System architecture for audio signal processing in headsets

King Tam
AMI Semiconductor Canada Company
611 Kumpf Drive, Waterloo, Canada N2V 1K8
519-884-9696
king_tam@amis.com

Etienne Cornu
AMI Semiconductor Canada Company
611 Kumpf Drive, Waterloo, Canada N2V 1K8
519-884-9696
etienne_cornu@amis.com

Abstract

In this paper we describe an audio processing system whose architecture is optimized for improving the quality of the audio signal in headsets. This platform is an integral part of the architecture for advanced headsets. It provides the means to implement various signal processing algorithms in wired and wireless headsets, including system control functions. The audio processing system's architecture and signal processing algorithms are described, as well as two types of headsets in which the system is deployed. The performance of the algorithms is also discussed.

1. Introduction

Headsets must perform well in a number of different environments such as at home, at the office, in cars, and in noisy public places. Because of this wide range in operating conditions, the need arises for advanced digital signal processing to deliver the performance expected by both the local headset user and the remote user. For example, signal processing allows for the implementation of features such as echo cancellation, acoustic shock control and limiting, noise reduction, and speech intelligibility enhancement.

In this paper we present the architecture of the BelaSigna 200 audio processing system, a DSP system optimized for audio applications in ultra low-power environments such as headsets. In the audio processing system, common audio processing functions such as sample management and signal transformations to and from the frequency domain are performed using a dual-core system, ensuring that as little power as possible is consumed. The audio processing system incorporates a 16-bit DSP core and specialized weighted overlap add (WOLA) coprocessor on which signal processing algorithms can be implemented.

Section 2 of this paper describes some approaches that are used today to improve audio signal quality in headsets. Section 3 describes the architecture of the audio processing system. In section 4, the role of the audio processing system in two typical headset applications is described. Section 5 discusses the evaluation of the audio processing system. Conclusions and future work are presented in section 6.

2. Current headset technology

The headset market includes two types of headsets: consumer headsets, which are mainly targeted to applications such as mobile phones, personal computing, and gaming, and specialty headsets for industrial, medical, and security applications. Consumer headsets are typically small and use wireless technology, whereas specialty headsets have fewer physical constraints and must provide a high level of dependability.

A headset without an embedded DSP typically has a very limited set of features. If these headsets perform noise reduction on the signal heard by the user, it is usually in the form of some active noise cancellation implemented by an analog circuit, or by traditional passive means. For the signal transmitted from the headset, some form of noise reduction can usually be achieved through placement of the microphone very close to the user's mouth (thus improving signal-to-noise ratio), or by using a directional microphone specifically designed for rejecting some types of unwanted noise.

While these non-DSP solutions have their merits in some applications, they are generally not as flexible or capable compared to their digital counterparts. In addition, non-DSP solutions are more difficult to customize to a particular user, and provide limited means to change or upgrade functionality. In some cases, digital signal processing makes possible some functionality which was previously impossible. For example, the cancellation of acoustic echo (caused by the leakage of the sound from the headset's speaker into the microphone) through adaptive estimation and subtraction of the echo signal makes it possible to design headsets which are less constrained in terms of their mechanical design, thus enabling smaller, more portable, and more aesthetically appealing headsets.

In wireless headsets that use Bluetooth technology, the audio is typically handled by a relatively simple audio codec chip that only performs analog-to-digital (A/D) and digital-to-analog conversion (D/A). Typically, despite obvious technical advantages (e.g., correcting an echo problem at the source), there is little or no audio signal processing performed in the headset. A DSP can serve as a replacement for the codec in these headsets, connecting to the Bluetooth chip via the PCM audio interface. It can also process both the incoming and outgoing signals with few changes in the configuration and functionality of the Bluetooth chip.

The trend towards advanced signal processing in headsets has been addressed in several Bluetooth wireless chips that contain an integrated, general-purpose programmable DSP core. While this is a step in the right direction toward more advanced signal processing in headsets, the memory and computational overhead associated with some common audio processing functions can, in this type of DSP, significantly reduce the type of algorithms that can be deployed.

