Fraunhofer IIS

FRAUNHOFER INSTITUTE FOR INTEGRATED CIRCUITS IIS

WHITE PAPER

Fraunhofer IIS Audio Communication Engine – Raising the Bar in Communication Quality

INTRODUCTION

Developed to enable highest quality communications in a wide variety of professional and consumer applications, the Fraunhofer IIS Audio Communication Engine is a fully integrated solution that combines all relevant components of an audio communication system. A high-quality/low-delay audio codec, robust acoustic echo control and a low-delay IP-streaming system are the primary features of a solution which sets a new standard for quality in audio communications over IP.

ABOUT THIS PAPER

In the pages that follow, this white paper will introduce the basic principles of the Audio Communication Engine, explaining its core components, applications and benefits. In chapter 3 and beyond, this paper will explore in greater depth the specific technologies that make up the Fraunhofer Audio Communication Engine.

1. AUDIO COMMUNICATION ENGINE: AN INTRODUCTION

1.1 CORE COMPONENTS

Three main components enable the Audio Communication Engine to provide a comprehensive, high quality communication experience (see figure 1):

1. Low Delay Audio Codec: MPEG Enhanced Low Delay AAC (AAC-ELD) – the successor to Low Delay AAC (AAC-LD) – is the state-of-the-art MPEG audio codec for highest speech and audio quality with very low coding delay.

2. Acoustic Echo Control: This innovative Fraunhofer IIS technology guarantees reliable and consistent prevention of echoes that can otherwise render conversation frustrating, inefficient and prone to disruption.

3. Low-delay IP-streaming: Based on advanced error concealment and adaptive playout algorithms, this component provides robust audio quality at low delay under highly variable network conditions such as packet loss and jitter.

Collectively, these components – which are also available separately – satisfy the highest performance requirements for IP communication systems.

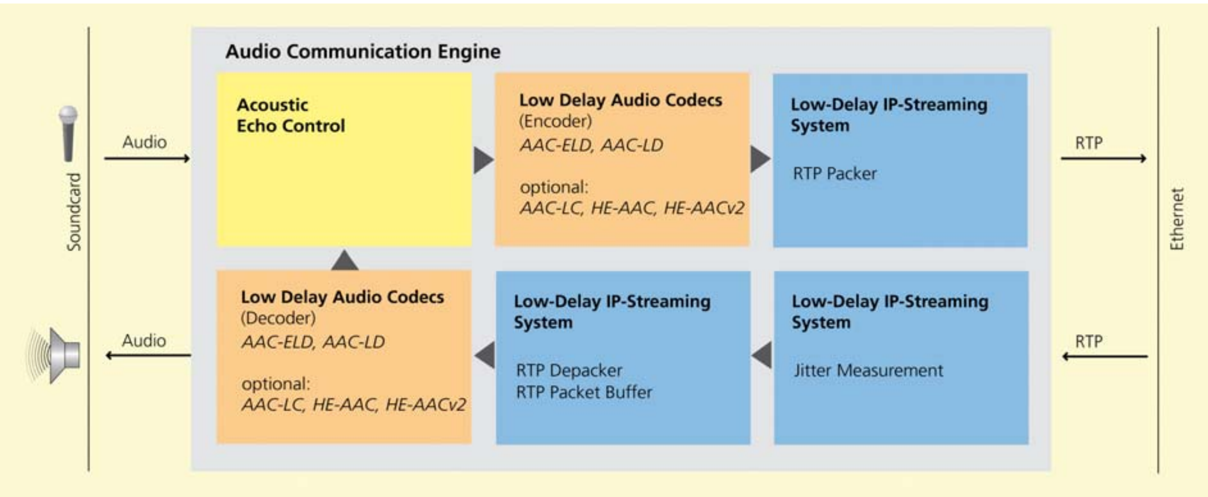

Figure 1: Key components of the Audio Communication Engine

1.2 APPLICATIONS

The Audio Communication Engine has been developed not only to satisfy current requirements, but also to offer the flexibility required to adapt to those that will emerge in the future. Here are some key examples:

Video- and Teleconferencing: Today's users of video- and teleconference systems expect superb audio quality with no compromises in usability and utility. By employing Fraunhofer's Audio Communication Engine, manufacturers of video- and teleconferencing systems, devices and software can deliver CD-like quality with hands-free, full-duplex ease-of-use. Using stereo or multi-channel set-ups, conference participants can be localized, which leads to an even more natural and appealing communication experience. AAC-LD audio codecs have been deployed in many telephone and videoconference systems, including those from leading suppliers such as Cisco and Tandberg.

