

An Acoustic Shock Limiting Algorithm Using Time and Frequency Domain Speech Features

T. Soltani, D. Hermann, E. Cornu, H. Sheikhzadeh and R. L. Brennan

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

Abstract

The phenomenon of acoustic shock occurs when a headset user is subjected to an audio disturbance at an uncomfortable or unsafe level. In this paper, a new acoustic shock limiting algorithm is proposed which reduces the output level by utilizing both time and frequency domains limiting features. During acoustic shocks, the algorithm improves the intelligibility of the underlying speech and minimizes the artifacts by reducing the shock in the subbands based on the narrowband and broadband characteristics of the input signal. The algorithm is implemented on a low power DSP system where the input data is analyzed in both time and frequency domains. The method is tested for sinusoidal and speech inputs. The results demonstrate shock compression and limiting, with reduced speech distortion during shock onset and good overall speech quality.

1. Introduction

As specified in international standards such as ITU-T P.360 [1], headsets used in telecommunication environments are required to be equipped with output limiting features for safety reasons. This is also the case for headsets worn in very noisy environments such as airports and battlefields. Usual methods limit the absolute sound pressure level (SPL) at the eardrum to a pre-defined maximum, based on recommended safety thresholds and without distinguishing between shock and speech signal. This can cause the output speech signal to be heavily masked by the shock.

Choy et al. [2] describe a low-delay subband-based approach which attempts to maintain the speech quality while compressing the acoustic shock. In their approach, acoustic shock is detected independently in each frequency band and shock limiting is applied only in the bands in which the shock occurs. In the case of dual tone multiple frequency tones (DTMF), for example, only a few bands are affected by shock and consequently applying this method results in maintaining a relatively high degree of speech intelligibility during the shock. However, for some shock conditions limiting in individual subbands might not be sufficient. For example for a given threshold value, the actual peak level of the output signal depends on the position of the frequency of the input signal relative to the center frequencies of the filterbank. Also, because of the inherent delay through the analysis filterbank, the shock is not observed in the frequency domain until several samples after its actual occurrence. As a result, the first few shock-carrying samples are not compressed and an overshoot is observed at the beginning of the shock, often resulting in a loud click. We refer to this as the transient effect. Finally, in [2], when the shock is very loud and occurs in a small number of subbands (like DTMF tones), leakage to adjacent subbands is compensated by additional compression

of the subbands in shock. This sometimes distorts the speech and creates undesirable artifacts.

In this paper, we present an algorithm that addresses these issues by applying a number of novel methods for shock detection and limiting in both the time and frequency domains. The transient effect is also addressed, resulting in instantaneous shock attenuation with no overshoot. The subband attenuation is calculated as a combination of the narrowband and broadband characteristics of the shock signal. This method produces a smooth transition between non-shock and shock states and minimizes audible artifacts.

Section 2 of this paper presents an overview of the acoustic shock detection and limiting algorithms which are explained in more detail in Section 3. In Section 4, the algorithm implementation on a DSP system, using a weighted overlap-add (WOLA) filterbank, is described and results are presented. Finally, in Section 5 conclusions and future work on this algorithm are stated.

2. Algorithm overview

The acoustic shock limiting algorithm is divided into two main sections: shock detection and gain calculation, as explained in Section 3. In general, the information generated by the shock detectors determines the existence and nature of a shock and the required amount of limiting. The block diagram shown in Figure 1 illustrates the data flow and parameter extraction. At the beginning of signal processing in the time domain, the algorithm determines three broadband parameters (broadband shock flag, broadband gain, and transient counter) for each block of R input samples. An over-sampled WOLA filterbank [3] with prototype filter of length (window size) L is then used to obtain the subband signals. Subband shock detection and attenuation calculations are performed based on the individual subband energy. As shown in Figure 1, the final gain in each subband (a value between 0 and 1) is evaluated as a combination of the subband and broadband gains, based on overall shock detection flags and a gain ratio factor. Gains are then applied to each band and the resulting subband signals are converted back to the time domain using the synthesis filterbank.

Both the shock detection and gain calculation are based on the instantaneous or average energy levels of the broadband signal and the subband signals. Instantaneous energies are used during periods of large signal changes, i.e. at the start of sudden shocks; otherwise, the average energies are applied to minimize distortion from gain modulations.

