Reducing Latency in Digital Wireless Audio Systems

Sheersound™ White Paper

Author; Peter Mariager, CTO, RTX A/S
sales@rtx.dk

1 Scope

This white paper will introduce common challenges in modern wireless systems for "live" signals, such as microphones, in-ear monitors and gaming headsets. We will introduce you to what we believe is the best available technology on the market today, which effectively counters these challenges and provides high audio quality with the lowest possible latency.

Topics include:

- Trends in wireless microphones for professional and semi-professional use
- Wireless technologies for digital systems
- Audio compression techniques

Author; Peter Mariager, CTO, RTX A/S
sales@rtx.dk

2 Introduction: From Analog to Digital

Wireless transfer of audio signals, such as voice or music, is in widespread use. Special requirements originate from this application of wireless transfer of “live” signals, where the latency (delay) must be kept low, and significantly lower than other wireless application areas such as telephony and intercom systems for communication only.

These *low-latency* systems range from highly professional stage microphones, in-ear monitors (IEM) and music instrument pick-ups over pro-sumer garage band systems to karaoke and gaming microphones and headsets.

For many years, analog technology has made use of portions of the UHF band, specifically 500 – 800 MHz, to provide excellent audio quality and low latency.

However, these UHF bands are gradually being allocated for other uses, initially for re-allocation for digital TV roll-out and latest for broadband cellular access. These frequencies are gradually being auctioned off to mobile operators for the continuing roll-out of 4G and later 5G services. This has and is forcing manufacturers of wireless microphones to develop new systems capable of providing the needed wireless performance in other parts of the wireless spectrum.

2.1 Why Digital?

As a part of this new development, digital microphone systems are being developed for several reasons in favor of their analog counterparts: The spectral efficiency can be higher (more channels per MHz of used bandwidth), coding can be applied to increase performance and reduce risk of drop-outs and finally encryption can be applied to make it impossible to eavesdrop on the microphone signal.
These digital systems have the potential to be superior to their analog counterparts in most performance aspects, except for one: The critical latency requirement.

Author; Peter Mariager, CTO, RTX A/S
sales@rtx.dk

3 The Latency Challenge

The use of in-ear monitors give performers more mobility as well as less acoustical interference from other sources, such as singers or instruments. Professional wireless stage microphones, instrument pick-ups and in-ear monitoring systems aim at a transmitter-to-receiver latency of less than 3 milliseconds (ms). Such a low latency allows for a much tighter control of the full system round-trip time from e.g. a performer's microphone to the mixing console and back to an in-ear monitor. The (result of which) is a system delay that is much more demanding than e.g. wedge monitoring / stage monitors due to the direct feedback and lack of acoustical delay by the IEM. Several studies have been performed to determine the effect of latency (originating with the Hass effect). Most systems today aim to keep the round-trip latency well under 10 ms.

For gaming headsets, the requirements are slightly less demanding. Good gaming headsets provide under 15 ms of transmitter-to-receiver latency, but less than 10 ms is desirable. Hard-core computer game players require very low delays of game audio to stay competitive and fully immersed in the game play. Virtual reality gaming further adds to this requirement.

The figure below illustrates a simplified architecture for a typical digital wireless microphone and receiver:

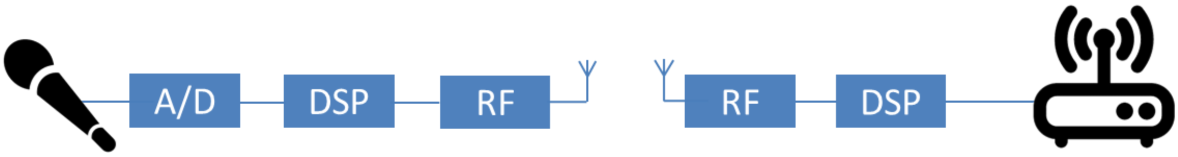

Figure 1. A simplified architecture for a typical digital wireless microphone transmitter and receiver

A digital wireless system will add additional latency over the analog counterpart in three parts of the signal chain:

The A/D Converter: The Analog-to-Digital (A/D) converter will always introduce some latency depending on the technology used. This latency is considered out of scope for this white paper, as it will either be constant or similar for all types of digital microphones.

