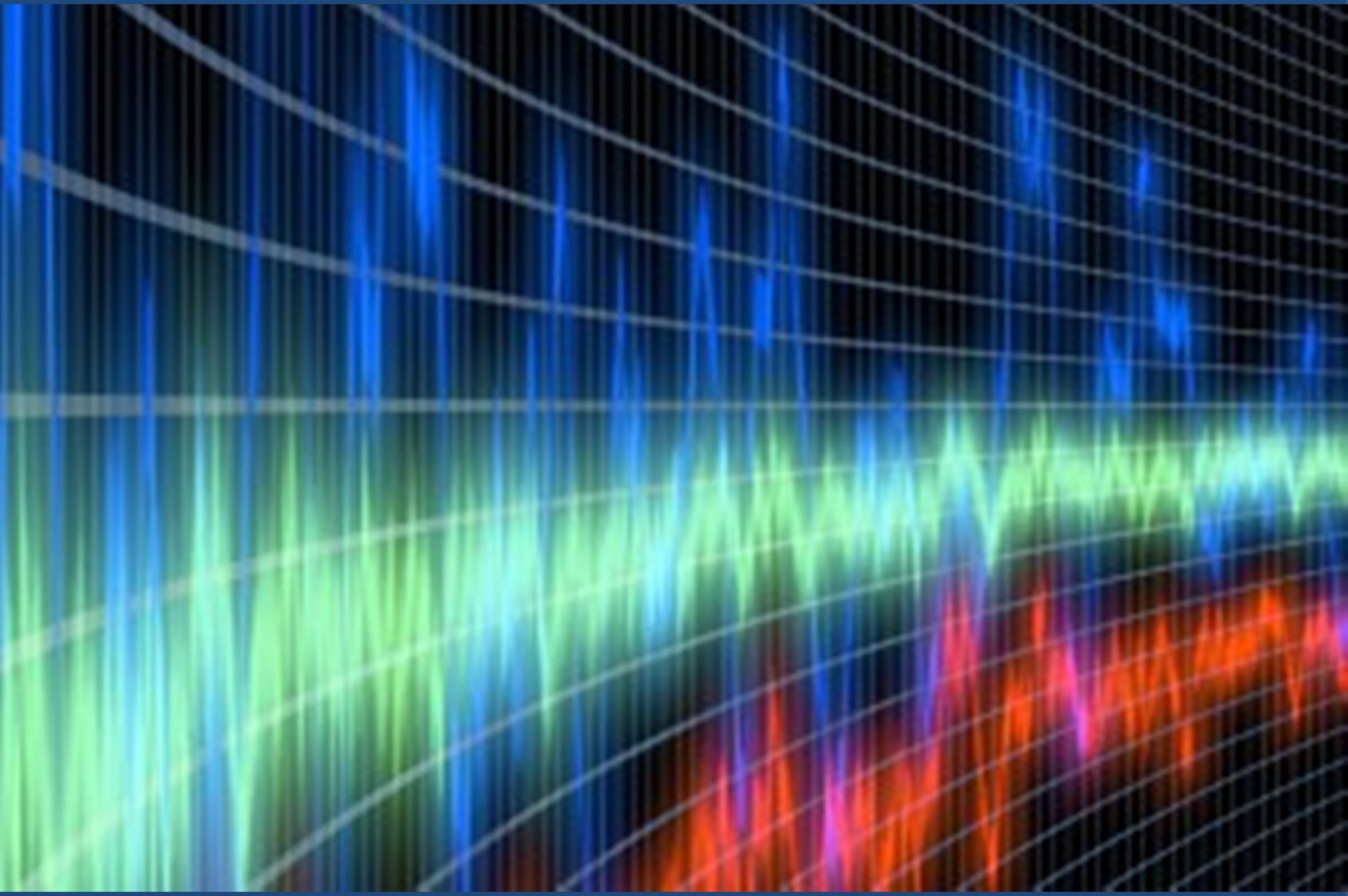
DIGITAL AUDIO COMPRESSION



Contents

Stay secured, Leave assured

Digital Audio Compression

For audio the most basic forms of data compression involve reducing the number of bits and sampling rate of the audio. Lowering the sample rate reduces the high frequency content of the original while reducing the bits per sample lowers the fidelity.

Different types of audio file compression formats exist because each file format offers certain advantages. The three most popular types of audio file compression schemes used for saving audio as digital files are lossless, lossless compression, and lossy compression.

Lossless audio files: The highest-quality audio files are called *lossless* because they never lose any audio data. As a result, lossless audio files offer the highest-quality sound, but they also create the largest file sizes.