Telepresence at Home: Globalisation is having a profound effect on individuals' business and private lives. The changing patterns of work and life are leading to a greater need for high-quality, easy-to-use personal communication solutions that allow people to keep in touch with friends, family and colleagues. The Fraunhofer Audio Communication Engine can provide the basis for comprehensive solutions integrated into broadband connected devices, including PCs, TVs, set-top boxes and mobile phones. Staying in touch should be as natural as having a conversation with all participants present in the same room. The Audio Communication Engine allows service providers and hardware manufacturers to make this a reality.

(Mobile) Voice over IP Services: The Audio Communication Engine enables operator of 3G and especially 4G (LTE and WiMax) networks, cable operators, Internet service providers and telecommunications companies to offer highly differentiated, CD-quality phone services. Using the latest high-quality perceptual MPEG audio codecs, voice, sounds and music are accurately reproduced, allowing rich, natural and productive communication far superior to anything used in phone services today. This - coupled with low bit-rates, good error concealment characteristics and high service quality even under adverse network conditions - allows the roll-out of advanced Voice over IP (VoIP) services with minimal effort, be it for cable VoIP, FTTH, IPTV, mobile or internet communications.

Broadcast Contribution: Live reports from remote locations have historically relied on QoS-proven ISDN network connections. ISDN is being replaced, however, by internet connections and, even less reliably, mobile and stationary wireless links. This represents a major challenge for broadcast equipment manufacturers. The Audio Communication Engine provides the necessary robustness even under challenging conditions, and is therefore an ideal component of broadcast contribution products. For example, Telos has implemented components of the Audio Communication Engine in its acclaimed Zephyr/IP codec.

1.3 COMMERCIAL BENEFITS

Communication products using the Fraunhofer Audio Communication Engine provide a natural, convenient and efficient experience that is comparable to talking to someone in the same room.

Audio Communication Engine's integration of all the major components required for high-quality communication systems in one single solution means that it can assist manufacturers to speed up their R&D processes and bring new products to market more quickly. With the pace of communication technology development continuing to accelerate, the Audio Communication Engine can give manufacturers a commercial edge in an increasingly competitive global marketplace.

Additionally, the close integration of components ensures the best possible performance and minimal implementation effort, underlining the Audio Communication Engine's credentials as a convenient route to effective and reliable audio communication.

2. LOW DELAY AUDIO CODEC

Enhanced Low Delay AAC (AAC-ELD) is the latest addition to the MPEG Advanced Audio Coding family. This growing range of technologies is designed to satisfy the most demanding requirements in high-quality communication applications.

High sound quality, low delay and a flexible range of bit- and sampling-rates are among the primary characteristics of AAC-ELD, a further improvement of the versatile and widely deployed AAC-LD (also included by Fraunhofer IIS as a component of the Audio Communication Engine).

In essence, AAC-ELD further boosts the bit-rate efficiency of its forerunner by combining it with Spectral Band Replication (SBR) technology while maintaining a low algorithmic delay (as low as 15 ms at 64 kbit/s and 32 ms at 24 kbit/s). SBR is widely deployed in High-Efficiency AAC (HE-AAC).

The following section looks at the way in which this super wide band codec delivers all the requirements of a modern communication codec, and describes some of the underlying technologies.

Low delay – essential for productive, real-time communication

In a face-to-face conversation, delays in getting a response are usually and intuitively interpreted in a variety of ways (hesitation, needing time to think, not wanting to give an answer). However, if the other party actually responds immediately but a delay is introduced by technical shortcomings, misunderstandings can develop very quickly and conversation can become awkward and frustrating. It is, therefore, very important to keep these delays, also called latencies, to an absolute minimum; 150 to 200 ms are considered upper limits for a natural flow. AAC-ELD combined with the other components of the Audio Communication Engine always delivers the lowest possible latency. AAC-ELD itself contributes only 15 to 32 ms.

The low delay capability of the codec was demonstrated during the formal MPEG Verification Test[1], which preceded the formal MPEG standard approval of AAC-ELD in July 2008. In this test, the performance of AAC-ELD was compared to that of the latest available super wide band ITU codec, G.722.1-C. Among other results, the test revealed that AAC-ELD has a nearly 25 percent lower algorithmic delay for speech when compared to the ITU codec at the same bit-rate (32 kbit/s). The assessment also highlighted AAC-ELD's capacity to bring the algorithmic delay for audio signals down to 15 ms in comparison to MPEG Low Delay AAC (see figure 2).