The RF: Implementing an RF interface in a digital microphone will always introduce latency. It can originate from the lower layer protocols, where audio data is packed into frames and transmitted in a Time-Division Multiple Access (TDMA) fashion. TDMA systems have the advantage of sharing the spectrum with other systems and hence increasing the capacity. It also allows for retransmissions for improved interference immunity as well as preserving microphone battery by shutting down the transmitter when not active.
Even though the latency is kept low by not using a TDMA frame format, additional latency will be added due to e.g. Forward Error Correction (FEC) schemes.

The latency typically originating from the RF portion of the transmitter and receiver can be from 1 ms and higher, depending on the required capacity and robustness.

The DSP: Even though advanced radios today can effectively transfer huge amounts of data (e.g. LTE and Wi-Fi), it is desirable to reduce the number of bits/second (bps) that is transferred. Reducing the *bps* requirement leads to cheaper and less power-hungry RF parts (resulting in longer battery life), less occupied RF spectrum (more capacity per area) and more immunity to interference.

Consider the raw bps requirement from an *uncompressed* audio signal sampled at 48 kHz with 24 bits/sample:
48,000 samples / sec x 24 bits / sample = 1,152,000 bits/sec = 1,152 Mbps

Transferring such an uncompressed audio signal over a wireless link will easily occupy 2 MHz of RF spectrum which is 10x more than allowed by regulatory bodies such as EN 300 422. Furthermore, the number of active microphones in a certain area would be greatly reduced. It is therefore desirable to compress the digital audio signal using some form of *audio coding*, which will be described in the next chapter.

4 Introducing Audio Coding

As mentioned above, it is desirable or even necessary to reduce the number of bits/sec. This is accomplished using audio coding, which is a well-known technique used in many application areas from cell phone communication to music streaming.

Audio coding algorithms (denoted *Coders* in the remainder of the document) can be divided into two classes:

- **Lossless audio coders:** These algorithms preserve a bit-exact version of the original audio, meaning any effects of the coding / decoding are inaudible. Examples of lossless coders are FLAC and ALAC, which is analogous to a *zip* compression for audio signals. The disadvantage is that the compression rate is often low, a typical FLAC compression reduces the signal to 50-60% of the original rate.
- **Lossy audio coders:** Lossy audio coders compress the audio signal through various compression techniques. Typical speech coders use models of the human vocal tract to predict information in the signal. These coders often compress with very high rates and are used for less performance centric applications such as telephony. Other, more generic coders do not rely on information of the audio signal and can compress any type of signal, such as voice, music and musical instruments. These coders, called *perceptual* coders, rely on knowledge of psychoacoustic properties (the way the human perceive sound) to effectively remove portions of the signal that can't be heard anyway. An example of such a coder is the well-known MP3 coder.

5 The SHEERSOUND™ Audio Coder

RTX has developed solutions for wireless audio for 15+ years, and has worked with a broad range of audio coders. Although coders for low latency exist (described later), RTX felt that there was a need for a better solution and we started developing our SHEERSOUND™ codec in 2014.

During the development, we had the following design criteria:

1. **Ability to compress every audio signal:**
   SHEERSOUND™ is a lossy codec that is based on psychoacoustic properties to effectively compress the audio signal. It can therefore be used in any application.
2. **Lowest possible latency:**
   SHEERSOUND™ belongs to a class of coders known as non-block-based,

Author; Peter Mariager, CTO, RTX A/S
sales@rtx.dk

which effectively gives it the lowest possible algorithmic delay of only one sample, which is 20 us at 48 KHz sample rate.

3. **Excellent audio quality:**
   Throughout the development, focus was put on excellent audio quality over compression rate. SHEERSOUND™ has been developed using both psychoacoustic models as well as extensive listening tests for evaluation and tuning of performance. The result today is a coder that has excellent performance with minimal or even inaudible effect on the signal it is compressing.
4. **Flexibility:**
   SHEERSOUND™ can be configured for a wide range of sample and coding rates, which makes it very flexible and can be adapted to many wireless protocols.

