Fraunhofer IIS

TECHNICAL PAPER

THE FUTURE OF COMMUNICATION: FULL-HD VOICE POWERED BY EVS AND THE AAC-ELD FAMILY

We have grown accustomed to "HD Everywhere" by consuming high fidelity content in most aspects of our lives. State-of-the-art audio and video codecs such as MPEG AAC and H.264 have set our expectations by assuring the highest rich media quality at very low bit-rates. The only real exception to the omnipresence of high-quality sound is the phone call, which is still largely tied to the limitations of technologies derived from the last century.

With Full-HD Voice, a new era of audio quality for the telecommunications market has begun. Full-HD Voice offers an unsurpassed level of quality, resulting in calls that sound as clear as talking to someone in the same room, or listening to high-quality digital audio.

The codecs enabling Full-HD Voice audio quality include the AAC-ELD audio codec family, tailored for over-the-top (OTT) voice services and already used in millions of calls today, and the new next-generation 3GPP communication codec Enhanced Voice Services (EVS), specifically designed to improve mobile phone calls in managed networks.

Fraunhofer Institute for
Integrated Circuits IIS

Management of the institute
Prof. Dr.-Ing. Albert Heuberger
(executive)
Dr.-Ing. Bernhard Grill
Am Wolfsmantel 33
91058 Erlangen
www.iis.fraunhofer.de

Contact
Matthias Rose
Phone +49 9131 776-6175
matthias.rose@iis.fraunhofer.de

Contact USA
Fraunhofer USA, Inc.
Digital Media Technologies*
Phone +1 408 573 9900
codecs@dmt.fraunhofer.org

Contact China
Toni Fiedler
china@iis.fraunhofer.de

Contact Japan
Fahim Nawabi
Phone: +81 90-4077-7609
fahim.nawabi@iis.fraunhofer.de

Contact Korea
Youngju Ju
Phone: +82 2 948 1291
youngju.ju@iis-extern.fraunhofer.de

* Fraunhofer USA Digital Media Technologies, a division of Fraunhofer USA, Inc., promotes and supports the products of Fraunhofer IIS in the U. S.

1. PHONE CALLS TODAY - PHONE CALLS TOMORROW

It is no secret that the vast majority of phone calls sound muffled compared to other sources of audio. Calls today have shortcomings that can make it difficult to understand conversations especially in noisy or reverberant environments, listen to talkers with soft or whispered speech and follow conversations in a non-native language or with an accent. For example, the low audio bandwidth makes distinguishing between certain consonants such as "f" and "s" quite difficult. Both share a similar low frequency spectral envelope, but the "s" phoneme is characterized by its significant energy in 10 kHz frequency range.

Most telephone calls employ low audio bandwidth speech codecs which model the human speech system and can only reproduce the human voice reasonably well. Phone calls are therefore limited to speech only, shutting out more natural communication options that include multiple speakers, ambient sounds, or music. Also, delays lead to involuntary interruptions, impacting natural conversation flow.

Codecs used in telecommunication applications such as POTS, ISDN and mobile phone calls are mostly limited to an upper frequency of 3.4 kHz - narrowband. Because people are able to hear audio signals of much higher frequencies (between 14 and 20 kHz at normal listening levels), most phone services are simply deleting at least three quarters of the audible spectrum. This causes the muffled sound in everyday telephony.

The HD Voice services recently introduced by some telephony providers have raised audio bandwidth from 3.4 to around 7 kHz, resulting in wideband transmission. The speech codecs used in these services provide audibly better quality compared to legacy calls, but still only transmit less than half of the fully audible audio spectrum. Now, with new technologies providing audio bandwidths of 14 kHz and higher, Full-HD Voice is available. The new codecs offer coding of superwideband speech and audio (14-16kHz bandwidth) or even fullband audio (20 kHz bandwidth) and deliver an unprecedented level of performance for voice calls, raising the quality to the level experienced in most digital media today. This is achieved by using optimized audio codecs (AAC-ELD) or hybrid solutions uniting both worlds of audio and speech coding (EVS).