The diagram below – drawn from the results of the 82nd MPEG Meeting on AAC-ELD – provides further evidence of the strengths of AAC-ELD when compared to AAC-LD and G.722.1-C.

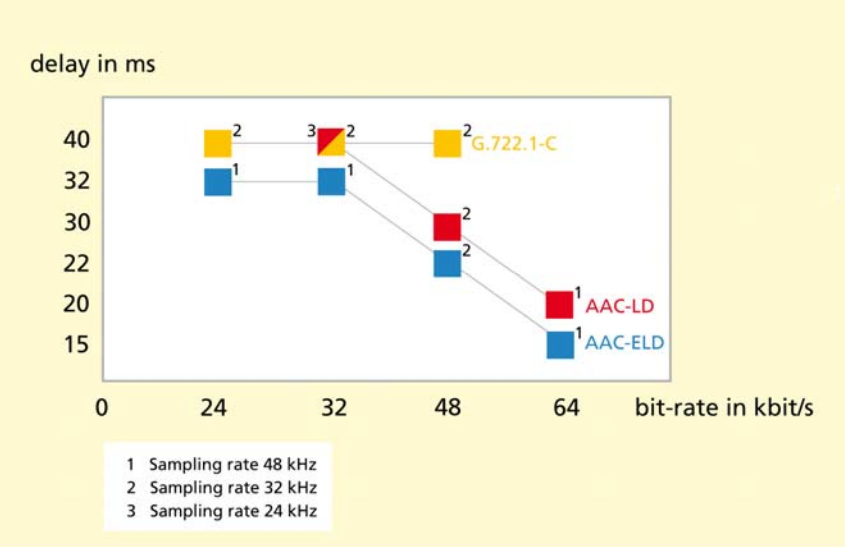

Figure 2: Comparison of algorithmic delay of super wide band audio codecs at different bit-rates

High quality audio for natural communication sound

Historically, some communication configurations have fallen short of the highest expectations because, whilst they might have provided acceptable reproduction of speech, they have been unable to deliver as 'natural' a communication experience as participants might have wished.

Unlike traditional communication codecs, a high-specification audio codec such as AAC-ELD is able to ensure both excellent speech quality and transmission of any input signal – for example, music and background sounds – without altering their fundamental characteristics. The result is a sensation of 'virtual presence' so compelling that in some test scenarios, participants have entered a room expecting to encounter a participant who was actually at a location several thousand kilometres away!

[1] MPEG Verification Tests are the official listening tests of the ISO standardization body to examine the technological progress of a new audio standard. The official set of test items includes music, speech and critical single instruments. http://www.chiariglione.org/mpeg/working_documents/mpeg-04/audio/aac-eld-vt.zip

AAC-ELD supports mono, stereo and all common multi-channel configurations. Once again, the codec's performance was assessed during an MPEG Verification Test that involved direct comparisons with ITU codec G.722.1-C. Significantly, AAC-ELD was shown to achieve comparable audio quality at 24 kbit/s to G.722.1-C at 32 kbit/s (see figure 3), while the codec also achieved "excellent" audio quality at 48 and 64 kbit/s for the critical MPEG speech and audio test signal set. It was also shown that the AAC-ELD is able to deliver better speech quality than its comparison codec at the same bit-rate (both 32 and 48 kbit/s) (see figure 4).