3. The audio processing system in headsets

The audio processing system described in this paper is designed for ultra-low power applications, and as such it is particularly suitable to applications in which there is no outside source of power such as a wall outlet or a large battery. The digital part of the audio processing system consists of three major components: a weighted overlap-add (WOLA) filterbank coprocessor [1, 2], a 16-bit fixed-point DSP core, and an input-output processor (IOP) that manages incoming and outgoing audio samples in a set of FIFOs. These three components operate in parallel and communicate through shared memory. The parallel operation of these components allows for the implementation of complex signal processing algorithms with low system clock rates, low resource usage, and low power consumption. Typical power consumption of the audio processing system is under 7 mW for the applications described in this paper. The system is particularly efficient for subband processing in the frequency domain: the configurable WOLA coprocessor efficiently splits the fullband input signals into subbands, leaving the core free to perform other algorithm calculations.

Two types of headsets are considered. The first type is where the audio processing system is directly connected to an RF chip (such as Bluetooth) using serial and parallel digital interfaces. The second type is where the audio processing system's interfaces are all analog and the headset is connected to the analog input/output of a two-way UHF or VHF radio often used in industrial applications. In both cases the audio processing system processes the signal in two directions: from an external source to the headset's loudspeaker (usually called the receive direction or simply RX), and from the headset's microphone to an external destination (usually called the transmit direction or simply TX).

The following sections describe the features of the audio processing system and how they are used within wired and wireless headsets.

a. Audio signal interfaces

In Bluetooth chips, the transmission of the audio signal to the speaker and from the microphone occurs through the PCM interface. The input received by the audio processing system through the PCM interface is formatted as either uncompressed 16-bit linear-format samples, or blocks of compressed data that must be decoded by the audio processing system. The uncompressed data is convenient because it can be processed directly without having to pass it through a decoding stage; compressed data incurs the cost of decoding but allows for a higher bandwidth audio signal. For a typical telecom application, 16-bit samples at a sampling rate of 8 kHz are commonly used. However, applications that require higher audio fidelity need higher bandwidth and therefore receive compressed data over the PCM interface.

The analog input and output stages of the audio processing system contain on-chip A/D and D/A converters, and are designed for connection of certain types of microphones and speakers without external components for amplification. This system-on-a-chip approach, with on-board mixed-signal blocks, has the advantage of reducing cost because of reduced component count, and generally makes the design simpler.

The flow of the audio samples is managed by a dedicated hardware unit called the Input/Output Processor (IOP). The IOP coordinates the data flow between FIFO memory and the A/D and D/A converters without intervention from the DSP core. The DSP core does not begin processing of the samples until a specific number of samples have been stored in the FIFO. The IOP raises a periodic interrupt to signal the DSP core that a new block of samples has been written to the FIFO, and that a new block of samples must be ready in the FIFO to be read for output.

b. Signal processing

A diagram of the generalized signal flow for a wireless headset configuration is shown in Figure 1. The signal processing algorithms take advantage of the system's WOLA filterbank coprocessor to perform the forward and inverse transformation between the time domain and the time-frequency domain. The filterbank's analysis efficiently splits the incoming time-domain signal into a number of subband signals which each correspond to a particular frequency region. These subband signals are processed either individually or in groups and are subsequently passed to the synthesis filterbank to be transformed back to a time domain signal for output.

The filterbank coprocessor can be easily reconfigured to realize different configurations of the WOLA filterbank. It has an architecture that is optimized for the operations that are performed regularly, such as multiply-accumulate and Fourier transforms [2]. The hardware architecture is also

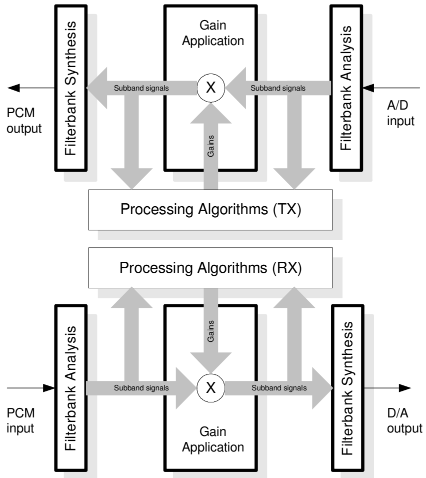

Figure 1. Framework for processing

optimized for the processing of two audio streams in parallel.

