Introduction and Comparison of Common Videoconferencing Audio Protocols

I. Digital Audio Principles

Sound is an energy wave with frequency and amplitude. Frequency maps the axis of time, and amplitude maps the axis of level. Generally, there are three types of sound:

- Audible sound: This sound is clearly audible to the unaided ear, and the frequency ranges from 20 Hz to 20 kHz.
- Infrasound: The frequency is smaller than 20 Hz.
- Ultrasonic sound: The frequency is larger than 20 kHz.

Multimedia technologies only focus on audible sound.

For audible sound, the frequency band of speech signals ranges from 80 Hz to 3400 Hz, and the frequency band of music signals ranges from 20 Hz to 20 kHz. Multimedia technologies focus on processing of speech and music.

Because analog voice is continuous in terms of time, the voice signals collected from microphones must be digitized and then sent to the computer for processing. Generally, the pulse-code modulation (PCM) technology is used to convert a continuously variable analog signal to a digital signal by means of sampling, quantization, and coding.

1. Sampling

   Sampling: reading amplitude of voice at a regular interval. The number of samples collected per second is represented by the sampling rate. Obviously, the larger the sampling rate, the closer the data points of the discrete amplitude values to the continuous analog signal curve. The total number of samples is larger.

   Based on the sampling principle, the sampling rate must be larger or equal to two times of the maximum frequency in the analog signal frequency band, ensuring that a digital audio can be accurately transformed to an analog audio.

   Common audio sampling rates: 8 kHz, 11.025 kHz, 22.05 kHz, 16 kHz, 37.8 kHz, 44.1 kHz, and 48 kHz.

   For example, assume that the frequency of a speech signal ranges from 0.3 kHz to 3.4 kHz and a sampling rate of 8 kHz is used, the signals that are used to replace the original continuous speech signals can be sampled. Generally, CD audio has a sampling rate of 44.1 kHz.

2. Quantizing

   Quantizing: converting amplitudes of voice signal to digits to represent the signal strength.

   Quantizing precision: The maximum number of bits that each sample can be represented, also known as sampling resolution. Generally, the sampling resolution of voice signal can be 4 bits, 6 bits, 8 bits, 12 bits, or 16 bits.

   Based on the sampling rate and quantization precision, natural voice signals can be indefinitely closer but cannot be analogized completely by using an audio codec. In computer application, the highest fidelity of sound can be achieved by using the PCM, which is regarded as a lossless data compression.

3. Coding

   Coding: assigning a unique digital code to input signals.

   If the sampling rate is 44.1 kHz, quantization with 16 bits is used, and two-way audio is output by using the PCM, the bit rate is 1411.2 kbit/s (44.1 kHz x 16 x 2 = 1411.2 kbit/s). Therefore, the required storage space in one second is 176.4 KB and in one minute is 10.34 MB. Digital audio signals must be compressed so that the transmission and storage cost can be reduced.

Currently, the bit rate of an audio signal can be reduced to 32–256 kbit/s after compression, and the bit rate of a speech signal can be reduced to a value smaller than 8 kbit/s.

Digital audio information is compressed so that its data amount can be minimized without impacting on the use of the information. Generally, the following six attributes are used to measure the data amount of digital audio information:

- Bit rate
- Signal bandwidth
- Subjective and objective voice quality
- Delay
- Algorithm complexity and storage requirement
- Sensitivity to channel bit error

Audio coding is based on the standard algorithm. In this way, the coded audio information can be widely used. Traditional videoconferencing audio devices mainly adopt standards such as G.711, G.722, G.728, and AAC_LD. These standards are released by the International Telecommunication Union Telecommunication Standardization Sector (ITU-T).

II. Common Audio Protocols

1. ITU-T G.728

   G.728 is a telephone speech codec standard released by ITU-T in 1992. G.728 adopts the low-delay code excited linear prediction (LD-CELP) coding mode, uses a sampling rate of 8 kHz, and transmits voice signals at a bit rate of 16 kbit/s with a 0.625 ms algorithm delay.

2. ITU-T G.711

   G.711 was released for usage by ITU-T in 1972. Speech signals are coded by using the non-uniform quantization PCM methods. G.711 uses a sampling rate of 8 kHz and uses a non-uniform quantization with 8 bits to represent each sample, resulting in a 64 kbit/s bit rate. This narrowband codec supports compression of 300 Hz to 3400 Hz audio. G.711 defines an excellent compression quality. However, a relatively lager bandwidth is occupied. G.711 is applicable to digital telephones in the PBX/ISDN system.