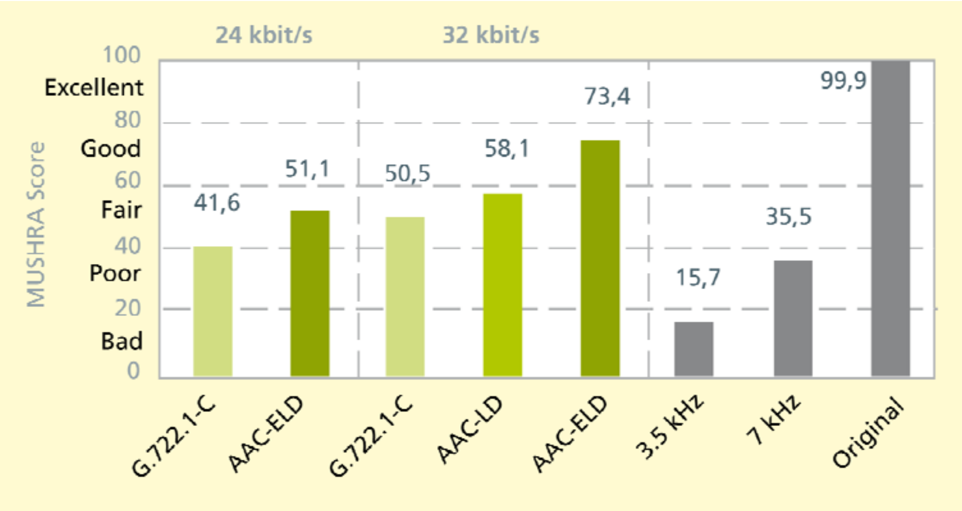

Figure 3: Audio quality of super wide band audio codecs at different bit-rates for audio signals

Finally, the comprehensive test demonstrated that AAC-ELD provides substantially better speech quality than a 7 kHz band-limited perfect anchor and, furthermore, any wide band speech codec. It was also documented that audio quality and intelligibility of well-designed speech devices increase very significantly with an audio bandwidth of more than 7 kHz.

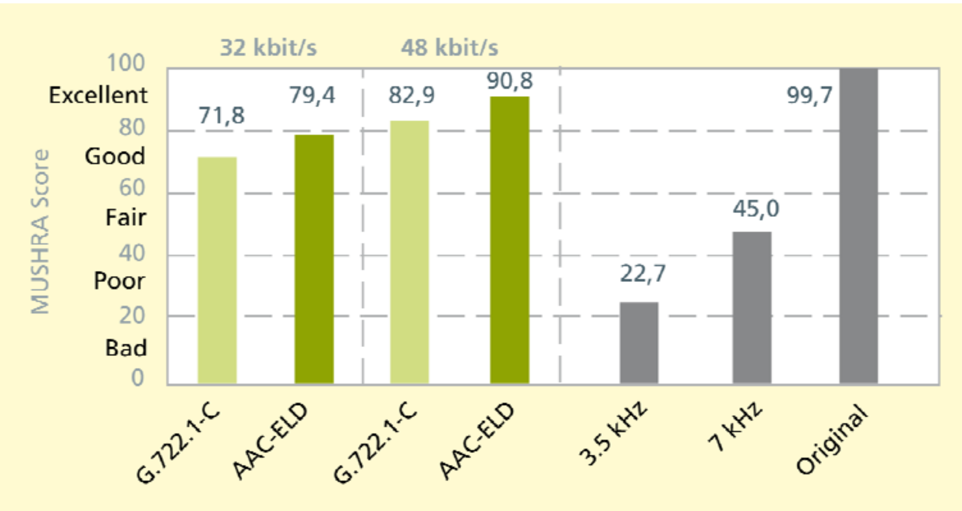

Figure 4: Audio quality of super wide band audio codecs at different bit-rates for speech signals

Enabling technologies: SBR and low delay filter bank

One of the defining characteristics of the AAC-ELD codec is its seamless incorporation of the SBR (Spectral Band Replication) tool and low-delay filter bank technology.

A generic parametric coding tool for high frequencies, SBR manages to overcome problems associated with the limiting of perceptible coding artefacts to a subjectively acceptable level. Source entropy limitation and coding gain optimisation are traditionally achieved by reducing the codec audio bandwidth and sampling frequency, but SBR takes another approach: simply put, the decoder reconstructs higher frequency components with the help of the low-frequency base band and a very compact parametric description of the high band. The low-frequency base band of the signal is coded by a conventional core coder, whilst SBR reconstructs the high frequency band by transposing the low frequency band signal followed by a level adjustment. This algorithm exploits the harmonic structure of natural audio signals.

Simple implementation of the SBR within AAC-ELD would lead to unacceptable algorithmic delay, however, which is why the AAC-ELD also employs low delay filter banks in the AAC core codec and the SBR tool.

The LD-MDCT filter bank in the AAC-ELD core adopts an asymmetrical shape for windowing functions which simultaneously allowing the reduction of overlap towards future values and the extension of impulse responses towards past samples. As a result, the filter bank delay is reduced by 240 samples – from 959 down to 719 (the impulse response, meanwhile, is extended to the past by 960 samples). A similar delay saving window allows the SBR tool to add only 1.3 ms delay at a sampling rate of 48 kHz.