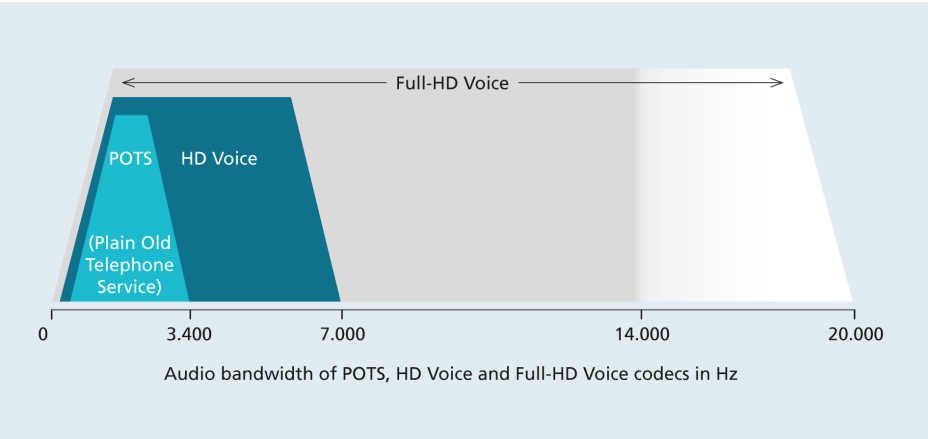

Figure 1: Bandwidth comparison of POTS, HD Voice and Full-HD Voice

For medium bit rates, as used in professional applications and OTT audio and video services, the AAC-ELD family consisting of AAC-LD (Low Delay AAC), AAC-ELD (Enhanced Low Delay AAC) and AAC-ELD v2 fulfills these high demands. All three codecs are based on the highly successful AAC audio codec and are used for video conferencing systems as well as IP based telecommunication services such as Apple FaceTime.

However, the AAC-ELD family is not suitable to enable Full-HD Voice at the low bitrates typically used by operators of managed mobile networks. With the new communication codec EVS developed within 3GPP, there is now a highly efficient audio solution for mobile communication available.

2. UNDERSTANDING THE CODECS BEHIND FULL-HD VOICE

2.1 EVS for mobile communication

2.1.1 Mobile telephony on the fast lane

Since Long Term Evolution (LTE), the fourth generation of mobile network standards, has been introduced, cellular phone networks are starting to switch to IP based transmission. LTE is based on the older, established GSM and UMTS standards, offering an all-IP architecture and low latencies. It requires the deployment of all-IP voice services or Voice-over-LTE (VoLTE) and in turn opens up the prospect of moving all voice services onto IP networks, eventually phasing out the legacy-switched services based on GSM, UMTS, CDMA networks, and wired public switched telephone networks.
With the help of Full-HD Voice technologies service providers can shake off the limitations of these legacy services, including very limited audio bandwidth and the use of speech-centric codecs.
With this goal in mind, 3GPP, the international standardization organization for mobile telephony, decided to develop and standardize the new communication codec EVS. It targets packet based systems such as VoLTE (Voice over LTE) or VoWifi (Voice over Wifi), but may be available for 3G/circuit switched systems as well in the future.

2.1.2 EVS – the all-rounder

Introduced in 2014, EVS is the first 3GPP codec to cover the complete audible audio spectrum of up to 20 kHz, pushing mobile phone calls to a new level. Combining state-of-the-art speech and audio compression technologies, EVS enables unprecedented audio quality for speech, music and mixed content.
EVS offers a wide range of bit rates from 5.9 kbit/s to 128 kbit/s, allowing service providers to optimize network capacity and call quality as desired for their service. Bit rates for narrow- and wideband start at 5.9 kbit/s, while superwideband Full-HD Voice audio quality is supported from 9.6 kbit/s on. EVS significantly improves the audio quality over legacy codecs at popular mobile bit rates such as 13.2 kbit/s and 24 kbit/s.
Another advantage is the codec's backward compatibility to AMR-WB which is enabled by an interoperability mode. It allows seamless switching between VoLTE and circuit switched networks when network conditions warrant a transition.

Mobile network services are often victims of packet loss issues with an unmistakably negative effect on speech intelligibility. Several unique concealment techniques, including a so-called channel-aware mode (CAM) using partial redundancy to improve concealment technologies, have shaped EVS to be a robust audio codec that minimizes errors and quickly recovers from lost packets. Its highly efficient jitter buffer management tops off the all-rounder package for a high quality communication experience.

The following diagrams compare codec quality on the basis of the mean opinion score (MOS).