3. Shock processing

Using a broadband gain is advantageous because it preserves the shape of the spectrum, thereby reducing artifacts or speech distortion. As will be shown, it is also applied for

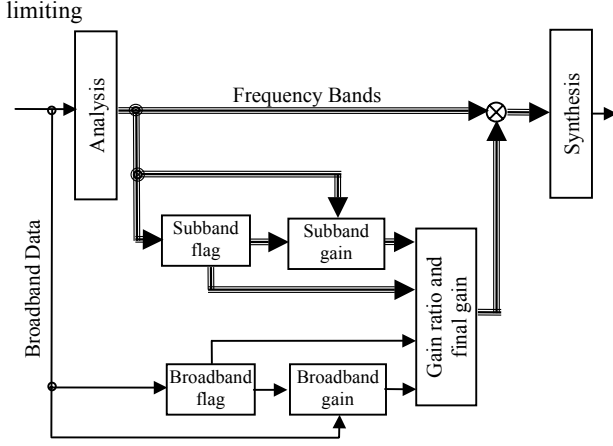


Figure 1: Block diagram of the acoustic shock limiting algorithm

the overshoot effect at the beginning of a shock. The drawback of using only the broadband gain for shock limiting is that the shock and the speech embedded in the shock are equally attenuated and consequently, the speech may no longer be audible. The opposite approach, consisting of using only subband gains as described in [2], has the advantages and disadvantages explained in Section 1. In the proposed acoustic shock limiting algorithm, a combination of both gain types is applied to the input signal to minimize speech quality degradation during shock. Therefore, the final gain applied to each individual subband, $G_{cb}(i,k)$, is evaluated as:

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

where $G_{bb}(i)$ is the broadband gain of block i and $G_{nb}(i,k)$ is the subband gain of block i and band k . Both are expressed in dB. The function $r(i)$ is the gain ratio, indicating the contribution of each gain value in reducing the shock.

The gains $G_{bb}(i)$ and $G_{nb}(i,k)$ are non-zero only when the energy levels are higher than the threshold values. $G_{bb}(i)$ is calculated as follows:

$$\begin{cases} G_{bb}(i) = 0 & \text{if } L(i) \leq T_{bb} \\ G_{bb}(i) = L(i) - T_{bb} & \text{if } L(i) > T_{bb} \end{cases} \quad (2)$$

where $L(i)$ is the total block energy, and T_{bb} is a calibrated threshold value. Similarly, the subband gain $G_{nb}(i,k)$ is calculated as follows:

$$\begin{cases} G_{nb}(i,k) = 0 & \text{if } A(i,k) \leq T_{nb}(k) \\ G_{nb}(i,k) = A(i,k) - T_{nb}(k) & \text{if } A(i,k) > T_{nb}(k) \end{cases} \quad (3)$$

where $A(i,k)$ is the energy in band k and $T_{nb}(k)$ is a calibrated threshold value.

As shown in Eq. 1, the behavior of the system is dependent on the value of the function $r(i)$. This function is evaluated based on the characteristics of the input signal. Each input sample block is classified based on a set of flags derived by the shock detection modules in both time and frequency domains, and its position relative to the beginning of the shock. The shock detectors used in this algorithm are identical to those employed in [2]. In summary, the time

domain shock detector sets the value of a flag, $F_t(i)$, based on

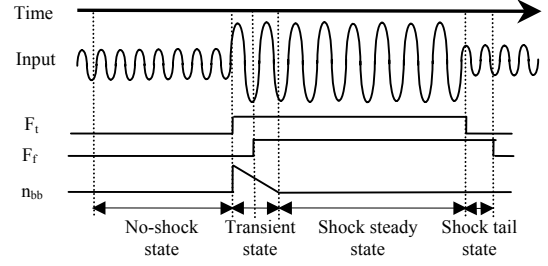


Figure 2: Shock detection path

the energy level of the block i relative to a given threshold value and the energy level of the previous blocks. Similarly, the frequency domain shock flag, $F_f(i,k)$, for band k of block i is determined by comparing its energy level to a preset threshold value and the energy level of band k in the previous blocks. The flags are set to 1 when a shock is detected and to zero otherwise. The combination of the values of these flags defines four states for the system: no-shock, transient, steady shock, and shock tail. Figure 2 shows different states of the system with their corresponding flag conditions and Figure 3 illustrates the relationship between the system states, the gain values and the gain ratio function.