Delay is thereby reduced to such an extent that it is possible to bring the bit-rate-saving capabilities of the SBR to bear on bidirectional communication.

Feature-set

Thanks to the aforementioned technologies, AAC-ELD is able to offering the following features:

Highest sound quality. This super wide band audio codec delivers CD-like audio quality for voice, music and ambient sounds.

Flexible bit-rate range. AAC-ELD is optimized for a wide bit-rate interval – from 24 kbit/s all the way to 64 kbit/s (or even higher if required).

Low delay. AAC-ELD guarantees low algorithmic delay – as low as 15 ms at 64 kbit/s and 32 ms at 24 kbit/s.

Effective error concealment. This codec provides good intelligibility at up to 30 percent packet loss rate.

3. ACOUSTIC ECHO CONTROL

As many everyday users will be able to attest, echoes can constitute a frustrating impediment to phone calls, be they for business or personal matters. The quality and value of the communication can be adversely affected – to the extent that participants may be moved to abbreviate or even abruptly terminate their conversations.

The origin of these acoustic echoes is easily explained: an acoustic coupling between the loudspeakers and microphones of telecommunication devices gives rise to an acoustic feedback signal that is transmitted to the far-end subscriber, who hears a delayed version of his/her own speech.

To avoid these distracting echoes and other associated effects – including 'howling' and instability of the acoustic feedback loop – an effective echo control is an evident necessity. Consequently, the Audio Communication Engine incorporates an Acoustic Echo Control, which addresses these issues and eliminates the coupling between loudspeakers and microphones in any full-duplex hands-free telecommunication system. This new approach relies on Fraunhofer's know-how about the human perception of sounds. It supports full audio bandwidth to provide high audio quality for all kinds of sound signals, including speech, music and background noise. This enables natural, more convenient and efficient hands-free communication. In addition, Fraunhofer's Acoustic Echo Control supports multi-channel signals and offers a more robust performance.

Basic approach

In the Fraunhofer Acoustic Echo Control, the frequency spectrum of the microphone signal is modified so that the undesired echo components are removed from the signal transmitted to the far-end.

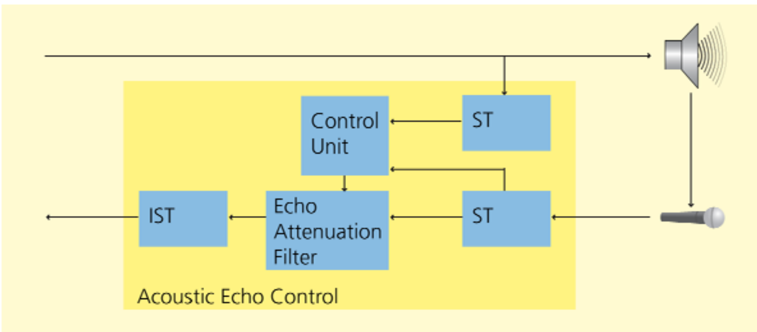

Figure 5: Basic approach of the Fraunhofer Acoustic Echo Control

The first stage in this technique entails the transformation of both the loudspeaker and microphone signals into the frequency domain by a spectral transform (ST). Based on these input signals, the Acoustic Echo Control determines an optimum gain factor for each individual frequency band separately. These gain factors – also referred to as echo attenuation factors – are selected on the basis of application and requirement. Once the echo attenuation filter has been applied to the spectral representation of the microphone signal, the echo-free signal is transformed back to the time domain by a corresponding inverse spectral transform (IST) (see figure 5). Usually, it is possible to achieve robust and consistent attenuation of the echo by 60 dB.

In addition, the Acoustic Echo Control offers an optional noise reduction module to enable the removal of undesired stationary background noise captured by the microphone.

Feature-set:

High quality audio. Extensive sampling rate support and a sophisticated filter bank are among the features that allow the Acoustic Echo Control to deliver high sound quality.

In addition to allowing 8 kHz telephone speech, Acoustic Echo Control supports a host of sampling rates (16, 32, 44.1 and 48 kHz) and employs a filter bank that shares important features of filter banks commonly used in perceptual audio coding. Human perceptions with respect to time/frequency processing are reflected in the spectral modification of the microphone signal by the echo attenuation filter.

Acoustic Echo Control also provides for the removal of non-stationary speech and stationary background noise echo components separately in order to more accurately reflect their contrasting requirements.

