

An Ultra-Low Power, Miniature System for Detecting and Limiting Acoustic Shock in Headsets

Etienne Cornu

etienne.cornu@dspfactory.com

Tina Soltani

tina.soltani@dspfactory.com

Dave Hermann

dave.hermann@dspfactory.com

Hamid Sheikhzadeh

hsheikh@dspfactory.com

Robert Brennan

robert.brennan@dspfactory.com

Dspfactory Ltd., 611 Kumpf Drive, Unit 200, Waterloo, Ontario, N2V 1K8, Canada
+1.519.884.9696

1. Introduction

In a number of applications, headset users may be subject to undesirable loud sounds that can produce a number of negative physiological effects. The phenomenon of a sudden loud noise and the resulting user reaction is sometimes referred to as acoustic shock. A shock may originate from a telecommunications line, for example in the form of loud DTMF tones or in environments such as airport runways and battlefields. A simple solution to controlling acoustic shock is to limit the overall signal energy to below a given threshold. Unfortunately, this method does not differentiate the shock from the speech signal and both get attenuated equally. Speech is then lost in the acoustic shock signal and can no longer be understood.

This paper presents a system that detects and limits acoustic shock while attempting to preserve a high level of speech intelligibility during the shock. The system is based on an ultra-low power, miniature DSP system designed specifically for speech processing. We present the architecture of the DSP and describe its processing units. Of particular interest is the weighted overlap-add (WOLA) co-processor [1], which efficiently implements an over-sampled filterbank that performs transformations between the time and frequency domains. This DSP architecture allows for an efficient shock detection and limiting algorithm that operates in both time and frequency domains. The algorithm first detects shock situations by comparing broadband and subband energies to calibrated thresholds. Depending on the shock state (onset, steady-state or tail of the shock), an adaptive combination of broadband attenuation and subband attenuation is applied in each subband. The method for calculating and combining the

two types of gains are designed to ensure smooth gain control and preservation of speech quality. In the time domain, the algorithm ensures that no acoustic shock is present at the system output; in the frequency domain, subband gains are calculated so that speech quality is optimally preserved by attenuating only the shock carrying bands; and in the time domain, the input-gain relationship provides a smooth transition between the linear region (when output equals input) and the limiting region (when the output energy level must remain below a threshold) in order to avoid undesirable artifacts during fluctuations of the input signal energy. The shock detection and limiting algorithm is mapped to the DSP system including the WOLA co-processor, achieving a group delay of 8 milliseconds and a power consumption of about 2.5 mW at 1.25 V. Experimental results show how speech intelligibility is improved when using this system compared with a system that simply compresses the entire signal to make it fit under a given threshold.

In section 2, we describe the problems associated with the processing of acoustic shock in real-time, and in section 3 we present a DSP system onto which our algorithm is implemented. Sections 4, 5 and 6 describe the details of the algorithm. In section 7 we present the methods used for achieving real-time performance on the DSP. Finally, in section 8 we address the speech quality issues and in section 9 we present the conclusions.

2. Acoustic Shock Processing

In a sense, an acoustic shock is a loud noise added to a signal. The usual method for dealing with acoustically perturbed signals is to use noise reduction algorithms. These algorithms are usually optimized for relatively slowly varying

noise and SNR levels near 0 dB. They are not suited for acoustic shock processing for several reasons. First, acoustic shocks occur suddenly and unexpectedly. It is necessary to protect the headset user from these shocks immediately. Even the moderate amounts of leakage through the shock protection system may damage the user's ear. Shocks are also very loud; well beyond the processing capabilities of noise reduction algorithms. Due to these properties, alternative methods must be employed to ensure the best possible working conditions for headset users. In order to meet these demands, an acoustic shock limiting system should have the following characteristics:

- a. A wide input dynamic range so that only extremely loud signals are clipped.
- b. Sufficient dynamic range in the low amplitudes to preserve speech quality in normal conditions.
- c. Low group delay, to avoid introducing problems such as echo.
- d. Immediate response in case of shock.
- e. Unity gain response (output equals input) at low amplitudes.
- f. Infinite dynamic range compression at high amplitudes, i.e the output level be limited to a pre-defined maximum.
- g. Good speech quality in case of shock.

