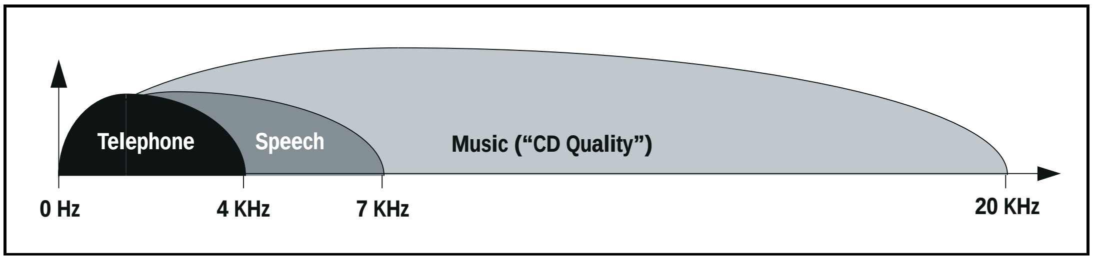
Audio and Voice Coding Fundamentals

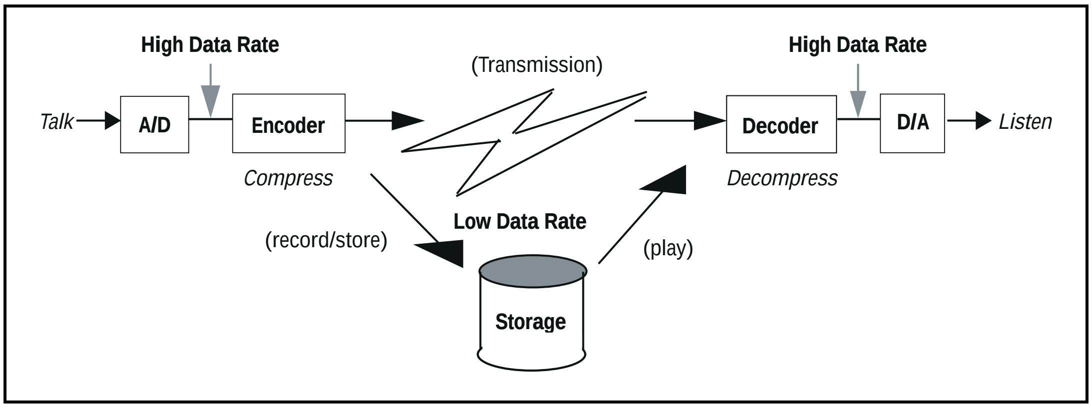
Overview



Audio and voice coding is used to compress audio and voice information for efficient digital storage or transmission. Humans can hear sound from a minimum of 5 Hz to a maximum of 20 KHz. As a result, audio coders encode digitized sound signals up to 20 KHz and “CD quality” audio systems provide 20 KHz of frequency bandwidth. This relatively large bandwidth is desired for high fidelity music listening, especially for classical music. However, the majority of audio information is located in the lower half of this 20-KHz band. Human speech rolls off between 7 and 10 KHz, and telephone lines only provide approximately 3.5 KHz of bandwidth. This is why voice and music sound muffled over the telephone.

Many voice-coding software products focus on telephone bandwidth audio compression (~4 KHz). Some multimedia and teleconferencing applications provide higher voice quality than the telephone. This is accomplished by offering 7 KHz of audio bandwidth. This adds presence to the signal because speech naturally utilizes approximately 7 KHz of bandwidth. The audio and voice signals processed by telephone and speech bandwidth compression algorithms are the same signals found on wider channels, but with the higher frequencies filtered out.

Compression is achieved by removing predictable, redundant, or pre-determined information from a signal. The compression function is called encoding and the decompression function is called decoding. After