3. ITU-T G.722

   ITU-T G.722 is the first standard wideband speech codec that uses a sampling rate of 16 kHz. This codec was approved by the International Telegraph and Telephone Consultative Committee (CCITT) in 1984 and is still in use nowadays. G.722-based codec receives 50 Hz to 7 kHz audio data by using a 16 kHz sampling rate and a 16-bit quantization, and then compresses the audio data to reach a bit rate of 64 kbit/s, 56 kbit/s, or 48 kbit/s. The total delay is about 3 ms, and a high-quality audio is provided.

   G.722 has the following advantages:

   - Low delay
   - Low transmission bit error rate
   - No patented technology
   - Low cost

   Therefore, G.722 is widely used in Voice over IP (VoIP) services, personal communication services, and videoconferencing applications.

4. G.722.1

   G.722.1 is a third-generation Siren 7 compression technology developed by Polycom. G.722.1 standard was approved by ITU-T in 1999. G.722.1 uses a sampling rate of 16 kHz and a 16-bit quantization. G.722.1 supports sampling of 50 Hz to 7 kHz audio data, and compresses the audio data to reach a bit rate of 32 kbit/s or 24 kbit/s. G.722.1 encapsulates voice signals into frames at 20 ms and provides a 40 ms algorithm delay.

With comparison to G.722, G.722.1 can implement lower bit rate and higher quality of audio data compression. G.722.1 aims to provide audio with the same quality defined in G.722 by operating at a half of the bit rate. To use this codec, you must obtain the authorization of the codec from the Polycom.

5. G.722.1 Annex C

   G.722.1 Annex C is a Siren 14 compression technology developed by Polycom. It uses a sampling rate of 32 kHz, supports sampling of audio data with frequency ranging from 50 Hz to 14 kHz, and compresses the audio data to reach a bit rate of 24 kbit/s, 32 kbit/s, or 48 kbit/s. G.722.1 Annex C encapsulates voice signals into frames at 20 ms and provides a 40 ms algorithm delay.

   In 2005, the Siren 14™ technology of the Polycom was approved by ITU and was a new standard for 14 kHz wideband speech codecs. The mono version of Siren 14 became ITU-T G.722.1 Annex C.

   G.722.1 Annex C has the following advantages:

   - Defines a low requirement for calculation capability and bandwidth.
   - Be applicable to the processing of speech, music, and natural voice.

6. AAC_LD

   Advanced Audio Coding (AAC) is an audio compression format developed by the Fraunhofer Institute (also known as the developer of the MP3 format), Dolby Laboratories, and AT&T Corporation. It is a part of the MPEG-2 standard and was approved to be an international standard for audio compression technologies in March 1997. With the wide use of MPEG-4 in 2000, MPEG2 AAC was brought into use as another core codec technology and added with some new codec features. MPEG2 AAC is also known as MPEG-4 AAC.

   The MPEG-4 AAC family contains nine codec specifications. The MPEG-4 Low Delay Audio Coder (AAC_LD) is used in the case of low bit rate. AAC_LD uses a sampling rate of 8 kHz to 48 kHz and provides CD-quality audio at a bit rate of 64 kbit/s. In addition, AAC_LD supports multi-channel output and provides a 20 ms algorithm delay.

   AAC adopts modular design and provides more powerful functions.

7. Parameter Comparison of Audio Protocols

   Table 1-1 lists parameter comparison of audio protocols.

Table 1-1 Parameter comparison of audio protocols

	Sampling Rate	Supported Audio Frequency	Output Bit Rate	Minimum Algorithm Delay
G.711	8 kHz	300 Hz to 3400 Hz	64 kbit/s	< 1 ms
G.722	16 kHz	50 Hz to 7 kHz	64 kbit/s	3 ms
G.722.1	16 kHz	50 Hz to 7 kHz	24 kbit/s and 32 kbit/s	40 ms
G.722.1 C	32 kHz	50 Hz to 14 kHz	24 kbit/s, 32 kbit/s, and 48 kbit/s	40 ms
AAC_LD	48 kHz	20 Hz to 20 kHz	48 kbit/s to 64 kbit/s	20 ms