Several existing methods have been used to prevent headset users from experiencing acoustic shock. They include high-level limiting using automatic gain control [2] and clipping of high-level signals using diodes or similar devices [3]. While these methods provide some protection against acoustic shock, they do not maintain speech quality or have a wide dynamic range. In particular, they fail to identify and attenuate the portions of the spectrum affected by the shock disturbance, thus causing audio dropouts, intermodulation and harmonic distortion. The system presented in the following sections addresses these issues.

3. The DSP System

The acoustic shock detection and limiting algorithm is implemented on an Application Specific Signal Processor (ASSP) that includes all the necessary digital and analog components to operate as a single-chip audio processing solution. It features two A/D and two D/A converters, pre-amplifiers, a timer, peripheral

interfaces such as serial and GPIO pins, data/program memory, a 16-bit DSP core, a dedicated Input/Output (IO) processor and the WOLA filterbank co-processor. Power consumptions for typical applications ranges from less than 1 mW for hearing-aid applications up to around 5 mW for complex headset applications. Figure 1 illustrates the overall architecture of this DSP system.

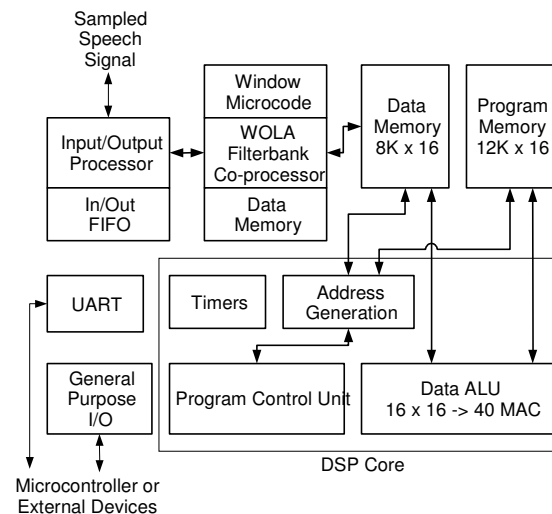


Figure 1 – DSP system architecture

The DSP core, IO processor and WOLA co-processor all operate in parallel. The IO processor provides the interface between the analog and digital components of the system. It reads digitized input samples and stores them into the input FIFO. The input samples are provided in frames of R samples, where $R=8$ for this application. These frames are the unit of processing for the system. The DSP core and WOLA co-processor both have access to the input FIFO. The WOLA co-processor transforms the samples in the input FIFO into the frequency domain and stores them in shared memory. For a configurable transform size of N , the frequency domain data consists of $N/2$ complex numbers where $N=64$. The algorithms on the DSP core can process the signal in the time domain (input FIFO) or in the frequency domain (shared memory). The WOLA co-processor applies frequency domain gains calculated by the DSP core to the frequency domain signal, converts it back to the time domain and stores the output time domain samples in the output FIFO. The algorithms on the DSP core have the opportunity to perform operations on the out-going signal

before it is read out by the IO processor, which then sends it out to be converted back to an analog signal in frames of the same size. Figure 2 summarizes the signal path through the system.

One major advantage of this architecture is that several signal processing algorithms can operate at the same time on the DSP without affecting each other. For example, an echo cancellation algorithm may operate on the DSP at the same time as the acoustic shock algorithm, although they may be developed independently.

The group delay through the system, i.e. the time it takes for samples to be processed, is a function of the analysis window length (L_a), the synthesis window length (L_s), the frame size (R) and the sampling frequency. Depending on how these parameters are chosen, it can be as low as 5 milliseconds; typical applications where the DSP is used require a group delay below 10 milliseconds.