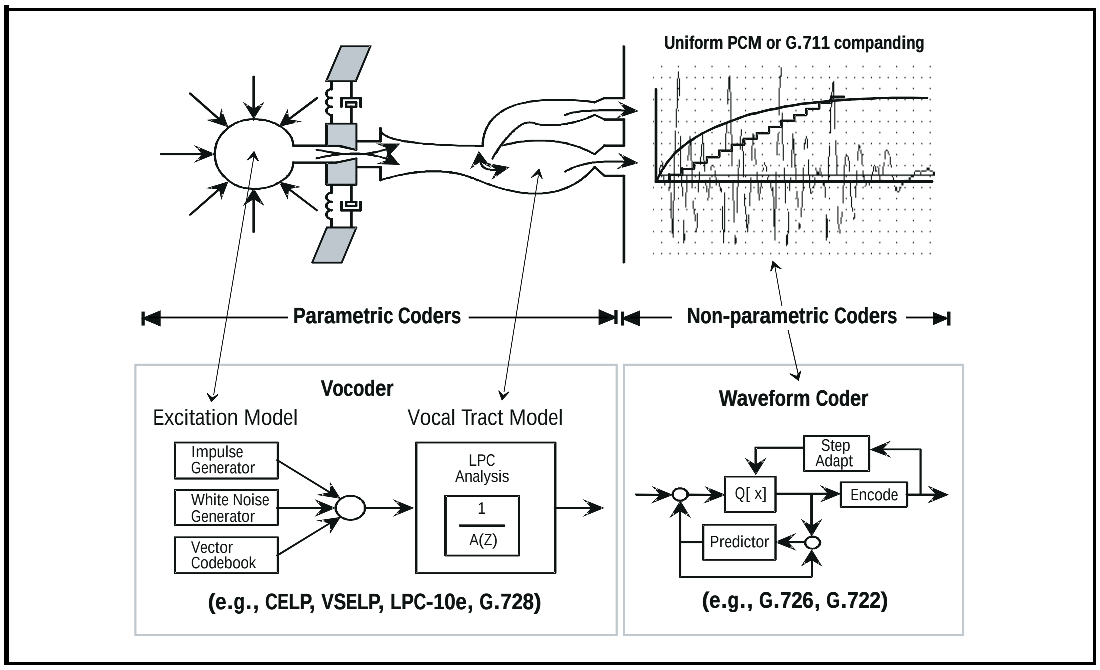
encoding, the compressed information is transmitted to a receiver or to a storage device. The signal is reconstructed by the decoder at the receiver or playback station. A compression algorithm may lose some of the original information in this process, which may be perceptible by the listener. Such algorithms perform what is called "lossy" compression. If a compression algorithm loses no information, providing perfect reconstruction, it is said to be "lossless." Lossy compression algorithms achieve higher compression than lossless algorithms.

Audio Versus Voice Compression

Audio compression is typically performed by non-parametric compression algorithms called waveform coders. Waveform coders are used to compress telephone, speech, and music bandwidth signals. Voice compression is typically performed by parametric compression algorithms called voice coders. Voice coders are used to compress telephone bandwidth speech signals. Waveform coders (non-parametric) are designed to efficiently quantize any audio waveform by quantizing only new and unpredictable information in the signal. Such coders are robust for diverse signal sets, but do not accomplish high compression. Voice coders (parametric) are designed to encode a human voice and are also called vocoders. Vocoders use a prior knowledge of the signal's source to achieve higher compression than that achieved by waveform coders.

Vocoders map speech signals onto a mathematical model of the human vocal tract. Instead of transmitting efficiently-quantized speech samples, voice encoders transmit model parameters. The decoder applies the received parameters to an identical mathematical model and generates an imitation of the original speech. The process of determining model parameters is called analysis and the process of generating speech from the chosen parameters is called synthesis.

Vocoder sound quality varies greatly with the input signal because vocoders are based upon a vocal tract model. Signals from sources that do not fit the model may be coded poorly, resulting in lower quality signal reproduction after decompression. For this reason, waveform coders are sometimes chosen over vocoders for applications that process signals from a diverse set of sources.



When selecting an audio / voice compression algorithm there are six key criteria to consider; not all criteria apply to all situations. The six key criteria are: interoperability, digital bandwidth, speech quality, signal set diversity, delay, and hardware resources.

Interoperability

DSPSE offers both standard and proprietary audio / voice compression algorithms. In order to allow telecommunications equipment to interoperate, standard algorithms are defined. These algorithms specify a set of operations that must be performed to process a signal in accordance with the standard. For example, the USFS 1016 CELP standard allows different manufacturers' secure phones to communicate with each other. When interoperability with other communication systems is not required, as is the case for voice mail, proprietary algorithms are often preferred. Proprietary algorithms are customized to provide higher quality solutions for a given application. If a standard algorithm is required for inter-operability, then no further analysis is required to make a selection.

Digital Bandwidth

Most applications require a particular digital communication / storage bandwidth (bit rate). Therefore, the most obvious criterion for algorithm selection is the algorithm's bit rate. DSPSE offers audio / voice compression algorithms for a wide spectrum of digital bit rates, 2.4 to 64 kbps.