Compressed lossless audio files: Lossless audio files take up large amounts of space, so compressed lossless audio files are designed to squeeze audio data into a smaller file size.

Compressed lossy audio files: A lossy audio file compresses audio files by stripping certain audio data to shrink the file size. The greater the audio quality, the more audio data the file needs to retain and the bigger the file. The smaller the file, the less audio data the file can hold, and the lower the audio quality. As a result, most audio file formats strive for a balance between the audio quality and file size.

MPEG Audio Compression

MPEG audio compression algorithm is an International Organization for Standardization (ISO) standard for high-fidelity audio compression. This is a lossy compression but this lost can hardly be noticed because the compression method tries to control it. By using several quite complicate and demanding mathematical algorithms it will only loose those parts of sound that are hard to be heard even in the This leaves more space for information that is important. This way you can compress audio up to 12 times (you may choose compression ratio) which is really significant. Due to its quality MPEG audio became very popular. As well as using a psychoacoustic model to mask the encoded signal, an MPEG encoder also feeds the input signal through a time-to-frequency mapping device that constructs sets of spectrum components for subsequent encoding. A certain number of data bits is used by the encoder to form a template for the entire length of the input signal, and subsequent information is stored in relation to the initial data, rather than as data in its own right. This concept is similar to the idea of storing ten letters "A"s as A(x10) or 10A rather than as AAAAAAAAAA - the nine subsequent letters are stored in relation to the first to take up less storage space.

Pulse Code Modulation

Voice and audio signals are analogic, whereas data network is digital. The transformation of the analogic signal to a digital one is made by Analog-to-Digital Converter (ADC). This process of Analog-to-Digital Converter or Pulse Code Modulation (PCM) is done in three steps:

- Sampling

- Quantization

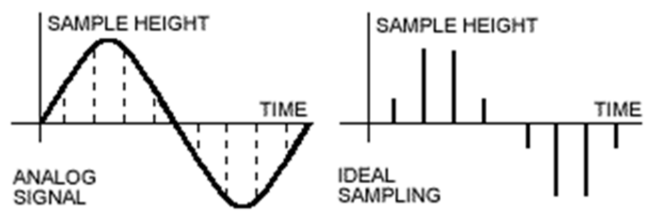
- Codification (codification)

In the quantization process a compression of the voice could be used as it will be explained in this section:

Sampling

Sampling is the process of encoding an analog signal in digital form by reading (sampling) its level at precisely spaced intervals of times. The obtained values are called samples. This process is shown in the following images. Sampling usually happens at equally separated intervals; this interval is called the sampling interval.

The reciprocal of sampling interval is called the sampling frequency or sampling rate. The unit of sampling rate is Hz.



Quantization

Quantization is the process of converting the height of the obtained samples to a finite number of discrete values. There are several methods to quantify that we will explain according to its complexity. As the follow, the two basic category of quantization step are described.

Uniform Quantization

It is necessary to use a finite number of discrete values to represent approximately the amplitude of the samples. All the amplitude ranges that the samples can take are divided in an equal number of intervals. All the samples whose amplitude falls within an interval, take the same value.

Not uniform quantization

In a uniform quantization the distortion or noise does not depend on the sample amplitude. Therefore, when the amplitude is lower the influence of the error or quantization noise is greater. The situation is critical for signals whose analogical amplitude is near the one of a quantification interval.

Encoding Law

The not uniform quantization process follows a certain feature called encoding law. There are two types of encoding laws: continuous and segmented. In continuous encoding laws, the quantization intervals have different width, growing from small values (corresponding to low level signals), to greater values, (corresponding to high level signals). In segmented encoding laws, the operation range is divided into a finite number of groups. Each interval of the same group has the same width, being different from other groups. Normally, the encoding laws used are segmented.

The two main encoding laws used nowadays are a-law and μ-law, which are also known as G.711 codec. This codec is introduced as below.

G.711 Codec

The G.711 codec is the most used and supported codec in IP telephony. It is produced using pulse code modulation at an uncompressed sampling rate of 8000 samples per second. The bandwidth required for the G.711 codec is 64,000 bits per second. This is the bandwidth of the payload (not including IP, RTP, and UDP headers).

There are two versions or formats of the G.711 codec. G.711 μ-Law is the codec used in North America. The G.711 a-Law is used outside North America. Even though both codecs have the same bit rate of 64,000 bits per second, they perform a completely different sampling of pulse code modulation to arrive at their respective digitized samples. Therefore, the codecs are not directly compatible and require transcoding between G.711 μ-Law and G.711 a-Law. However, both of these codecs produce a high-quality audio stream. Table 1 shows different types of G.xxx codecs.