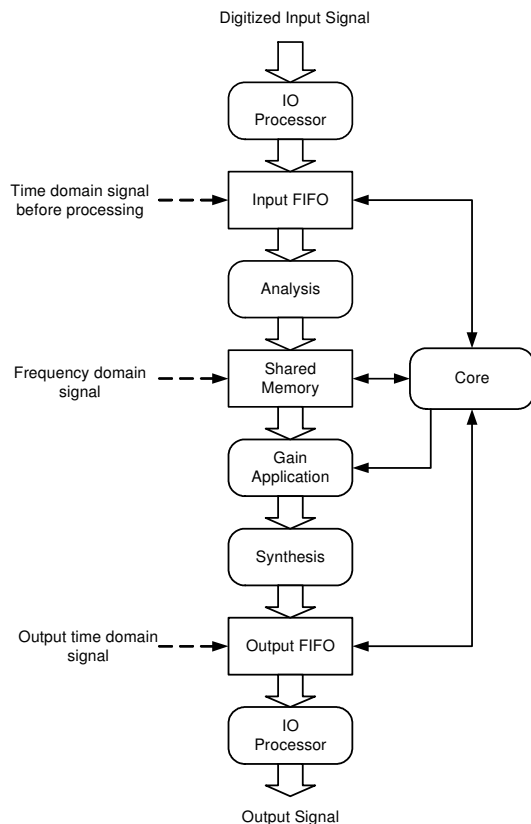


Figure 2 – Signal path

4. Subband Shock Processing

In some cases, a broadband acoustic shock event can completely obscure the signal because it is significantly louder than the underlying signal. In other cases, the shock is still very loud, but is concentrated in specific frequencies, such as with DTMFs and police sirens. In other cases, shock processing can be performed on a subband basis. The shock carrying subbands may be attenuated while the other subbands are left untouched. In the worse case, a broadband shock will appear to occur in all subbands, and then all will be attenuated. Through this subband processing scheme, speech quality will be optimized. .

In the DSP system, shocks are detected by comparing the energy in each subband with a pre-defined threshold. The general approach consists of applying a gain that attenuates the subband energy to a point below this threshold. If the shock occurs in a limited number of subbands, most subbands will not be affected and speech quality can be improved to a certain level. Subband energy levels are calculated by the DSP core using the frequency domain data generated by the WOLA co-processor (stored in shared memory). The DSP core also calculates the gain to apply to each subband and transmits it to the WOLA co-processor to be used by the gain application process

The subband gains are calculated solely based on the energy carried by each band, regardless of the overall shape of the spectrum. As the subband energy level increases, leakage to adjacent subband starts to occur. Since the energy level in the adjacent subbands may still be lower than the threshold, shock is only reduced by attenuating the main shock-carrying subband. As a result, the maximum output energy increases as the shock level increases, until the leakage in adjacent bands reach the threshold as well.

Another source for maximum output variation is the location of the shock frequency relative to subband centre frequencies. When the shock frequency is between two adjacent subbands, its energy is distributed evenly between the two bands; if it is located at the centre of a subband, the energy is observed primarily in that subband, with the exception of the energy leakage to adjacent subbands as previously mentioned. For a fixed threshold value, 3 dB more input shock energy is needed to reach the threshold when the

shock occurs between the subbands. As a result, the maximum output energy varies up to 3 dB as the shock frequency shifts from the centre of the subband to its edge. Figure 3 shows this variation by plotting the limiting behaviour of a tone in the centre of a channel, and the limiting behaviour of a tone between two adjacent channels.

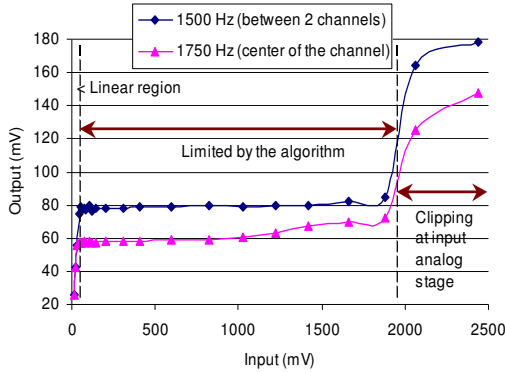


Figure 3 – System output using only subband gains

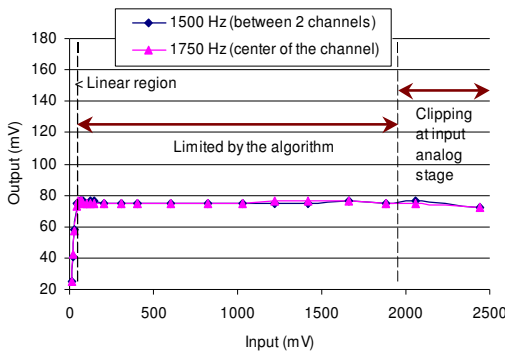


Figure 4 – System output using only broadband gains

To solve this problem we apply a second gain to each band, this one determined from the overall energy of the signal. Applying this gain ensures that adjacent subbands as well as other subbands get attenuated by a certain amount, causing all bands to have the same response. Figure 4 shows how frequencies within a channel and on channel edges are limited equally using only this gain. The combination of this gain with the subband gain is described in section 6.

