieving low group delay. **(3)** A simple calibration procedure is proposed to configure the ASL algorithm such that different SPL limit requirements can be satisfied for different communication equipments.

In this paper, we will concentrate on the suppression of narrowband shock disturbances. In the following sections, we first present a description of the DSP architecture, followed by a description of the ASL algorithm. In Section 4, we describe the calibration procedure and test setup for the ASL algorithm. Finally, results and conclusions are presented. Throughout this paper, we will use the term "shock" to describe a rapid increase of signal energy or high signal energy level that will cause acoustic shock phenomenon at the DSP output.

2. The DSP System

The DSP system [4, 5] consists of three major components: a weighted overlap-add (WOLA) filterbank coprocessor, a 16-bit fixed point DSP core, and an input-output processor (IOP). These three components run in parallel and communicate through shared memory. The parallel operation of the system allows for the implementation of complex signal processing algorithms in low-resource environments with low system clock rates. The system is especially efficient for subband processing since the configurable WOLA coprocessor efficiently splits the fullband input signals into subbands, leaving the core free to do adaptive processing on the subband signals.

The ASL algorithm is implemented on the DSP system using a 32-band oddly stacked oversampled WOLA filterbank configuration [6]. With an analysis window length of 128 samples and input block size of 8 samples, the system group delay is 6.5 msec. For this application, the system uses an 8.96 MHz clock and a sampling rate (F_S) of 16 kHz. The system operates on 1.25 V with a power consumption of 2.6 mW.

3. Acoustic-Shock Limiting

The acoustic-shock limiting (ASL) algorithm is implemented on the DSP system described in Section 2. Figure 1 shows the block diagram of the ASL system. The input signal is passed through the analysis filterbank and is split into N uniform subbands. The output of the analysis filters are adjusted by digital gains $\{g_1(i)\ldots g_N(i)\}$, which are determined by the ASL algorithm to eliminate acoustic shock phenomenon at the DSP output. The ASL algorithm set the value of $g_k(i)$ to less than 1 when a shock occurs in subband k. Otherwise, $g_k(i)$ is set to 1. The processed analysis filterbank outputs are assembled back to a time-domain signal using a synthesis filterbank.

The ASL algorithm modules consist of three major components: the parameter extraction module, the shock detection module and the gain calculation module. These modules process the input signal in both the time and frequency domains. Specifically, the parameter extraction block detects sudden increases in input energy level by measuring the time-domain energy of the input signal, while the shock detection and gain calculation modules suppress the signal in specific subbands by controlling the gains $\{g_k(i)\}$. Details of each major block are provided below.

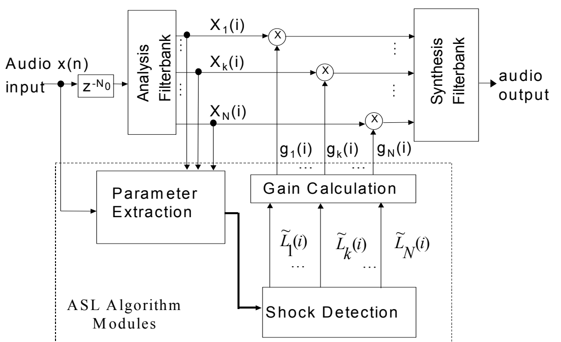

Figure 1: Acoustic-shock limiting (ASL) system.

3.1. Parameter Extraction

The parameter extraction module measures the energy of the input signal both in time-domain and within each subband. These energy measurements, together with their exponential-time averages, will be used for shock detection and gain calculation.

The time-domain energy is calculated as:

$$L(i) = 10\log\left(\frac{1}{R}\sum_{j=1}^{R} x^2(Ri+j)\right), \tag{1}$$

where $x(n)$ is the input signal; i is the data block index; $L(i)$ is the energy of the i^{th} data block; R samples is the data block length. With a 16 kHz sampling rate, $L(i)$ measures the input signal energy with 0.5 msec resolution. This allows sudden increases in the sound level to be detected. An artificial digital delay of N_0 samples (i.e. z^{-N_0} in Figure 2) is inserted between the $L(i)$ measurement and the WOLA analysis. This delay allows the time-domain measurement to notify the shock-detection module about the presence of high-level input before this shock reaches the filterbank, thus providing extra time for the shock-detection module to react to the shock. This delay is typically set to a multiple of R samples.