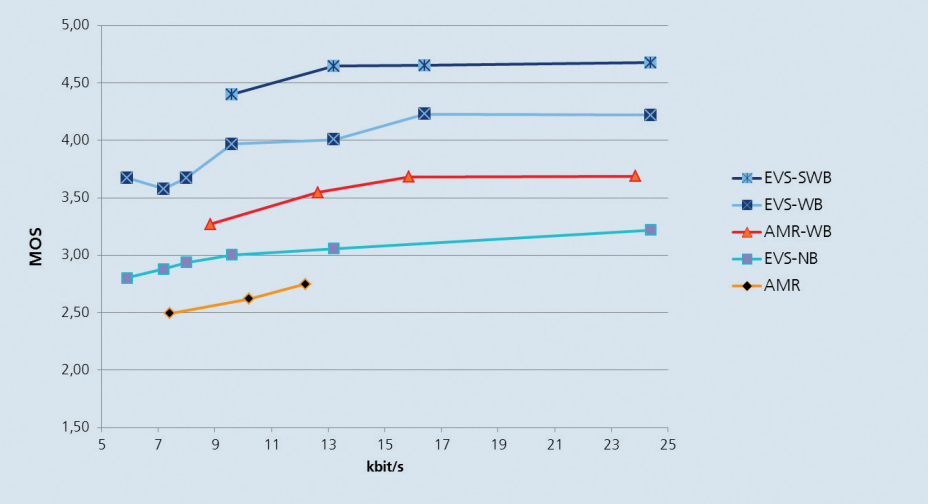

Figure 2: Quality comparison for clean speech between EVS (narrow-, wide- and superwideband) and AMR-NB (narrowband) and AMR-WB (wideband). Today's standard quality equals the pictured AMR-NB values.
AMR-WB applies to today's HD Voice quality. The Full-HD Voice superwideband service enabled by EVS outperforms AMR-WB even at very low bit rates such as 9.6 kbit/s.
Source: 3GPP TR 26.952, EVS Performance Characterization, Experiment M1.

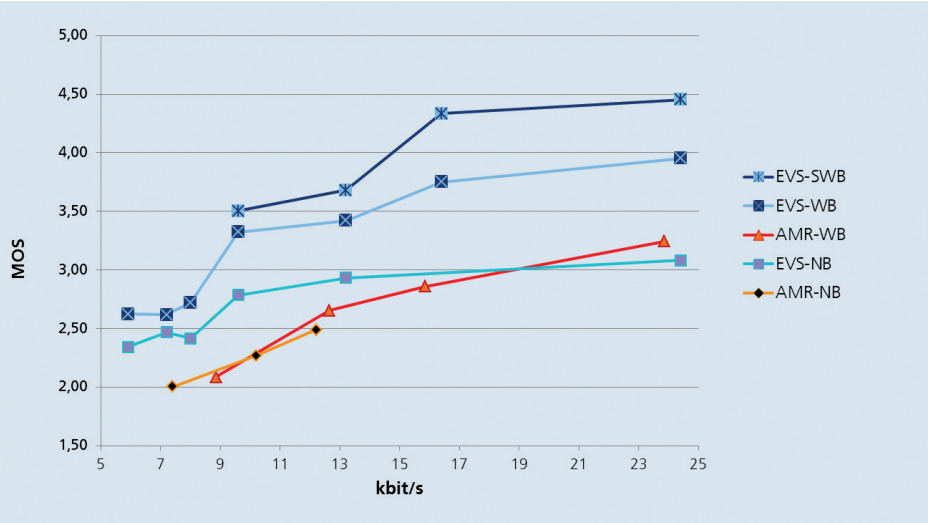

Figure 3: Quality comparison between EVS (narrow-, wide- and superwideband) and AMR-NB (narrowband)/AMR-WB (wideband) for music and mixed content. There's an evident quality gap between AMR/AMR-WB and EVS at low bit rates. EVS-WB and EVS-SWB significantly outperform AMR-WB, e.g.: The quality delivered by EVS at 9.6 kbit/s is still higher than that of AMR-WB at 23.85 kbit/s. Source: 3GPP TR 26.952, EVS Performance Characterization, Experiment M3b.

Figure 4: Robustness comparison between AMR-WB and EVS (CAM on and off). EVS maintains its high quality even at a frame error rate (FER) of up to 3% (CAM off). EVS' performance at a FER of 6.2% is comparable to AMR-WB's performance at a FER of 3.3%. The same comparison applies for CAM enabled EVS at 9.2%.
Source: 3GPP TR 26.952, EVS Performance Characterization, Experiment W1.

2.2 The AAC-ELD codec family

The AAC-ELD family consisting of AAC-LD (Low Delay AAC), AAC-ELD (Enhanced Low Delay AAC) and AAC-ELD v2 fulfills the high demands of communication applications in terms of quality and latency. All three codecs are based on the highly successful AAC audio codec, which has been used by Apple in its iTunes music store since 2001 and today is deployed in billions of devices.

2.2.1 High efficiency, low latency

They support mono, stereo and multichannel signals – all with latencies as low as 15-20 ms. Furthermore, the audio codecs extend the application area from clean voice to a broad variety of source material, including voice and singing, music and ambient sounds. The result of all this is natural, real-time communication.