G.729 Codec

The G.729 codec is also used extensively in IP telephony and also widely supported. This codec uses a compression algorithm to attain a payload bandwidth of 8000 bits per second. Because of bandwidth conservation, this codec is used for remote IP telephony communications and where bandwidth oversubscription is a concern. A number of versions of the G.729 codec exist. Two of these codecs, G.729a and G.729b, incorporate additional options and features. The sound quality produced using G.729 is not as high quality as G.711 but is still considered to be toll quality (similar to a residential phone service or traditional landline services). These lower bandwidth codecs are used primarily to save the bandwidth for lower speed WAN circuits. In these cases, the overhead calculation is still approximately 16 k, providing a total bandwidth calculation of 24 k.

G.722 Code

The G.722 codec produces is a high quality audio signal and is supported on many of the newer IP telephony devices and IP phones. G.722 uses its own compression algorithm called Sub-Band Adaptive Differential Pulse Code Modulation (SB-ADPCM) and can produce a digital signal using a number of bandwidths (48 k, 56 k, and 64 k). The G.722 codec requires 64,000 bits per second as the payload bandwidth for this codec; although it can adapt the compression algorithm based on changes in the network.

G.726 Codec

The G.726 codec uses Adaptive Differential Pulse Code Modulation (ADPCM) to produce a payload bandwidth of 16 k, 24 k, 32 k, or 40 k bits per second, although the most widely supported codec used is 32 kbps. Using half the bandwidth of G.711, this codec is used for many phone service providers, VPIM networking, and Simple Mail Transfer Protocol (SMTP) communications.

Table 1 Audio Compression standards

Name	standardized by	description	bit rate (kb/s)	sampling rate (kHz)
G.711	ITU-T*	Pulse code modulation (PCM)	64	8
G.721	ITU-T	Adaptive differential pulse code modulation (ADPCM)	32	8
G.722	ITU-T	7 kHz audio-coding within 64 kbit/s	64	16
G.722.1	ITU-T	Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss	24/32	16
G.723	ITU-T	Extensions of Recommendation G.721 adaptive differential pulse code modulation to 24 and 40 kbit/s for digital circuit multiplication equipment application	24/40	8
G.723.1	ITU-T	Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s	5.6/6.3	8
G.726	ITU-T	40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM)	16/24/32/40	8
G.727	ITU-T	5-, 4-, 3- and 2-bit/sample embedded adaptive differential pulse code modulation (ADPCM)	var.	?
G.728	ITU-T	Coding of speech at 16 kbit/s using low-delay code excited linear prediction	16	8
G.729	ITU-T	Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)	8	8

***ITU (International Telecommunication Union) Telecommunication Standardization Sector

iLBC

iLBC sits between G.711 and G.729. iLBC uses slightly more bandwidth than G.729 but results in better call quality. It's very hard to distinguish an iLBC call from a G.711a call. iLBC uses on average 24kbps (so 4kbps more than G.729) and uses more CPU than G.729, but it results in noticeably better audio quality than G.729. iLBC also is a free codec and does not require any licenses.

The drawbacks of using the iLBC codec are:

- iLBC is not as widely supported by VSPs as G.711a and G.729.

- iLBC uses even more CPU cycles than G.729 / a-Law.

AMR

In digital audio, the letters AMR are short for **A**daptive **M**ulti-**R**ate and relate to the AMR audio format. This audio file format, which was first released in 1999, is especially efficient at compressing and storing voice recordings compared to common formats like MP3, WMA, and AAC for example. It is a lossy format with files commonly identified with the. AMR extension -- the exception to this rule is that the 3GP container format can also be used to store AMR streams along with video. Incidentally, this type of voice coding technique is sometimes referred to as *vocoding*. There are essentially two AMR format standards which are AMR-NB and AMR-WB. The first one (AMR-NB), is a narrowband version which is commonly used in situations where low bitrates are sufficient -- such as a basic voice recording facility you may have on your MP3 player. The frequency range used for AMR-NB is 300-3400 Hz which is capable of producing sound quality that is comparable to traditional telephone. This narrowband version uses the following bitrates:

Narrow band ARM bitrates: 04.75 kbps, 05.15 kbps, 05.90 kbps, 06.70 kbps, 07.40 kbps, 07.95 kbps, 10.20 kbps, 12.20 kbps.