The programmable DSP core implements the algorithm block, as shown in Fig. 1, while the coprocessor handles the filterbank operations and gain application stage (i.e., scaling of subband signals). This is the basic framework for a variety of algorithms.

c. Algorithm integration

The system architecture provides performance benefits for many signal processing blocks because of the frequency-selective nature of the algorithms. Some examples of the signal processing algorithms are blocks for noise reduction, echo cancellation, musical tone generation, acoustic level limiting and dynamic range compression [3, 4, 5, 6]. The programmable nature of the system allows for the integration of several of these algorithms into one system, provided that the total memory and processing time constraints are not exceeded.

d. Group delay

The digital processing performed between the inputs and outputs is dependent on the requirements of the particular application. However, an important constraint in many applications is that the delay through the system must be low enough that the user does not perceive it as distracting or intolerable. This is particularly important when synchronization is expected between audio and video. Even in an audio-only telephone conversation, too much delay which can cause echo leads to rapid user fatigue. If acoustic echo is present on the headset, a long delay through the system makes this echo noticeable at the other end of the telephone link. The audio processing system's architecture is particularly well-suited for processing audio with little impact in terms of delay because of the oversampled nature of the WOLA filterbank processing [2]. In a typical configuration, the signal delay through the system is less than 10 milliseconds at a 16 kHz sample rate.

e. Other interfaces

The audio processing system's architecture is capable of performing other functions in addition to audio signal processing. The auxiliary low-speed A/D converters can be used to monitor the battery level or the voltage across a volume-control potentiometer. The general-purpose digital inputs can be connected to buttons to allow the user to control some system settings such as algorithm parameters. The digital outputs can be programmed to control an external LCD unit used for displaying status messages to the user. Additionally, the UART interface can be used to receive control commands from the Bluetooth chip.

4. Typical headset configurations

In this section we present two typical headset configurations in which the audio processing system is integrated.

a. Bluetooth headset

A typical advanced Bluetooth headset may provide two operating modes: telecom and audio streaming. Telecom mode is used when the headset is paired with a mobile phone or similar device for a two-way conversation. When the system is switched to audio streaming mode, the headset receives compressed audio data on the Bluetooth link and outputs the decoded audio. This mode is intended for listening only, and allows for high-bandwidth audio using the Sub-band Codec (SBC) from the Bluetooth Advanced Audio Distribution Profile (A2DP) [7]. These two modes need to co-exist because switching from one mode to another must occur seamlessly and as quickly as possible.

Figure 2 shows the signal flow through the audio processing system in telecom mode. The noise reduction blocks for receive (RX, audio stream coming to the headset user) and transmit (TX, audio stream coming from the headset user) signals analyze their respective input signals and calculate a set of time-varying scale factors to be applied by the coprocessor's gain application function. These time-varying scale factors attenuate the subbands based on the signal-to-noise ratio in each frequency region. The Audibility Enhancement block improves audibility of the incoming signal across a wide variety of conditions ranging from a quiet room to a noisy car. It makes a continual automatic adjustment of the level of the incoming signal based on a frequency-dependent analysis of the noise conditions in the user's local environment.

In audio streaming mode, only the SBC Decoder and Tone Generator block are active (Figure 3). The decoder supports

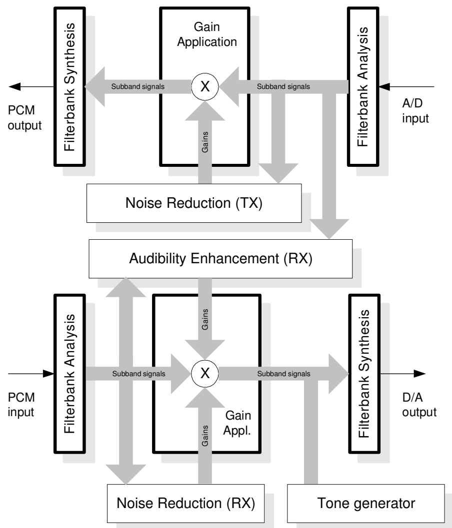

Figure 2. Processing in Telecom Mode

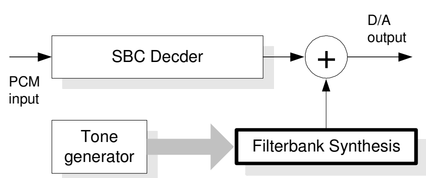

Figure 3. Processing in Audio Streaming Mode

a maximum sampling rate of 48 kHz for audio quality comparable to that of a compact disc [8].

In addition to using the PCM interface, the Bluetooth chip and the audio processing system also communicate through a UART interface. At system start-up, the Bluetooth chip boots the audio processing system by sending the application code over the UART interface, thus saving the cost of an EEPROM dedicated to the audio processing system. The interface is then used for control commands from the Bluetooth chip to the audio processing system, which allows for mode selection (i.e., telecom or audio streaming), volume control, activation of tone generation, pre-amplifier configuration, and control of the signal processing algorithms.