III. Advantage and Disadvantage Comparison of AAC_LD and G.722.1 Annex C

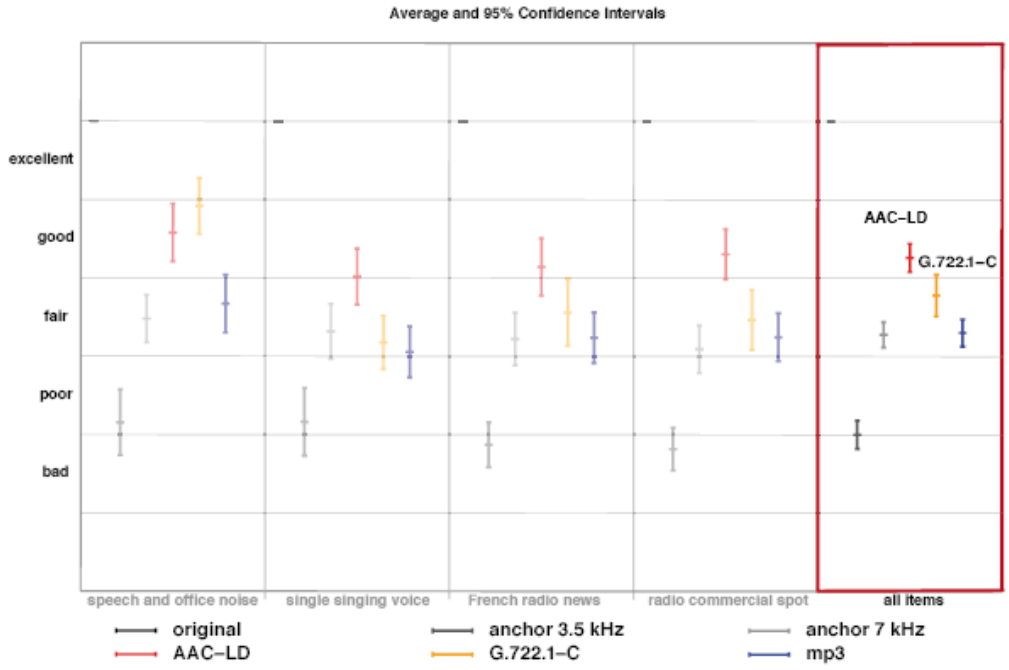
Table 1-2 lists advantages and disadvantages of AAC_LD and G.722.1 Annex C.

Table 1-2 Advantages and disadvantages of AAC_LD and G.722.1 Annex C

	G.722.1 Annex C	AAC_LD
Frequency of audio samples	Supports 50 Hz to 14 kHz audio. Defines CD-quality audio but fails to sample audio data with high frequency.	Supports sampling of 20 Hz to 20 kHz audio data. Defines better CD-quality audio than G.722.1 Annex C.
Bit rate	Supports a bit rate of 24 kbit/s, 32 kbit/s, or 48 kbit/s. Uses a smaller bandwidth than AAC_LD. Does not support audio output with high frequency.	Uses a bit rate of 48 kbit/s to 64 kbit/s and supports audio output at a bit rate larger than 64 kbit/s. Defines audio with higher quality.
Algorithm complexity	Uses an algorithm with low complexity. Occupies less CPU than AAC_LD.	Adopts modular design and provides more powerful functions. Defines specialized chips such as TI.
Minimum delay	Encapsulates voice signals into frames at 20 ms and provides a 40 ms algorithm delay.	Defines a 20 ms algorithm delay.
Multi-channel	Supports dual audio channel.	Supports a maximum of 48 tracks and 15 low-frequency tracks.
Commonality	Developed by Polycom. Used with authorization obtained from Polycom. Currently adopted only by Polycom and few videoconferencing vendors.	Obtains support from Apple, Nokia, and Panasonic as a core MPEG-4 standard. Adopted by a variety of videoconferencing vendors such as Ted. Has a wide prospect of application.

Figure 1-1 shows comparison of AAC_LD, G.722.1-C, and MP3 using speech items.

Figure 1-1 Comparison of AAC_LD, G.722.1-C, and MP3 using speech items



As shown in Figure 1-1 (provided by Fraunhofer Institute), AAC_LD provides higher quality audio than that defined in G.722.1 Annex C and MP3 in the case of same sampling rate. AAC_LD provides the smallest delay among wideband speech codecs. In addition, AAC_LD ensures the CD quality of audio and provides a best combination of audio quality, bit rate, and delay. Therefore, it is the best choice for videoconferencing vendors.