International Journal on Recent and Innovation Trends in Computing and Communication
Volume: 5 Issue: 5

ISSN: 2321-8169
637 – 640

Speech Compression Techniques: An Overview

Juhi singh
Department of computer science, Amity University, Haryana

Abstract: Speech is the natural phenomena of human for communication purpose. The aim of speech coding is to compress the speech signal to the highest possible compression ratio but maintaining user acceptability. In this paper basically two major approaches for speech compression techniques are discussed like waveform coder: pulse code modulation, adaptive differential pulse code modulation, Sub-band coding, transform coding and vocoder: linear predictive coder, formant coder/synthesis.

Keywords*: speech compression, voiced and unvoiced speech, time and frequency domain*

*****

I. INTRODUCTION

The aim of speech compression is to reduce the number of bits required to represent speech signals by removing the redundant bits so-that the less bandwidth is required for transmission. Before discussing the speech compression coding techniques, it is important to understand the digitization process.The speech signal is represented in its digital form, that is, the process of speech signal digitization. There are basically two the key features of speech signal are voiced and unvoiced speech and their characteristics. In broader terms, speech compression techniques are mainly focused on removing short-term correlation (in the order of 1ms) among speech samples and long-term correlation (in the order of 5 to 10 ms) among repeated pitch patterns. In this section, we will start with speech signal digitization and then discuss speech signal features and speech compression techniques.

II. SPEECH SIGNAL DIGITIZATION

Speech signal digitization is the process to convert speech from analog signal to digital signal in order for digital processing and transmission. The main phases in speech signal digitization are shown in fig 1 a) sampling, and in fig 1b) quantization and coding.

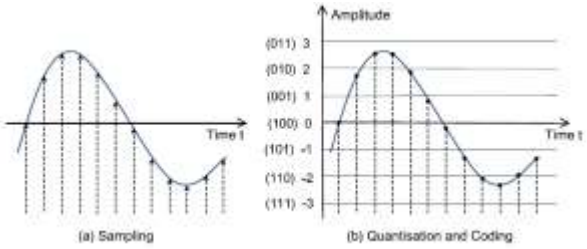

Fig1. Example of voice digitization

III. HUMAN SPEECH PRODUCTION

The production of speech is a natural phenomenon of human being by inhaling the air through mouth. In fig 2, a conceptual diagram of human speech production physical model. When we speak, the air from lungs push through the vocal tract and out of the mouth to produce a sound. Speech compression, especially at low bit rate speech compression, explores the nature of human speech production mechanism. In this section, we briefly explain how human speech is produced.

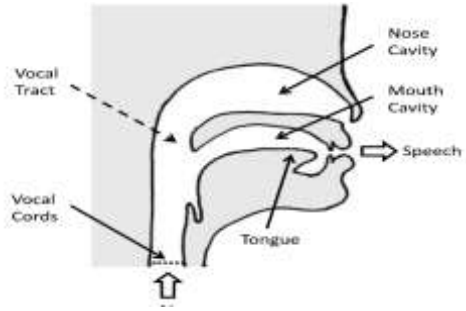

Fig 2: Conceptual diagram of human speech production[1]

a) **Voiced Sound:** For some sounds for example, a voiced sound, or vowel sounds of 'a', 'i' and 'μ', as, the vocal cords vibrate (open and close) at a rate (fundamental frequency or pitch frequency) and the produced speech samples show a quasi-periodic pattern.

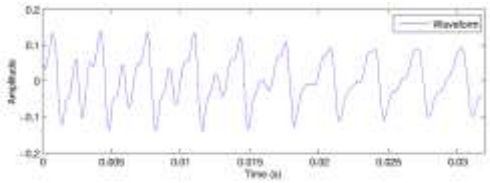

Fig 3a: Sample of voiced speech—waveform[1]

b) **Unvoiced Sound** :For other sounds (e.g., certain fricatives as 's' and 'f', and plosives as 'p', 't' and 'k' , named as unvoiced sound, the vocal cords do not vibrate and remain open during the sound production.

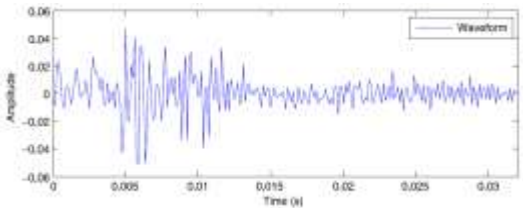

Fig 3a: Sample of Unvoiced speech—waveform[1]

Note: The waveform of unvoiced sound is more like noise.