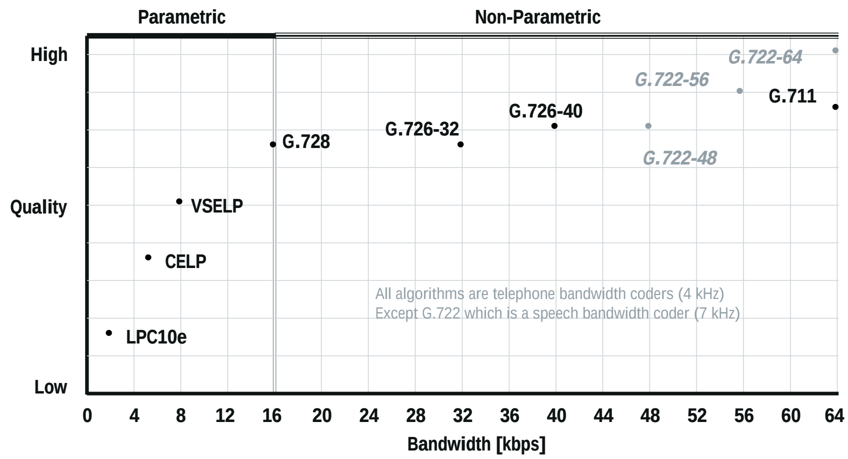
Speech Quality

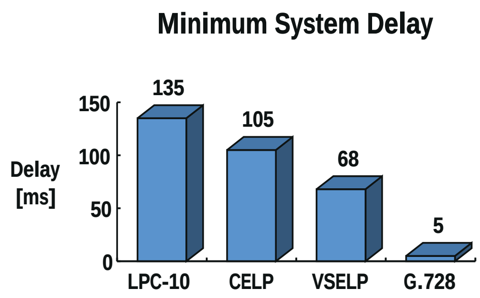
Speech quality is another very important criterion for algorithm selection. All of the algorithms are lossy, so speech quality generally degrades as the bit rate decreases. The analog bandwidth supported by an audio or voice coder also directly affects its speech quality (i.e., telephone, speech, or music bandwidth). The quality requirements vary as some applications require natural sounding speech, some require that the speaker be identifiable and others only require intelligible speech. An algorithm's speech quality is a function of its bit rate and its mathematical approach.

Algorithm quality and digital bandwidth are not linearly related. Algorithms that operate at twice a given bandwidth do not necessarily provide twice the quality. Parametric algorithms generally achieve much higher quality per bit rate than non-parametric.

Signal Set Diversity

Applications that communicate or store only voice are ideal for parametric voice-compression algorithms. Applications that communicate or store a variety of signals such as multiple speakers, music, city, or industrial noises may require non-parametric voice-compression algorithms. Some parametric algorithms, however, do perform significantly better than others on signals that are not generated by a human vocal tract. As a result, a parametric algorithm





may be used for some diverse signal applications providing a lower bit rate than would be achieved otherwise.

Delay

Delay is very important in a digital communication system and is of no concern for a digital storage system. The algorithmic delay is defined as the amount of time it takes for a signal sample to travel through the coder (compression and decompression). The processing delay is the algorithmic delay plus the algorithm's processing block size in time. Users can be irritated by end-to-end delays in an audio communication system as conversations become unnatural for such delays over a quarter of a second. Delays also affect multimedia systems that require synchronization between audio and video. Audio and voice compression algorithms have inherent delay associated with them. Non-parametric algorithms have very low delay and parametric algorithms generally have higher delay. Some parametric voice coders are specifically designed for low delay and these should be selected for delay-sensitive applications.

The end-to-end delay of a system using a vocoder is the sum of the algorithmic delay and approximately 3 times the processing block size (this multiplier is system dependent and ranges from 3 to 5). For example, the USFS 1016 CELP algorithm has an algorithmic delay of 15 ms and a processing block size of 30 ms. A system's end-to-end delay using CELP is 15 ms plus 3 to 5 times 30 ms, for a total of 105 ms to 165 ms. However, the end-to-end delay of a system using ITU G.728 LD-CELP is only 5 ms to 6 ms, hence the name "low-delay" CELP.

Hardware Resources

Hardware resources are key criteria for selecting an algorithm. A product's market typically determines the product's price. The product price determines the amount of MIPs, RAM, and ROM that can be used by the product's audio subsystem. Algorithm resource requirements vary drastically. Non-parametric algorithms are less complex than parametric and generally require fewer MIPs, RAM, and ROM. Some parametric algorithms are specifically optimized for minimal memory or MIPs utilization. The ideal algorithm for bit rate and quality may be impractical due to resource requirements. Products that require standard algorithms must be able to compete in the market given the costs associated with that algorithm's resource requirements.

Contact: DSP Software Engineering, Inc.
165 Middlesex Turnpike
Bedford, MA 01730
(617) 275-3733
Fax: (617) 275-4323
E-mail: sales@dspse.com