Speech Compression Field Overview and Latest Standards

Introduction:

Speech compression has been very important since the development of digital communication, always striving for better bandwidth and quality for audio calls. Since then speech compression plays a very important role in allowing applications such as Mobile Telephony and VoIP to work as we know them. This project aims to give an introduction to the field and analyze the state of the art of speech compression. Also, an implementation of speech compression with focus on speech recognition using latest algorithms is to be developed.

Speech could be compressed using the standard audio compression techniques available, however we want to achieve better compression ratios given the importance of the applications. For that end, speech compression focuses on encoding only information that is relevant for the human ear (not all sounds present in the recording). Since speech is a simpler signal than most other audio signals and because we know the properties of speech and how is produced, we can create stricter encoders/decoders to model speech.

The next section describes what are the known characteristics of speech used in compression systems, Section 3 introduces the types of speech compression and finally Section 4 presents the speech compression standards used today.

Basic Concepts

Speech is digitalized by the process of sampling, quantization and coding. Sampling is relatively simple, usually based on the Nyquist frequency of the voice signal to be sampled to allow the recovery of the analog signal. For example, for narrow-voice band (4 kHz) the sampling frequency would be 8 kHz. According to [1] speech amplitude is not evenly distributed, uniform quantization creates high levels of noise. Therefor, non-inform quantization is usually used for speech and with finer quantization applied to low speech signals. Coding is the process of representing the sampled and quantized values with bits, ideally using the minimum amount possible. We'll expand on this topic in Section 2 and 3.

Digitalized speech signals have some characteristics that are explored during coding and they occur due to how human speech is produced. The main characteristic used in coding is the fact that certain sounds produce quasi-periodic patterns in the signal (voiced sounds) due to how the vocal cords vibrate during the speech process. Other sounds that don't depend on the vibration of vocal cords produce a signal that resembles noise. These differences are used in frame-based speech coding.

The next section explains the main types of compression techniques, some of which use the characteristics described above to enhance compression.

Compression Techniques

There are three basic types of speech encoding. First we'll look into Waveform compression which is a simple approach with low compression rate. Secondly we'll analyze Parametric-based compression that uses knowledge of how speech is produced to generate a set of parameters to represent each speech segment. The decoder in turn would be able to reconstruct the speech based on these parameters. Finally the last type called Hybrid compression aims to unite the best parts of each method and has proven to generate very effective speech compression.

Waveform Compression

Waveform compression is based on Pulse Code Modulation (PCM) [3] which digitizes analog signals. Compression in PCM happens by quantization of the amplitude of a sample. Research has shown that lower audio signals have more information about speech than higher audio signals [1], and because of that PCM uses non-uniform quantization, with focus on low signals.

An improvement of PCM is called Adaptive differential Pulse Code Modulation (ADPCM) which further compresses the output of PCM, it does that by incorporating a predictor component and a adaptive quantizer in the encoder. The predictor generates an estimate of the next speech segment based on the previous speech segment, then only the prediction error is passed to the quantizer and coded. Since the quantized prediction error is smaller than the original PCM then ADPCM uses less bits to represent a speech segment.

PCM and ADPCM were initially designed for telephony but these techniques are also used in voice over IP communications.

Parametric Compression

Hybrid Compression

Compression Standards

1. Narrowband, wideband, super-wideband and fullband
2. Standards

Conclusion

References

1. Lingfen SunIs-Haka MkwawaEmmanuel JammehEmmanuel Ifeachor. Guide to Voice and Video over IP. Springer-Verlag London 2013.

2. A. Gersho. Advances in speech and audio compression. Proceedings of the IEEE (Volume: 82, Issue: 6, Jun 1994).

3. Y.Yatsuzuka, Highly Sensitive Speech Detector and High-speed Voiceband Data Discriminator in DSI-ADPCM Systems," IEEE Trans Commun., vol. COM-30, pp. 739-750, Apr. 1982.

4.