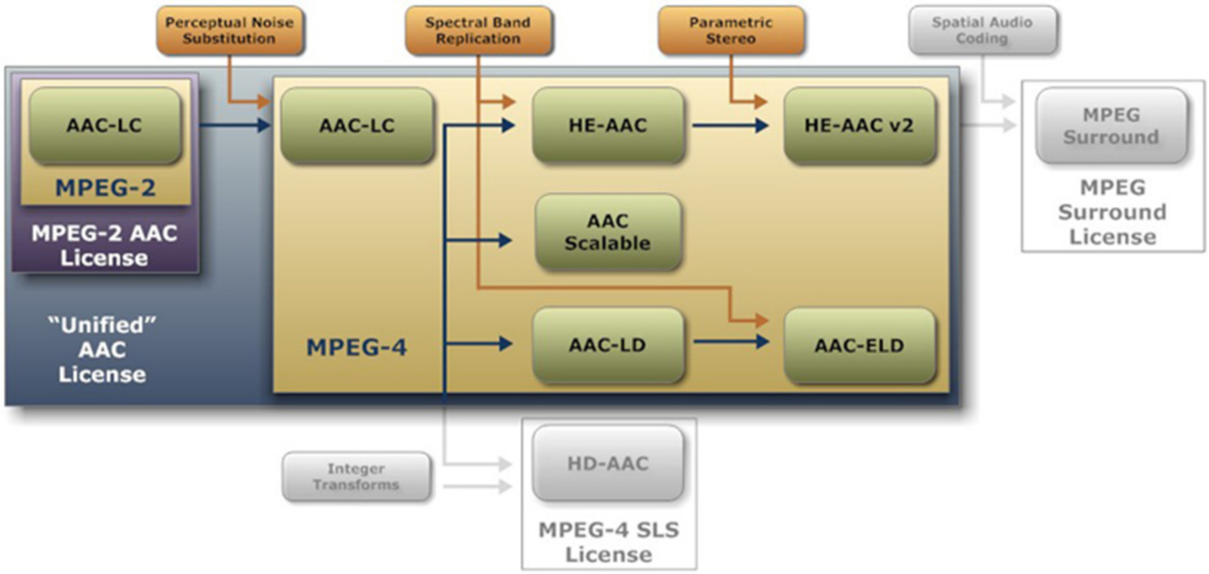
The second version of AMR is the wideband type which is represented by the acronym, AMR-WB. As the name would suggest, this is an enhanced vocoder that uses a wider bandwidth than AMR-NB in order to store voice at a much higher quality -- the frequency range used is for this is 50 -7000 Hz. The bitrates used for the wideband version of AMR are:
Wideband ARM bitrates: 06.60 kbps, 08.85 kbps, 12.65 kbps, 14.25 kbps, 15.85 kbps, 18.25 kbps, 19.85 kbps, 23.05 kbps, 23.85 kbps.

Due to its higher frequency range and therefore superior speech quality, AMR-WB is optimized for use in GSM (Global System for Mobile Communications) and UMTS (Universal Mobile Telecommunications System) technologies -- otherwise known as 2G and 3G mobile networks respectively.

Advanced Audio Coding -AAC

AAC is an audio compression scheme first standardized within MPEG in 1997. AAC was designed to provide high quality audio at lower bit-rates than previous MPEG audio compression formats. AAC was further refined through the MPEG-4 standardization process and has subsequently been enhanced with bandwidth extension technology yielding High Efficiency AAC (HE AAC), and with the addition of parametric stereo, resulting in High Efficiency AAC version 2 (HE AAC v2). Via Licensing administers a joint patent license which provides a convenient and cost-effective way to acquire the rights to practice the essential AAC patents from a set of licensors. Often referred to as the "Unified" AAC license, the program provides coverage for the AAC technologies identified in the Figure shown on the next page.

The core of this AAC "family" is a set of backwards-compatible audio coding technologies: MPEG-4 AAC LC (Low Complexity) decoders can playback MPEG-2 AAC LC bit-streams, MPEG-4 HE AAC (High Efficiency) decoders can playback both MPEG-4 and MPEG-2 AAC LC bit-streams, etc. In this way the AAC family can support a wide variety of applications ranging from extremely low bit-rates required for music delivery over cellular phone networks, to "transparent" quality (indistinguishable from the original source material) for the most discriminating listeners. Additional AAC technologies covered by the patent license include AAC-LD (Low Delay) and AAC-ELD (Enhanced Low Delay) which enable high quality audio for videoconferencing or other applications where low-latency performance is critical.



Unified AAC Description

Conclusions

In this whitepaper we reviewed some of the most widely used audio codecs and compression techniques used in video surveillance industry such as G series, iLBC, AMR and AAC. The characteristics and comparisons between several types of audio codecs are also discussed in the body of the whitepaper.