Multi-channel support. Acoustic Echo Control supports not only mono communication applications, but also stereo operation or a larger number of loudspeakers and microphones (for example standard 5.1 home theatre systems). CPU consumption scales reasonably and effectively in line with the increased number of channels thanks to an efficient combination of the loudspeaker and microphone channels.

Robustness. Fraunhofer Acoustic Echo Control offers consistent performance – not least because its estimation algorithms rely solely on the power or magnitude spectra of the loudspeaker and microphone signals; the phase information of the signal spectra is discarded. As a result, the technology’s performance is independent of any phase changes or distortions introduced by the acoustic echo path. This means that there is full immunity to effects including time drift caused by sampling rate mismatch between loudspeaker and microphone signals, and sample losses resulting from drop-outs during loudspeaker playback. The approach also offers protection against movements of the microphone or other changes in the acoustic environment.

Low complexity. The idea of ‘keeping it simple’ underlined the entire development of the Acoustic Echo Control, and it is with this in mind that the technology prioritises a low complexity implementation.

This approach is manifested in a reduced frequency resolution compared to some other techniques – for example, Fourier transform-based approaches. Spectral smoothing is performed in critical bands with regard to human perception, and is applied to the loudspeaker spectra, the microphone spectra and the model of the acoustic echo path. In some typical scenarios, it is possible for the echo components to be represented by less than a hundred parameters as opposed to the thousands of spectral values involved in an exact Fourier representation.

Low-Delay IP-Streaming System

Internet Protocol (IP) has become the most important network protocol for almost any application today. Not only are file downloads and web pages being carried over IP, but also an increasing amount of real-time services. In particular, VoIP applications continue to develop in their scope and popularity, with services such as Skype bringing flexible and highly cost-effective communications to both business and personal users.
Fraunhofer's Audio Communication Engine caters to the most demanding needs of the rapidly expanding IP communications markets.

The transmission of real-time content over IP networks generally includes the Realtime Transport Protocol (RTP) on top of the User Datagram Protocol (UDP). Despite its name, RTP does not assure real-time transport or any Quality of Service; instead, it only provides the necessary information to detect and handle unwelcomed network effects. As described in the following passages, the Audio Communication Engine makes best use of this information by advanced processing techniques in the application layer.

The first of the two primary effects is frame loss, wherein the IP frame is discarded by the network and so never delivered to the receiver. This effect is resolved through loss detection based on RTP sequence numbers and subsequent error concealment. Based on the overlapping structure of audio frames in AAC-ELD, it is possible to maintain good audio quality despite packet loss. The error concealment implemented in the Audio Communication Engine provides good intelligibility at packet loss rates of up to 30 percent.

The second of the two primary effects is network delay jitter, which describes the fact that packets undergo a variable delay as they traverse the network. Approaches to this issue vary, but Fraunhofer's preferred method is adaptive playout.

Adaptive Playout

One of the most challenging tasks in low-delay IP-streaming is to set the correct size of the de-jitter buffer. Normally, a fixed playout deadline is chosen, meaning that packets undergo a fixed end-to-end delay. However, if this fixed delay is set too low, then packets may arrive too late to be played out in time. Since these packets are not lost on the network but arriving late, the term *late loss* is used. By contrast, when the fixed delay is set too high, all packets are received in time but with the drawback of an unnecessary high *buffering delay*, i.e. the amount of time that a packet is kept in the buffer before playout. Because the variations in network delay are not easy to predict, it is very difficult to find a good trade-off between late loss and buffering delay when using fixed playout.

As a solution to this problem, adaptive playout offers an algorithm which estimates the jitter on the network and adapts the size of the de-jitter buffer in order to minimise buffering delay and late loss. For example, if the delay and/or jitter increases on the network, then the playout time is increased to reduce late loss. By contrast, if the delay and/or jitter decreases at a certain point, the buffer is reduced to minimize buffering delay (see figure 6).