In order to save power and extend battery life, the system enters a low-power sleep mode when it does not need to operate actively. For example, when the headset is in telecom mode but is not connected to a mobile phone or similar device, the audio processing system does not receive data over the PCM interface and can therefore enter a sleep mode. In this mode, the DSP core disables or reconfigures peripheral units such as the IOP and A/D and D/A converters to minimize power consumption. The audio processing system returns to normal operating mode when there is activity on the PCM interface.

The Bluetooth chip acts as master device in the system and is therefore responsible for handling the user's button presses and for determining if the system should be in telecom or audio streaming mode. For example, when the system is in audio streaming mode and an incoming phone call is detected, the commands are sent to the audio processing system to switch to telecom mode and start the tone generation module for any user signaling required. After the call has been accepted, the Bluetooth chip may detect a button press by the user to adjust the volume, and send the corresponding command to the audio processing system so it can perform the amplitude adjustment within its signal path.

b. Industrial headset

In a typical industrial headset application, the audio processing system is inserted between a VHF or UHF two-way radio and the headset's microphone and speakers. Even with the integrated audio processing system, the headset can be small enough to fit into the user's ear.

As an example, an industrial headset may be configured to take input from the radio and from two microphones, as seen in the overview diagram in Figure 4. Microphone A is used to transmit the user's voice over the radio, while microphone B is used to allow the user to hear the local environment.

The noise reduction and volume control blocks are similar to the ones seen in the Bluetooth headset example. A limiter block may also be integrated to protect the user from any potentially damaging high-level impulsive sounds. This block is needed so that the headset performs well in the industrial and military environments for which it was designed.

A button connected to one of the audio processing system's digital inputs is used to adjust the digital volume control applied to the signal from microphone B. The purpose of this control is to allow the user to adjust the amplification level of the local signal.

5. Algorithm performance

As described in this paper, signal processing in headsets may be used for at least three major purposes: improving audio signal quality at each end of the communication link for better user experience, protecting the headset user against potentially damaging signals, and allowing CD quality audio to be played on a headset despite the limited bandwidth of the wireless link. In many cases, such as audibility enhancement, algorithm performance is mostly subjective. Field tests have shown that enhancing audibility as a function of the ambient noise can significantly increase intelligibility in varying noise conditions. Evaluating the quality of signal received in audio streaming mode is also subjective, but certain conditions must be satisfied by the codec in order to be certified for use in the Bluetooth environment. For example, the Bluetooth standard specifies the acceptable margin error for signals that are encoded and then decoded by the codecs. The streaming audio algorithm described in the previous section satisfies these requirements. In the case of the limiter, the algorithm is judged by its ability to limit the level of the input to a certain loudness and by the speed in which it can react to impulsive signals. One type of algorithm for which objective results can be measured is noise reduction.

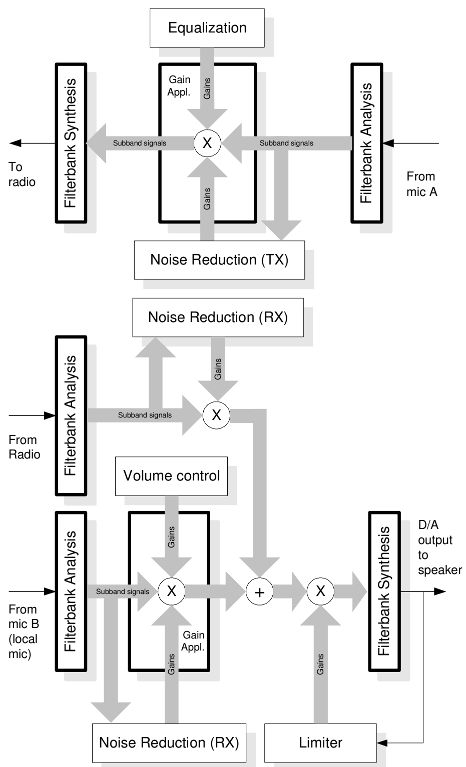

Figure 4. Industrial headset

Two noise reduction algorithms are considered. The algorithm referred to as NR1 is a low-resource and low-complexity algorithm that is based on minima and maxima tracking. It is used for RX processing on the Bluetooth headset, and for the processing of the microphone B signal on the industrial headset. The algorithm referred to as NR2 uses a noise reduction method similar to the method described by Ephraim and Malah [9]. It provides better noise reduction performance than NR1. NR2 is used for TX processing on the Bluetooth headset, and for the processing of both the microphone A signal and incoming radio signal on the industrial headset.