The energy of the input signal in each subband is calculated as:

$$L_k(i) = 20\log\left(\left|X_k(i)\right|\right), \tag{2}$$

where $X_k(i)$ and $L_k(i)$ are the WOLA analysis result and the energy for the k^{th} subband of the i^{th} data block, respectively. When a sinusoid of frequency $f_{c,k}$ is applied to the DSP input, where $f_{c,k}$ is the center frequency of subband k, $L(i)$ equals $L_k(i)$.

The parameter extractor also calculates the exponential-time averages of $L(i)$ and $L_k(i)$ using fast and slow time constants. For the rest of the paper, $L_{fast}(i)$, $L_{slow}(i)$, $L_{k,fast}(i)$ and $L_{k,slow}(i)$ denotes the fast and slow averages of $L(i)$ and $L_k(i)$, respectively. $L_D(i)$ and $L_{D,k}(i)$ represents $(L_{fast}(i) - L_{slow}(i))$ and $(L_{k,fast}(i) - L_{k,slow}(i))$ respectively.

3.2. Shock Detection

Based on the measurements obtained by the parameter extraction module, the shock detection module detects if shock has occurred in a particular subband, and determines the appropriate subband energy measurement to be used for the gain calculation.

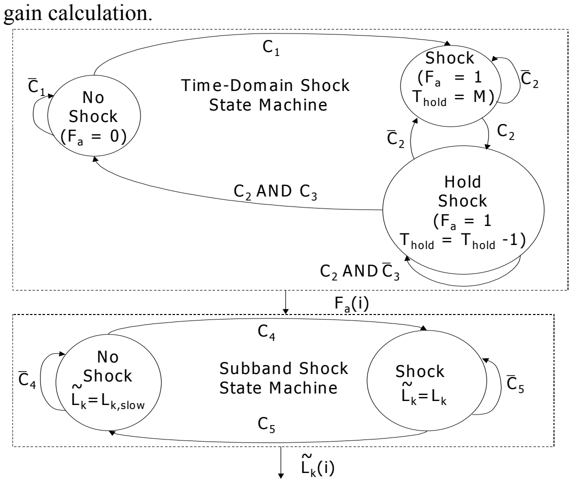

Figure 2: Shock detection state machines.

Condition in Figure 2	Description
C_1	$(L_D > D_{time})$ OR $(L > L_{max})$
C_2	$(L_D \le 0)$ AND $(L < L_{max})$
C_3	$T_{hold} == 0$
C_4	$(L_{D,k} > D_{band})$ OR $(F_a == 1)$
C_5	$(L_{D,k} \le 0)$ AND $(F_a == 0)$

Table 1: Descriptions of conditions C_1 to C_5 in Figure 2.

Figure 2 shows the state machines used to perform shock detection. Conditions C_1 to C_5 in Figure 2 are described in Table 1. The Time-domain Shock state machine checks if the input energy level has increased dramatically or has reached an excessive level. These two conditions are indicated by $L_D>D_{time}$ and $L>L_{max}$ in condition C_1, where D_{time} and L_{max} represent the values of $L_D(i)$ and $L(i)$ when acoustic shock occurs. If one of the above conditions is true, the Time-domain Shock state machine will enter the shock state and set the shock flag $F_a(i)$ to 1. This notifies all the subbands about the occurrence of loud sound. A "hold-shock" state and a hold time T_{hold} are introduced to compensate for the digital delay in the parameter extraction module. They keep the $F_a(i)$ asserted for the next M data blocks until the WOLA analysis results for the current data block are calculated.