SHEERSOUND™ can be used in any application that demands the best possible audio quality and lowest possible latency.

5.1 Technical Characteristics and Performance Figures

SHEERSOUND™:

- Is non-block based, and operates in a combination of both time and frequency domain, providing virtually zero-delay (1 sample) operation.
- Uses psychoacoustic properties of the human hearing to effectively conceal potential adverse effects of compression
- Has built-in transmission channel error resilience, making SHEERSOUND™ an excellent fit for wireless applications

Property	Specification
Sampling Frequencies	Configurable, 16 - 96 kHz
Input/Output Word lengths	16, 24 and 32 bit
Bit rate	Configurable compression rates, 2 – 4 bits/sample (typical values 160-210 kbit/s*)
Algorithmic Latency	1 sample (20 us*)
Frequency Response	flat from 10Hz to 22kHz*
Dynamic Range (testsignal 1kHz -140dBFS)	>128dB**
S/N (test signal 1kHz -1dBFS)	>89dB**
THD @ 1kHz -1dBFS	Better than -123dB**

*Specified at 48 kHz sample rate, **Specified at 4 bits per sample

Figure 2. Important technical parameters of SHEERSOUND™

Author; Peter Mariager, CTO, RTX A/S
sales@rtx.dk

As we mentioned above there are many other audio coders available. The following section will explore how SHEERSOUND™ compares with some of these:

5.2 Comparison to Other Audio Coders

Advancements in the speed of digital signal processors have enabled more and more complex audio coders during the past years. Especially in the field of audio for music streaming services, we have witnessed an evolution of better and better audio quality at lower and lower bitrates. However, music streaming has no latency requirement, which allows for a completely different type of compression. However, MPEG-1/2 Layer III (MP3) is included for comparison.

In this white paper, we therefore focus on typical coders used for real-time applications as explained above:

Organization	Coder Name
ISO / IEC	MPEG-1/2 Layer III (MP3) AAC-LD
ITU-T	G.711, G.722, G.726, G.729, G.729.1
IETF	OPUS (RFC6716)
3GPP	AMR-WB
Qualcomm	AptX, AptX Live
Broadcom	Broadvoice BV16, BV32

Author; Peter Mariager, CTO, RTX A/S
sales@rtx.dk

A comparison of SHEERSOUND™ with these codecs can be illustrated in the figure below:

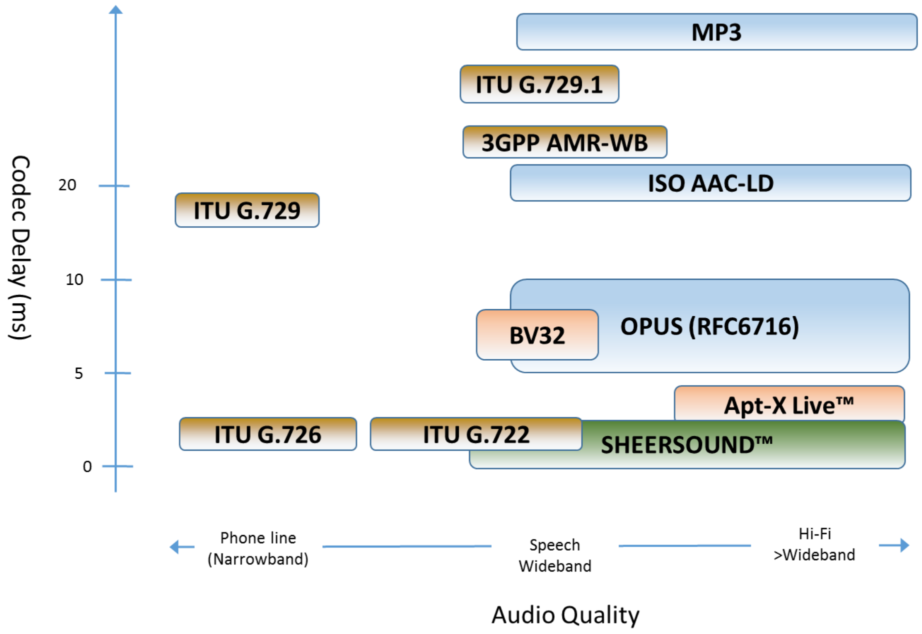

Figure 3. Comparison of the latency and audio quality of several well-known audio coders - bit-rate and complexity not considered.