At the beginning of a shock, the system goes into the transient state when the energy of R new time-domain samples exceeds the pre-defined threshold. This state lasts n_{bb} blocks, which is the time it takes for the shock to fully propagate into the subbands and is equal to approximately L/R blocks. As it is shown in Figure 2, this results in having two phases in this state, the first phase occurring before the subbands reach their threshold values and the second phase occurring after one or more of the subbands exceed the threshold but not yet reach the steady state shock level. During the first phase, the shock is not yet detected in the subbands, i.e. $F_t(i)=1$, $F_f(i,k)=0$ for all k subbands and consequently, the only available gain value is the broadband gain. Therefore the gain ratio $r(i)$ is set to zero and the final gain is forced to be equal to the broadband gain. As a result, during the first few blocks of the shock, all the subbands are attenuated by the same value, which reduces the transient effect.

The subband levels continue to rise as more blocks are accumulated in the analysis filterbank and the system eventually reaches the second phase of the transient state when the shock is detected in at least one subband, i.e. $F_t(i)=1$ and $F_f(i,k)=1$ for at least one of the subbands. This phase continues until the subbands reach their steady state shock energy levels. In this case, the subbands in shock are treated differently from the subbands not in shock by evaluating the gain ratio as the energy concentration factor over the spectrum,

$$r(i) = 10^{[A_{max}(i) - A_{total}(i)]} \quad (4)$$

where $A_{max}(i)$ and $A_{total}(i)$ are the maximum subband energy level and the total spectrum energy level for block i measured in dB, respectively. The broadband and narrowband gains for each subband are calculated based on Eq. 2 and Eq. 3, and the combinational gain is evaluated based on Eq. 1. As depicted

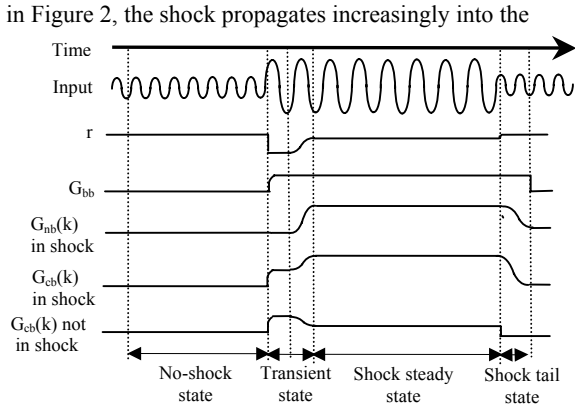


Figure 3: Gain calculation in different states

subbands as time advances and $r(i)$ becomes larger. Consequently, the contribution of the subband gain increases in the subbands in shock while the contribution of the broadband gain decreases in all subbands. This creates a smooth transition into the next state as a) subbands in shock are attenuated based mostly on their own energy level, b) subbands not in shock are attenuated based only on a portion of the total energy level, and c) the overall energy level is kept below the threshold.

At the end of the transient state, the shock steady state is entered and continues while $F_r(i)=1$ and $F_f(i, k)=1$ for at least one of the subbands. The ratio function, $r(i)$, is still calculated as in Eq. 4. As shown in Figure 3, subbands in shock undergo more attenuation than subbands not in shock.

The shock tail state includes the samples following the end of the shock. Figures 2 and 3 illustrate the flag settings and the gain variations during this state. In this case, the new R input samples do not carry any shock, but the input to the analysis filterbank still contains shock-carrying samples. This results in a higher apparent energy level of the signal in the frequency domain. As a result, the time domain flag, $F_r(i)$, is 0, while at least one of the time domain flags, $F_f(i, k)$ is 1. Since no broadband shock is present, $r(i)$ is set to 1 and only the subband gains are applied. The effect is to attenuate only the bands in shock. This creates a smooth transition between the shock and the moment when the shock disappears.

During normal speech, no shock is detected in the time and frequency domains, $F_r(i)=0$ and $F_f(i, k)=0$. Since one of the objectives of this algorithm is to not distort the non-shock signals, the final gain value is set to 0 dB by setting $r(i)=1$ and the broadband and subband gains to 0 dB.