5. Transient Effect

The transient effect is an overshoot of the output signal above the desired maximum threshold caused by the delay in detecting the shock in the

frequency domain. At every frame, the WOLA co-processor performs an analysis based on a window of 128 samples that includes 8 new samples. As the acoustic shock occurs, the samples become suddenly very large (in absolute value). Because of the windowing, this sudden surge in energy is only noticeable in the frequency domain a few frames after it occurs in the time domain. Until then, the shock cannot be detected and therefore no attenuation occurs, resulting in a loud click. This is shown for an acoustic shock signal in Figure 5a-b.

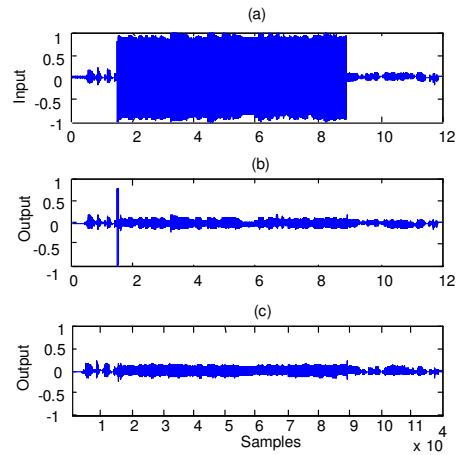


Figure 5 – (a) signal with shock, (b) after subband processing, (c) after transient effect processing.

To handle these transient effects at the beginning and at the end of a shock, a shock detector in the time domain is implemented. This detector looks at R samples at a time as they are stored in the input FIFO by the IO processor. The detector determines if the energy exceeds a threshold determined experimentally. When it detects a shock, it can determine an overall gain factor to be applied in each band in addition to the subband gain factor used for attenuating the subbands in shock. The result of this operation is illustrated in Figure 6c.

6. Gain Combination

As discussed in the previous sections, three types of gains are calculated:

- (a) A subband gain is used for attenuating subbands with energy higher than a specified threshold. This gain is calculated independently for each subband.

(b) An overall gain, used for attenuating all subbands when one or more bands are in shock and thereby reducing maximum output variations over frequency.

(c) A second overall gain, used for reducing the transient effects at the onset and tail end of an acoustic shock event.

The two overall gains are combined into a single gain. This gain is then combined at every frame with the subband gain depending on the status of the system:

- During the transient phase from non-shock state to shock state, the subband gains are set to zero since they are not in shock. The broadband gain, determined from the data in the input FIFO, is applied to all subbands.

- During the transient phase from shock state to non-shock state, only the subband gains are applied. The overall gain is not applied since there are no high energy samples in the input FIFO.

- During the shock steady phase, a factor r is used for weighting overall and subband shock in each subband. r is calculated as

$$r(i) = 10^{(A_{max}(i) - A_{total}(i))} , \quad (1)$$

where $A_{max}(i)$ and $A_{total}(i)$ are the maximum subband energy level and the total energy level for frame i measured in dB, respectively. $r(i)$ is one when the energy is concentrated in one subband, and decreases towards zero as the energy is spread over the spectrum. In each subband, the total gain $G_{cb}(i,k)$ is then calculated as

$$G_{cb}(i,k) = r(i)G_{sb}(i,k) + (1-r(i))G_{bb}(i) , \quad (2)$$

where $G_{bb}(i)$ is the broadband gain of frame i and $G_{sb}(i,k)$ is the subband gain of frame i and subband k . Both are expressed in dB.

The effect of the weighting factor is to put more emphasis on the subband gains for subbands with very high energy and more emphasis on the overall gain for frames with relatively low energy.

7. System Considerations

The acoustic shock algorithm presented in this paper is implemented on the DSP system described in section 3. In order to deploy the

algorithm in real-time, a number of system issues had to be considered.

The first major system consideration is the need for low group delay. The major part of the group delay is a result of the filterbank. The use of an over-sampled filterbank helps reduce the required window length [4] and hence reduces the overall delay. The total group delay for this algorithm was measured as 8 ms.

The second major system consideration is related to computational constraints. The low power consumption applications of the algorithm imply a low system clock, and hence there are limited computational resources for processing the signal. A number of features of the DSP system and algorithm optimizations help to reduce the computational requirements of the acoustic shock processing.

First, by utilizing the DSP core in parallel with the WOLA filterbank coprocessor, one can achieve greater computational efficiency. The DSP core is used to process the frequency and the time domain data while the WOLA coprocessor performs analysis, gain or synthesis and the IO processor manages the input and output FIFOs. To take advantage of this parallelism, the acoustic shock limiting algorithm is divided into seven functions, where three of them are performed in parallel with WOLA coprocessor and the rest are applied between the filterbank applications. Shared memory is used to transfer data between the three processors.