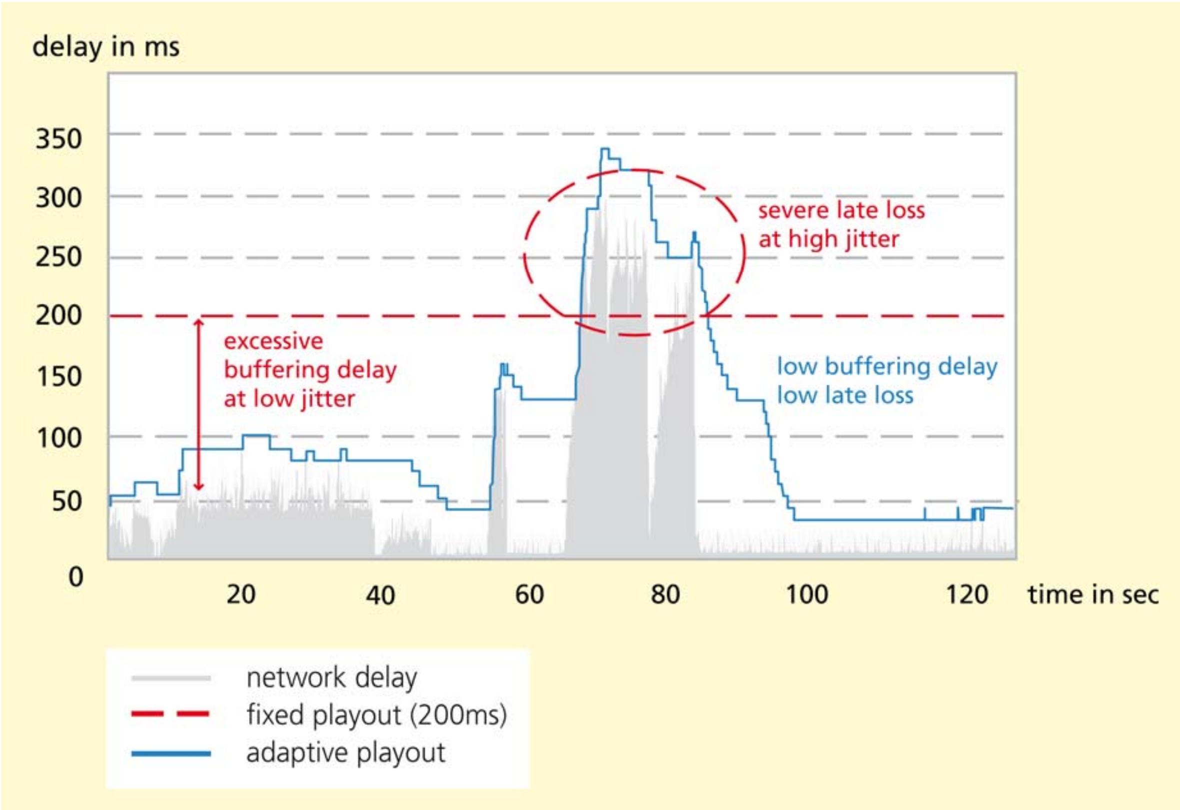

Figure 6: Adaptive playout enables the minimisation of buffering delay while simultaneously reducing late loss

A number of issues surround the effective implementation of adaptive playout. One important task is the effective modification of the time scale which is required to change the playout delay with the least possible subjective distortion. This is achieved by exploiting the effective error concealment properties of AAC without extra computational cost. To estimate the current jitter, the Audio Communication Engine maintains a histogram of previous delay values within a sliding window. As a result, it is possible to allow the estimation of delay percentiles in a flexible way and to derive appropriate playout times. Finally, a sophisticated and configurable playout control algorithm makes use of the jitter estimate and decides on the optimal jitter buffer size.

Fraunhofer's adaptive playout technology is able to guarantee reliable IP-based audio transmission with a significantly improved trade-off between late loss and buffering delay. Extensive testing has demonstrated the advantages of adaptive playout in comparison to fixed playout. Needless to say, applications and network situations will vary considerably, but the adaptive playout algorithm can reduce the loss rate by more than one magnitude compared to fixed playout when typical network conditions and same average delay are in play. Moreover, it is clear that adaptive playout provides robustness against the more critical situations; these are increasingly likely to occur the longer a given conversation endures.

Seamless Codec Switching

The IP-streaming system of the Audio Communication Engine is capable of switching codecs and their configuration seamlessly from frame to frame. For this purpose, not only is AAC-ELD from the AAC codec family supported, but also AAC-LC, HE-AAC, and HE-ACCv2. Furthermore, these codecs can operate in mono, or multi-channel configurations using various sample rates and bit-rates. This flexibility supports a wide range of application scenarios and bit-rate requirements. Since the configurations can be switched instantly it is possible to adapt to varying network resources during the ongoing session. For example, this can be advantageous when the bit-rate of an ongoing mobile connection suddenly drops after a handover.