This method has the following advantages: a) it preserves the shape of the spectrum while compressing the bands in shock b) it limits high-level broadband noise, even if not all the bands are in shock c) it compensates for the energy leakage between the adjacent bands by distributing the shock limiting over all bands, and d) the input signal is not affected when there is no shock.

4. System implementation

The acoustic shock algorithm presented here is implemented on a DSP system designed specifically for audio signal processing [3]. The DSP consists of three processing units operating in parallel: an input-output processor, a weighted overlap-add (WOLA) filterbank coprocessor, and a 16 bit

fixed-point DSP core. Figure 4 illustrates the block diagram of

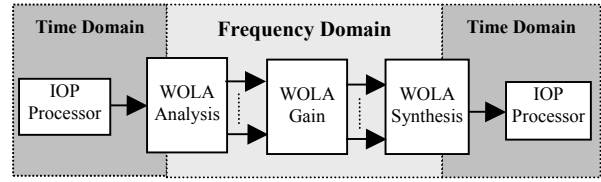


Figure 4: Signal flow path in the DSP system

the DSP and the signal path in the system.

The DSP system operates at 6.44 MHz and uses a sampling rate of 16 kHz. The size of the analysis window is $L=128$ samples. In each frame, $R=16$ samples are processed, and the analysis is done in 32 bands. The power consumption of the system is around 2.5 mW at 1.25 V.

Processing of the input signal occurs as follows. When the signal enters the DSP, the IO processor copies R digitized samples into the input FIFO and the DSP core reads these samples and determines the broadband energy, broadband gain and the broadband shock flag. The WOLA coprocessor then transforms the samples into the frequency domain. When this is done, the DSP core determines the value of the subband shock flags, calculates the subband gains, and evaluates the final gains. The WOLA coprocessor applies these gains to the subbands, performs the transformation back into the time domain and stores the synthesized samples in the output FIFO, where they are collected by the IO processor every R samples.

The algorithm is designed to produce an output signal with specific acoustic characteristics, while providing low group delay, fast shock response time, and compatibility with different headsets. Due to the real time application of the algorithm, the system provides a low group delay which is measured to be 8 msec. In addition, for safety reasons, the shock response time is minimized so that the user is not exposed to any shock signal. In this algorithm, during the transient period, the broadband gain is applied as soon as the shock is detected in the time domain. Since this value is calculated proportionally to the energy level of each new shock-carrying sample, the shock is reduced even for the first few samples and as a result, no shock is observed in the output signal. Thus, it has instantaneous response time. Also, since an acoustic shock limiter has to be integrated into many different headsets, it must be suitable to a wide range of electrical and acoustical environments. The system presented here is configurable for different applications by calibrating the threshold values, averaging coefficients and gain ratios based on the required characteristics of the output. As a result, the frequency response, speech quality and the intelligibility of the output signal can be adjusted based on the application and the characteristics of any given headset.

The acoustic shock limiting algorithm has been evaluated with both tones and speech inputs. Figure 5 demonstrates the input/output response of the system with different gain calculations for a sinusoidal input signal at the center of one of the subbands. The graph is divided into three regions: a) the linear region where the input is less than the threshold and the desired behavior is unity gain b) the limiting region where the input is more than the threshold but less than the input clipping voltage and the desired behavior is a constant output

level, and c) the input clipping region where the input is higher than the input clipping voltage of the system.

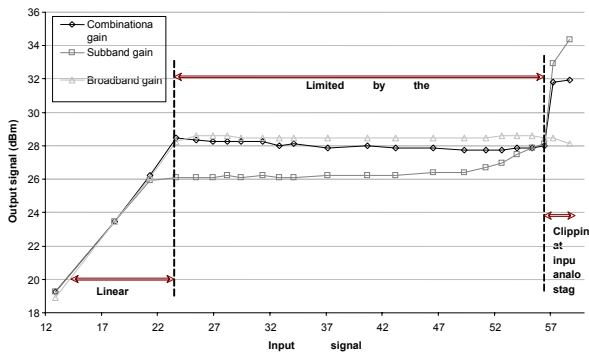


Figure 5: Input response of the system to different tones applying the broadband, narrowband, and combinational gains

As shown in Figure 5, applying broadband gain produces the required results, but as mentioned earlier it has the disadvantage of attenuating the speech as well as tonal shock signals. Conversely, the subband gain results in significant variation of output energy level in the second and third regions of Figure 5. Our algorithm mitigates this effect by calculating the combined gain described earlier, which applies a portion of the broadband gain to all the subbands. As can be seen in Figure 5, this new calculation maintains the desired flat input response in the limiting region. This behavior is observed for all the input tones with frequencies within the system’s bandwidth.