The Subband Shock State Machine determines if shock has occurred in each individual subband by examining the subband energy levels and the shock flag $F_a(i)$. The same state machine is used for each subband. D_{band} is the value of $L_{D,k}(i)$ when a narrowband shock disturbance occurs. The outputs of the subband shock state machines are $\{\tilde{L}_k(i)\}$, $k \in \{1 \ldots N\}$. Each $\tilde{L}_k(i)$ is used in the gain calculation module to calculate the gain $g_k(i)$ for the k^{th} subband. If shock occurs in subband k, $L_k(i)$ is used as $\tilde{L}_k(i)$. Since $L_k(i)$ is an instantaneous measurement of the subband energy, the gain determined using $L_k(i)$ ensures that shock is suppressed responsively. On the other hand, if subband k is not in shock, the slow average $L_{k,slow}(i)$ should be used as $\tilde{L}_k(i)$ such that the resulting gain is steadier and less affected by noise.

Currently, the values of D_{time} and D_{band} are heuristically determined. In the future, clinical experiments can be performed to determine these values.

3.3. Gain Calculation

Acoustic shock phenomenon is eliminated by applying the appropriate gain $g_k(i)$ to the signal in each subband k. The gains are determined by using static input-gain (IG) curves for each subband. The IG curve for subband k describes the relationship between the gain $g_k(i)$ and $\tilde{L}_k(i)$. By using a separate IG curve for each subband, frequency-dependent SPL limiting can be achieved. The IG curves are implemented as lookup tables in the DSP system. To calculate the gain $g_k(i)$ for a particular $\tilde{L}_k(i)$, linear interpolation is performed between table entries. In Section 4, we describe the design and calibration method for the IG curves.

4. Algorithm Testing and Calibration

4.1. Test setup

Figure 3 shows the test setup that is used to evaluate the performance of the proposed ASL algorithm. Using a desktop computer and an audio processing software, electrical test signals are applied to the DSP input via the computer's sound card. The gain of Amplifier 1 is adjusted to provide the DSP input with signals of different amplitudes. A supra-concha style headset is used to provide acoustic signal to an ITU-T type 3.3 artificial ear inside a head-and-torso (HAT) simulator (Brüel & Kjær Type 4128C). The electrical output of the artificial ear is digitally recorded by the desktop computer, where A-weighted narrowband SPL at different frequencies are calculated.

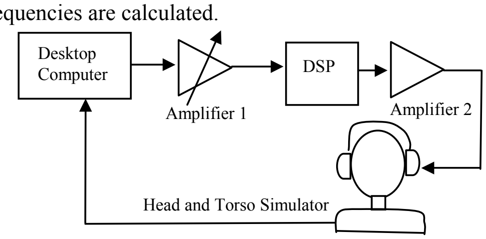

Figure 3: Acoustic Test Setup

To demonstrate the shock-limiting ability of the proposed algorithm, the DSP runs in two modes: the shock-limiting mode and the pass-through mode, which represent the cases where the ASL algorithm is activated and de-activated, respectively. In the pass-through mode, the same constant gain is applied to all the analysis filterbanks output. This gain is equal to the maximum gain used in the shock-limiting mode.

4.2. Algorithm Calibration

In this section, we propose a simple calibration procedure for the IG curve used in each subband. This calibration procedure takes into account the electro-acoustic characteristic of the communication equipment. Using this procedure and the test setup in Figure 3, the ASL algorithm can be calibrated such that the acoustic output of a specific communication equipment is limited to pre-defined frequency-dependent SPL limits.

Assume we want to limit the narrowband SPL of the headset output to below 65 dB at frequency $f_{c,k}$, where $f_{c,k}$ is the center frequency of the k^{th} subband. Using the test setup in Figure 3, we apply a sinusoid of frequency $f_{c,k}$ at the input of the DSP, which is running in the pass-through mode. Varying the gain of Amplifier 1 in Figure 3, we perform two measurements: 1) the narrowband SPL (in dB) of the headset acoustic output at frequency $f_{c,k}$ as received by the artificial ear; 2) the energy of the k^{th} subband in dB (i.e. L_k) as calculated by the DSP. Since the frequency of the input sinusoid corresponds to the center frequency of the k^{th} subband, both the narrowband SPL and L_k are almost constant. L_k and the narrowband SPL at $f_{c,k}$ satisfy the relationship:

$$\text{Narrowband SPL} = 20\log(g_{\mathrm{pt,k}}) + L_\mathrm{k} + P_\mathrm{k}\,, \tag{3}$$

where $g_{\mathrm{pt,k}}$ is the constant gain applied to the k^{th} subband in pass-through mode, and P_k is a calibration factor. In order to limit the narrowband SPL at $f_{c,k}$ to 65 dB, the gain g_k in shock-limiting mode must be chosen to satisfy the following relationship:

$$20\log(g_k) + L_k \le GL_{\max}(k)\,, \tag{4}$$

where $GL_{\max}(k)$ is calculated by solving for $20\log(g_{\mathrm{pt,k}}) + L_k$ in Equation (3) when narrowband SPL is 65 dB.