Feature-set:

In summary, Fraunhofer's low delay IP-streaming system is a reliable and fully-featured component for high-quality audio communication over IP networks. By employing the following features it ensures high audio quality even under adverse network conditions:

RTP compliance. IP-streaming is based on RTP and the Audio Communication Engine is fully compliant to this widely-adopted streaming protocol.

Low delay implementation. The Audio Communication Engine can achieve end-to-end ("mouth-to-ear") delays below 40 ms on a clean Local Area Network (LAN) when installed on standard Linux systems.

Error concealment. Based on the superior error concealment performance of AAC codecs, the Audio Communication Engine can tolerate packet loss rates of up to 30 percent.

Adaptive playout. Sophisticated jitter estimation and playout control ensures an improved trade-off between late loss and buffering delay.

Codec switching. Possibility to switch codecs (AAC-ELD, AAC-LC, HE-AAC, HE-AACv2) and configuration (mono, stereo, multi-channel, bit-rates, sampling rates) from frame to frame.

4. WORKLOAD OF THE AUDIO COMMUNICATION ENGINE

The overall workload and memory requirements of the Audio Communication Engine result mainly from the workload caused by the audio codec and the Acoustic Echo Control whereas the low-delay IP-streaming system does not generate remarkable workload.

The workload of the Low Delay AAC encoder and decoder on TI C64x (mono, 32 kHz) is below 25 MHz. The workload of the Acoustic Echo Control on ARMv7 (stereo, 48 kHz) is 95 MHz.

For workload and memory requirements of the current Audio Communication Engine software version on dedicated platforms, please contact us directly at amm-info@iis.fraunhofer.de.

5. SUMMARY

In its incorporation of Enhanced Low Delay AAC (AAC-ELD) or Low Delay AAC (AAC-LD), Acoustic Echo Control and low-delay IP-streaming technology, the Fraunhofer Audio Communication Engine provides a comprehensive, 'all bases covered' solution for a wide range of modern audio communication applications. The individual components of the solution have already garnered interest from major manufacturers, broadcasters and service providers – for example, Teliris recently implemented Acoustic Echo Control into its telepresence 6G solutions. The flexibility and configurability options indicate that the Fraunhofer IIS Audio Communication Engine is ideally-placed to satisfy both current and future requirements.

6. AVAILABILITY AND LICENSING

The Audio Communication Engine is available for evaluation as hardware kit as well as evaluation software for Windows, Mac OS and Linux.

The software is licensed as object and source code for PC platforms, including Windows, Mac OS and Linux, as well as for embedded processors, including TI, ARM and Analog Devices. It can be licensed as a complete system; the audio codec AAC-(E)LD can also be licensed as standalone software.

The ACE is also available on iOS in which case it uses the built-in AAC-ELD codec. Since in this case AAC royalties do not apply for each single download, very attractive business models become possible for app developers.

For more information and licensing questions, please contact amm-info@iis.fraunhofer.de.

Watch this video to see how the Audio Communication Engine was developed: http://www.youtube.com/watch?v=R1DLw9aDQJg.

ABOUT FRAUNHOFER IIS

Fraunhofer IIS, based in Erlangen, Germany, is the home of the Audio and Multimedia division, which has been working in audio coding technology for more than 20 years and remains a leading innovator of technologies for cutting-edge multimedia systems. Fraunhofer IIS is universally credited with the development of mp3 and co-development of AAC (Advanced Audio Coding) as well as technologies for the media world of tomorrow, including MPEG Surround, MPEG Spatial Audio Object Coding and the Fraunhofer Audio Communication Engine.

Through the course of more than two decades, Fraunhofer IIS has licensed its audio codec software and application-specific customizations to at least 1,000 companies. Fraunhofer estimates that it has enabled more than 5 billion commercial products worldwide using its mp3, AAC and other media technologies.

The Fraunhofer IIS organization is part of Fraunhofer-Gesellschaft, based in Munich, Germany. Fraunhofer-Gesellschaft is Europe's largest applied research organization and is partly funded by the German government. With nearly 20,000 employees worldwide, Fraunhofer-Gesellschaft is composed of 60 Institutes conducting research in a broad range of research areas. For more information, contact Matthias Rose, matthias.rose@iis.fraunhofer.de, or visit www.iis.fraunhofer.de/amm.