To illustrate the improvement made in the quality of speech, the output signal of the DSP system is compared to the input signal by using the log area ratio (LAR) distance [4]. In this case, the affected signal is compared to the original unperturbed speech signal on a sample per sample basis and the difference is expressed as the distance between the two signals. In this measurement, the smaller the distance is, the more similar the two signals are. Figure 6 demonstrates the improvement made in speech quality compared to the original shock carrying signal. Both signals are compared to the non-shock carrying speech signal.

Figure 7 illustrates the input and output signals with and without the transient reduction algorithm. The top graph shows the unprocessed input signal. The second graph is the output of the system when only the narrowband gain is applied to reduce the shock and the third graph shows the output of the system when the subband and broadband gains are combined. As it can be observed, the overshoot in the output signal at the beginning of the shock is controlled and reduced significantly when the broadband gain is used in parallel with the subband gain.

5. Conclusion

The acoustic shock limiting algorithm proposed in this paper limits the output signal to a constant value for different input signals, while minimizing the effects on speech quality. It also reduces the transient effect observed at the beginning of the shock in a subband-only limiting approach. The algorithm does not distort or otherwise effect signals with energy levels

lower than the desired threshold. The algorithm has been implemented on a low-power miniature DSP system that

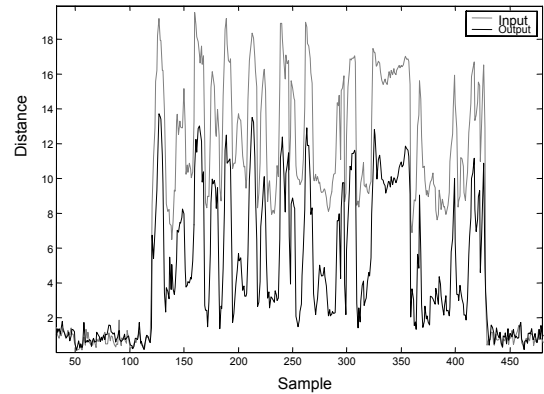


Figure 6: The LAR distance for the input and output signals, relative to the no shock carrying speech

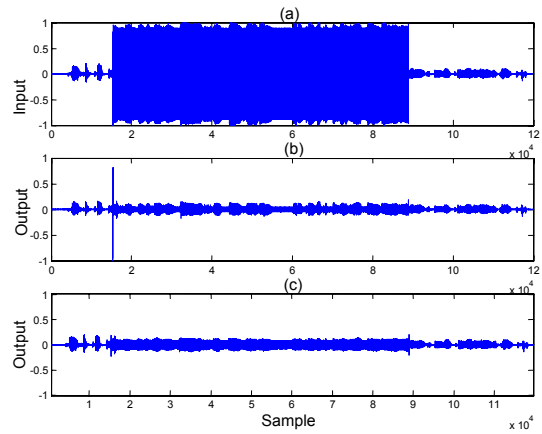


Figure 7: (a) Shock carrying signal (b) output signal without transient reduction algorithm (c) with transient reduction algorithm.

provides low group delay subband processing and it is configurable for integration into different headset.

Currently, the algorithm detects and limits shocks as they occur. Some standards, such as ITU-T P.360, also specify the necessity to monitor sound exposure over specific periods of time. Future work on the algorithm to meet these requirements will involve using adaptive thresholds based on accumulated energy.

6. References

- [1] ITU-T P.360, “Efficiency of Devices for Preventing the Occurrence of Excessive Acoustic Pressure by Telephone Receivers”, *International Telecommunication Union*, Dec. 1998.
- [2] Choy, G. and Hermann, D. “Subband-based acoustic shock limiting algorithm on a low resource DSP system”, *Proc. Eurospeech 2003*.
- [3] Schneider, T., Brennan, R, “An ultra low-power programmable DSP system for hearing aids and other audio applications”, *Proc. ICSPAT 1999*.

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