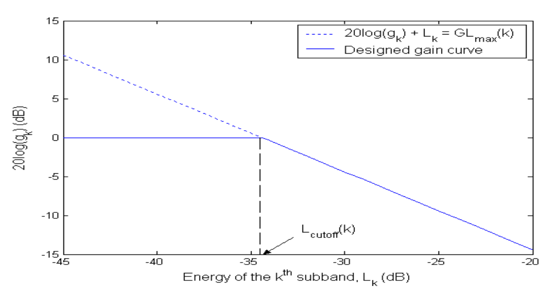

Figure 4: Design of input-gain (IG) curve.

Figure 4 shows the line $20\log(g_k)+L_k = GL_{\max}(k)$ (i.e. the doted line). To limit the narrowband SPL to 65 dB, the IG-curve should be on the left-hand side of the doted-line. The solid curve is the designed IG-curve. It overlaps with the $20\log(g_k)+L_k = GL_{\max}(k)$ line when $L_k > L_{\mathrm{cutoff}}(k)$. This IG curve implies that unity gain is used if L_k is less than $L_{\mathrm{cutoff}}(k)$. Furthermore, the narrowband SPL will be fixed at 65 dB when $L_k > L_{\mathrm{cutoff}}(k)$.

The above procedure can also be used to calibrate the IG curves such that the narrowband SPL at frequency f_e is below a limit, where f_e is close to the edge of the $(k-1)^{th}$ and the k^{th} subband. In this case, a sinusoid of frequency f_e is applied at the DSP input, and $BL_{cutoff}(k)$ and $BL_{cutoff}(k-1)$ are calibrated simultaneously. This is done to account for the leakage between adjacent analysis filters.

The above calibration procedure assumes that the energy of a shock disturbance is concentrated in a particular subband. Work is underway to address the case where shock disturbance energy is spread across several subbands.

5. Results

Using the test setup in Figure 3, the acoustic shock protection ability of ASL algorithm is evaluated. We apply a signal $s_{in}(t)$ to the DSP input, where $s_{in}(t)$ contains a shock disturbance embedded in speech. Four sinusoids (1.5, 2, 2.5 and 3 kHz) of 2 seconds duration are generated simultaneously to create the shock disturbance. When the shock occurs, the ratio between shock disturbance and speech energy is 24.5 dB.

Figure 5 shows the narrowband SPL of the sound recorded at the ear-drum reference point (DRP) of the artificial ear. A darker shade of gray represents a higher SPL value, while black represents an SPL value exceeding the 65 dB limit. The bottom graph in Figure 5 shows that when the DSP is in shock-limiting mode, the shock disturbance at 1.5, 2, 2.5 and 3 kHz is limited to below 65 dB, while the SPL values in the rest of the spectrum are relatively unaffected by the ASL algorithm.

Next, we evaluate the speech quality of the ASL algorithm output by measuring the log area ratio (LAR) distance [7] between $s_{lm}(t)$ and $s_{ref}(t)$, where $s_{lm}(t)$ is the DSP output during shock-limiting mode when $s_{in}(t)$ is applied at the DSP input, and $s_{ref}(t)$ is the DSP output during pass-through mode when the speech in $s_{in}(t)$ is applied at the DSP input. The LAR metric is used because of its high correlation with subjective speech quality. A small LAR distance indicates that the distorted speech ($s_{lm}(t)$) closely resembles the original speech ($s_{ref}(t)$). For comparison, we also measure the LAR distance between $s_{pt}(t)$ and $s_{ref}(t)$, where $s_{pt}(t)$ is the DSP output during pass-through mode when $s_{in}(t)$ is applied at the DSP input. The signals $s_{ref}(t)$, $s_{pt}(t)$ and $s_{lm}(t)$ are digitally recorded before Amplifier 2 in Figure 3. All the signals are normalized to have the same energy before LAR distances are computed.