We can further maximize this parallel processing by carefully scheduling the required system calculations. Since a significant part of the final gain calculation is performed and applied in the frequency domain, it is required to have all the subband calculations performed before the subband gain application. This would normally prevent the algorithm from using the WOLA coprocessor in parallel with the DSP core during the gain and synthesis stages of the WOLA processing, since all the calculations are already completed prior to these processes. In order to alleviate this problem, the subband gains, calculated from the band levels of a given block i , are applied to the subband signals of the following block, $i+1$. By doing so, the frequency data generated by the analysis filterbank can be processed during both gain and synthesis applications. This significantly improves the

computational efficiency of the system, since the analysis, gain application and synthesis stages of the filterbank can consume considerable amounts processing time. The only drawback to this scheduling is that the subband gain application lags the input to the system by one input frame, possibly delaying the response of the system when a shock occurs. However, the broadband gain is calculated and applied to the corresponding time domain samples at every frame.

Another method for reducing the computational complexity is to reduce the number of subbands that require processing. This reduces the frequency resolution of the system, leading to a less accurate processing of a shock. Instead, we group the existing subbands into a smaller number of channels in the frequency domain, where the channel groupings correspond roughly to the critical bands. The gain calculations are then based on the grouped channel levels.

8. Speech quality

One of the main reasons for choosing a subband approach is the fact that under certain circumstances it is possible to preserve most of the speech signal during a shock. In the case of DTMF tones or a police sirens, the shock signal is concentrated in a few frequency bands, and attenuating these bands while the shock occurs over a speech signal should result in improved speech quality. To illustrate this, we compare the output signal of the DSP system to the input signal by using the log area ratio (LAR) distance [5]. In Figure 7, both the input and output signal are compared to the original unperturbed speech signal sample by sample, and the difference is expressed as the distance between the two signals. As Figure 6 shows, the output signal matches the unperturbed speech closer than the input signal, thus indicating an improvement in speech quality.

9. Conclusion and Future Work

We have presented a system for detecting and limiting acoustic shock in headsets that are deployed in a miniature DSP consuming less than 2.5 mW of power. The group-delay though the system is only 8 ms, making it suitable for deployment in headsets in applications such as call centres. The algorithm performs acoustic shock detection and limiting in the time and frequency domains, allowing it to attenuate the

shock while preserving speech quality in a variety of situations.

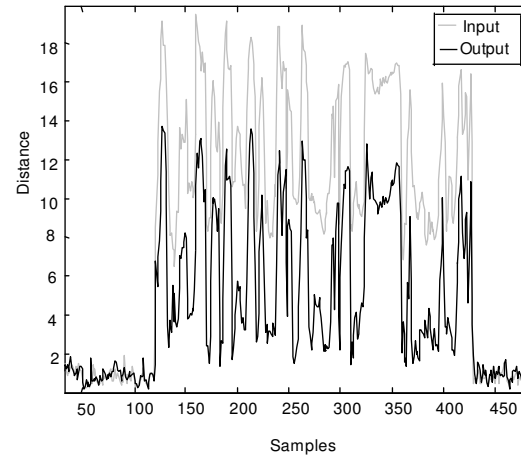


Figure 6 – Speech quality comparison

We expect that in the future there will be a demand for acoustic shock processing systems that not only protect the user from shocks but also keep track of the user's exposure to loud signals during a period of time. We plan to add this capability to the system by integrating long-term energy levels in the time and frequency domains.

10. References

- [1] Schneider, T., Brennan, R., "An ultra low-power programmable DSP system for hearing aids and other audio applications", *Proc. ICSPAT 1999*.
- [2] Lu, L., "A Digital Realization of Audio Dynamic Range Control", *Proc. IEEE Int. Conf. On Signal Processing*, pp. 1424-1427, 1998.
- [3] Park, S-K., "Preventing Acoustic Shock in a Telephone", UK Patent Application GB 2,231,236 A. Nov. 7, 1990.
- [4] R. L. Brennan, and T. Schneider, "A flexible filterbank structure for extensive signal manipulation in digital hearing aids," *Proc. IEEE Int. Symp. Circuits and Systems*, 1998.
- [5] Quackenbush, S.R. et al, *Objective Measures of Speech Quality*, Prentice Hall, New Jersey, 1988.