Measurements were taken using a typical headset and different types of noise and speech. The noise (white and convention) and speech levels were adjusted in order to test the performance of the noise reduction across a wide range of conditions. All measurements were performed in an acoustically controlled environment, placing the headset on a head and torso simulator and processing the signal in real-time on the audio processing system.

Table 1 summarizes the improvement in signal-to-noise ratio (SNR) provided by the two different noise reduction algorithms. The SNR improvement was calculated using the energy level of non-speech segments in the signal before and after noise reduction is applied. Results show that the algorithms provide a clear benefit which allows headsets to perform well in noisy environments. In white noise and convention noise, the NR1 algorithm provided at least 7.9 dB of noise reduction. The NR2 algorithm provided at least 11.0 dB of noise reduction in tests using similar test conditions.

Table 1. Noise Reduction Performance

Algorithm NR1		
Noise type	**SNR (dB)**	**Improvement (SNR, in dB)**
white	0	8.0
white	6	9.1
white	12	8.9
convention	0	7.9
convention	6	8.7
convention	12	9.5
Algorithm NR2		
Noise type	**SNR (dB)**	**Improvement (SNR, in dB)**
white	0	15.8
white	6	14.9
white	12	14.4
convention	0	16.1
convention	6	16.2
convention	12	11.0

6. Conclusions and Future Work

The BelaSigna DSP audio processing system is an integral part of the architecture for advanced headsets. It provides the means to implement a number of signal processing algorithms for wired and wireless headsets.

Two typical headset configurations were discussed. The signal processing algorithms implemented by the audio processing system, as well as the system's non-signal processing functionality, were described. Measurement of noise reduction performance indicates that SNR improvements between approximately 8 and 16 dB are achieved. Subjective testing has also shown that enhancing audibility as a function of the ambient noise can significantly increase intelligibility in varying noise conditions.

Future directions may include the design of the next-generation audio processing system and architecture. We expect to make many improvements based on experience gained from past headset development, possibly including improvements in word-length, enhanced flexibility, and improved PCM interfacings. We expect the outcome to be an improved audio processing system with greater flexibility and capability in terms of signal processing and interfacing with external modules.

References

[1] R. L. Brennan and T. Schneider, "Filterbank structure and method for filtering and separating an information signal into different bands, particularly for audio signals in hearing aids", U.S. Patent 6,236,731, May 2001.

[2] A. Drollinger, et al., "Efficient Sub-band Coder Implementation for Portable Low-power Applications", *Proceedings of Int'l COST 254 Workshop*, Neuchâtel, Switzerland, 1999.

[3] K. Tam, "Frequency-Domain Additive Synthesis With An Oversampled Weighted Overlap-Add Filterbank For A Portable Low-Power MIDI Synthesizer", *Proceedings of 117th AES Convention*, San Francisco, 2004.

[4] K. Tam, H. Sheikhzadeh, and T. Schneider, "Highly oversampled subband adaptive filters for noise cancellation on a low-resource DSP system", *Proc. of Int. Conf. on Spoken Language Processing (ICSLP)*, 2002.

[5] T. Soltani, D. Hermann, E. Cornu, H. Sheikhzadeh, and R. L. Brennan, (2004) "An acoustic shock limiting algorithm using time and frequency domain speech features", *Proceedings of ICSLP*, Korea, 2004.

[6] J. Johnson, E. Cornu, G. Choy and J. Wdowiak. (2004) "Ultra Low-Power Sub-Band Acoustic Echo Cancellation for Wireless Headsets", *Proceedings of the IEEE International Conference Acoustics, Speech and Signal Processing*, 2004.

[7] *Advanced Audio Distribution Profile*, Bluetooth Audio Video Working Group. 2002.

[8] D. Hermann, R. L. Brennan, H. Sheikhzadeh and E. Cornu. "Low-power Implementation of the Bluetooth Subband Audio Codec", *Proceedings of the IEEE International Conference Acoustics, Speech and Signal Processing*, 2004.

[9] Y. Ephraim, D. Malah, "Speech Enhancement Using Optimal Non-linear Spectral Amplitude Estimation", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing,* 1983.