Figure 6 shows the LAR distance between $s_{lm}(t)$ and $s_{ref}(t)$ (solid line), as well as the LAR distance between $s_{pt}(t)$ and $s_{ref}(t)$ (dotted line). When acoustic shock occurs (i.e. during the time intervals 150 to 280 and 350 to 490), the average LAR distance for shock-limiting mode and pass-through mode are 8.02 and 14.50, respectively. This indicates that the ASL algorithm improves the speech quality at the DSP output by suppressing the shock disturbance.

6. Conclusions

In this paper, we presented a subband-based acoustic-shock limiting algorithm that is implemented in an ultra-low power DSP system with low group delay. The implemented algorithm limits narrowband SPL at different frequencies to predefined limits. Furthermore, we presented a calibration procedure that configures the implemented algorithm to work with different communication equipment.

Future work will include testing the current algorithm with narrowband noise. Furthermore, broadband shock disturbance limiting ability can be added to the ASL algorithm. This can be achieved by making the gain calculation adaptive to the flatness of the signal spectrum. Finally, subband implementations of adaptive line enhancer [8] can be incorporated to detect and virtually eliminate periodic acoustic shock disturbance in each subband.

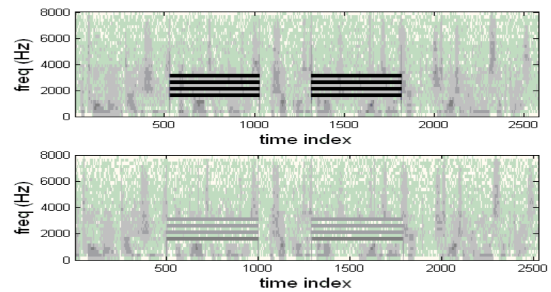

Figure 5: Narrowband SPL at the DRP displayed as a function of frequency and time when DSP is in pass-through mode (top) and shock-limiting mode (bottom).

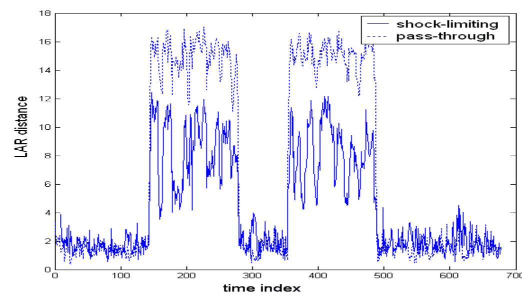

Figure 6: LAR distance between the reference speech and the DSP output signal when "speech + shock disturbance" is applied at the DSP input.

7. References

[1] Lu, L., "A Digital Realization of Audio Dynamic Range Control", *Proc. IEEE Int. Conf. On Signal Processing, pp. 1424-1427, 1998*.

[2] Park, S-K., "Preventing Acoustic Shock in a Telephone", *UK Patent Application GB 2,231,236 A*. Nov. 7, 1990.

[3] Brennan, R. and Schneider, T., "A Multichannel Compression Strategy for a Digital Hearing Aid", *Proc. ICASP-97, pp. 411-414, 1997*.

[4] Brennan, R. and Schneider, T., "Filterbank Structure and Method for Filtering and Separating an Information Signal into Different Bands, Particularly for Audio Signal in Hearing Aids", *United States Patent 6,236,731*. WO 98/47313. Apr.16, 1997.

[5] Brennan, R. and Schneider, T., "A Flexible Filterbank Structure for Extensive Signal Manipulations in digital Hearing Aids", *Proc. IEEE Int. Symp .Circuits and Sytems, pp.569-572, 1998*.

[6] Crochiere, R. E. and Rabiner, L. R., *Multirate Digital Signal Processing*, Prentice-Hall, New Jersey, 1983.

[7] Quackenbush, S. R. et. al., *Objective Measures of Speech Quality*, Prentice Hall, New Jersey, 1988.

[8] Haykin, S., *Adaptive Filter Theory*, Prentice-Hall, New Jersey, 3rd Edition, 1996.