Only a few other codecs can achieve the <1 ms delay offered by SHEERSOUND™, as they are also non-block based. These are the ADPCM based G.726 and G.722 which have been used for many years for telephony applications. However, these coders cannot fulfill the requirements for Hi-Fi use, such as professional stage microphones. The Qualcomm AptX Live codec also uses ADPCM principles combined with sub-band coding which bring the computational latency up to 1.8 ms which leaves very little latency room for the inevitable delay introduced by the wireless link.

When comparing audio coders, 3 contradicting properties are seen. These are: latency, audio quality and bit-rate / compression rate.

- It is easy to develop a coder with very low latency and excellent audio quality, but the challenge is to also obtain a suitable compression rate (i.e. the number of bits used on the wireless link).
- One can also develop a coder with high compression rate and low latency, but the audio quality would suffer - examples are coders used for telephony, such as G.72x.

Author; Peter Mariager, CTO, RTX A/S
sales@rtx.dk

- Finally, coders with a high compression rate and good audio quality also exists, but the latency is high. An example of such an audio coder is the well-known MP3.

This can be illustrated in the figure below:

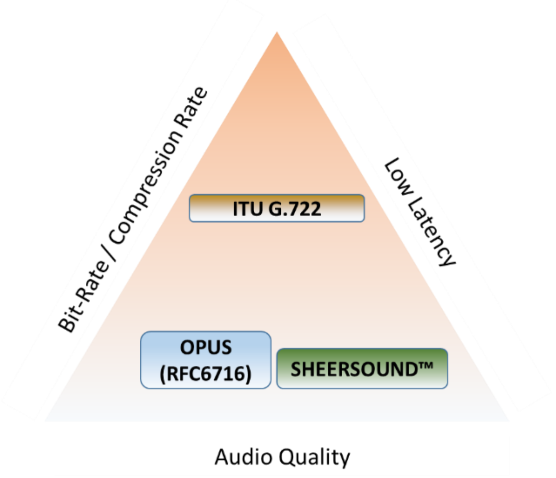

Figure 4. The trade-offs between high compression rate, low latency and good audio quality.

The SHEERSOUND™ coder has been optimized to provide the lowest possible delay and best possible audio quality. We obtain this with a typical compression rate of between 6:1 and 10:1 for a 24-bit input signal. This is equal to 2.5 – 4 bits per sample, and corresponds to 120 – 192 kbps for a 48kHz input signal.

6 Implementation Options

SHEERSOUND™ is available on a variety of off-the-shelf platforms today, such as:

- Tensilica LX4 / HiFi-3 (Dialog DA14195/DA14495)
- Analog Devices Blackfin ADSP-BF592
- Texas Instrument C54
- x86
- Fixed / floating point C

To evaluate the codec, we can offer to transcode your audio files or let you evaluate our analog-in/analog-out SHEERSOUND™ evaluation kit.

Author; Peter Mariager, CTO, RTX A/S
sales@rtx.dk

7 Conclusion and Further Information

RTX identified the need for an audio compression coder that would enable the application of digital wireless technology while still meeting the market requirement for excellent audio quality and end-to-end delays of 2 – 10 ms in performance critical applications. During the development of SHEERSOUND™, we focused on flexibility, lowest possible latency and highest possible audio quality.

We believe the SHEERSOUND™ codec offers the best available solution for performance critical audio applications today.

For more information, please contact sales@rtx.dk

Author; Peter Mariager, CTO, RTX A/S
sales@rtx.dk