The audio quality and operating point of the AAC-ELD family members is described in Figure 5 for stereo audio. While AAC-LD is a very good choice for bit-rates above 96 kbit/s, AAC-ELD improves the audio quality down to 48 kbit/s. Below this bit-rate, AAC-ELD v2 is the best choice to keep the audio quality high. For mono applications, a similar relationship between AAC-ELD and AAC-LD at half bit-rate can be expected, whereas AAC-ELD v2 delivers identical audio quality to AAC-ELD.

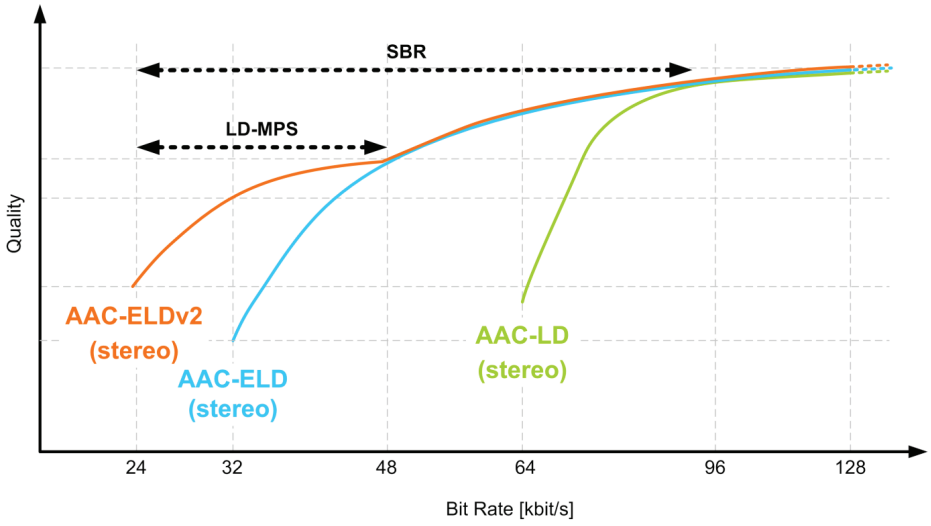

Figure 5: AAC-ELD stereo operating points

2.2.2 Application Scenarios

Over-the-Top Services and IP Video Telephony

Although OTT and VoIP services have been around since the mid to late 1990s, or even earlier with some Packet Radio experiments in Silicon Valley, credit for mass acceptance across services must go to Skype. However, there are limits to what a speech codec, such as Skype's SILK, can deliver in terms of quality.
To overcome the limitations of speech codecs, Apple's OTT peer-to-peer video telephony service FaceTime is based on the Full-HD Voice codec, AAC-ELD. As FaceTime is available on most Apple devices, such as iPhone, iPad, and Mac, the service can be used today on more than 200 million devices and is growing rapidly.
As AAC-ELD is natively supported in Android and iOS, Full-HD Voice is readily available in these operating systems.

Video Conferencing and Telepresence

Video conferencing and telepresence services have been all-IP-based technologies for many years. In these markets, the user expectations of both video and audio quality are demanding. As a result, Full-HD Voice has long been a default choice for these providers. Most companies are offering Full-HD Voice, and a majority of these products are based on the TIP standard, assuring interoperability between devices from different manufacturers. The TIP standard chose AAC-LD as the only mandatory codec besides G.711, which is the legacy voice codec used in narrowband telephony.

3. MORE INFORMATION

Read more about the AAC-ELD codec family in the EDN Network article "Full-HD Voice: Understanding the AAC codecs behind a new era in communication"
here: http://www.edn.com/design/consumer/4405424/Full-HD-Voice--Understanding-the-AAC-codecs-behind-a-new-era-in-communication.

ABOUT FRAUNHOFER IIS

The Audio and Media Technologies division of Fraunhofer IIS has been an authority in its field for more than 25 years, starting with the creation of mp3 and co-development of AAC formats. Today, there are more than 10 billion licensed products worldwide with Fraunhofer’s media technologies, and over one billion new products added every year. Besides the global successes mp3 and AAC, the Fraunhofer technologies that improve consumers’ audio experiences include Cingo® (spatial VR audio), Symphoria® (automotive 3D audio), xHE-AAC (adaptive streaming and digital radio), the 3GPP EVS VoLTE codec (crystal clear telephone calls), and the interactive and immersive MPEG-H TV Audio System.

With the test plan for the Digital Cinema Initiative and the recognized software suite easyDCP, Fraunhofer IIS significantly pushed the digitization of cinema. The most recent technological achievement for moving pictures is Realception®, a tool for light-field data processing.

Fraunhofer IIS, based in Erlangen, Germany, is one of 69 divisions of Fraunhofer-Gesellschaft, Europe’s largest application-oriented research organization.

For more information, contact amm-info@iis.fraunhofer.de, or visit www.iis.fraunhofer.de/amm.