IV. SPEECH COMPRESSION TECHNIQUES

There are two types of speech compression techniques as follows:

1) Waveform Coders: waveform coders attempt[5], without using any knowledge of how the signal to be coded was generated, to produce a reconstructed signal whose waveform is as close as possible to the original. This means that in theory they should be signal independent and work well with non-speech signals. Generally they are low complexity coders which produce high quality speech at rates above about 16 Kbits/s. When the data rate is lowered below this level the reconstructed speech quality that can be obtained degrades rapidly.

i) Time Domain[2]: Time compressed speech refers to an audio recording of verbal text in which the text is presented in a much shorter time interval than it would through normally paced real time speech.[6] The basic purpose is to make recorded speech contain more words in a given time, yet still be understandable. For example: a paragraph that might normally be expected to take 20 seconds to read, might instead be presented in 15seconds, which would represent a time compression of 25% (5 seconds out of 20).

a) PCM: Pulse-code modulation (PCM) is a method used to digitally represent sampled analog signals. It is the standard form of digital audio in computers, compact discs, digital telephony and other digital audio applications. In a PCM stream, the amplitude of the analog signal is sampled regularly at uniform intervals, and each sample is quantized to the nearest value within a range of digital steps. Some forms of PCM combine signal processing with coding. Older versions of these systems applied the processing in the analog domain as part of the analog-to-digital process; newer implementations do so in the digital domain. These simple techniques have been largely rendered obsolete by modern transform-based audio compression techniques.

Limitation of PCM

- There are potential sources of impairment implicit in any PCM system:
- Choosing a discrete value that is near but not exactly at the analog signal level for each sample leads to quantization error.
- Between samples no measurement of the signal is made; the sampling theorem guarantees non-ambiguous representation and recovery of the signal only if it has no energy at frequency $f_s/2$ or higher (one half the sampling frequency, known as the Nyquist frequency); higher frequencies will generally not be correctly represented or recovered.
- As samples are dependent on *time*, an accurate clock is required for accurate reproduction. If either the encoding or decoding clock is not stable, its frequency drift will directly affect the output quality of the device.[note 3]

b)ADPCM: Adaptive differential pulse-code modulation (ADPCM) is a variant of differential pulse-code modulation (DPCM) that varies the size of the quantization step, to allow further reduction of the required data bandwidth for a given signal-to-noise ratio. [4]Adaptive differential pulse code modulation (ADPCM) is a method used to convert analog signals to binary signals. The technique converts the analog signals by taking frequent samples of the sound and representing the value of the sampled modulation in binary form. The concept of ADPCM is to use the past behavior of a signal to forecast it in the future. The resulting signal will represent the error of the prediction, which has no significance. Therefore, the signal must be decoded to rebuild a more meaningful original waveform. The ADPCM technique is employed to send sound signals through fiber-optic long-distance lines. This is useful especially for organizations that set up digital lines between remote sites to broadcast both voice and data. The voice signals are digitized before they are broadcasted. The technique is a variation of the digitized method known as pulse code modulation.

ii) Frequency Domain: In frequency domain speech compression technique, to compress the speech mainly two techniques are being used as follows

a)Sub-band Coding: In signal processing, sub-band coding (SBC) is any form of transform coding that breaks a signal into a number of different frequency bands, typically by using a fast Fourier transform, and encodes each one independently. This decomposition is often the first step in data compression for audio and video signals.[3] The process of breaking the input speech into sub-bands via band-pass filters and coding each band separately is called sub-band coding. To keep the number of samples to be coded at a minimum, the sampling rate for the signals in each band is reduced by decimation. Of course, since the band-pass filters are not ideal, there is some overlap between adjacent bands and aliasing occurs during decimation. Ignoring the distortion or noise due to

compression, Quadrature mirror filter (QMF) banks allow the aliasing that occurs during filtering and sub-sampling at the encoder to be cancelled at the decoder. The coders used in each band can be PCM, ADPCM, or even an analysis-by-synthesis method. The advantage of sub-band coding is that each band can be coded differently and that the coding error in each band can be controlled in relation to human perceptual characteristics.

b)Adaptive Transform Coding: Adaptive transform coding is a promising technique for speech coding at low to medium bit rates. There remains however, much more work to be done in this field. The effect of channel errors was not considered in this thesis. This is another topic which could be investigated. The optimum quantization strategy for side information, whether all-pole or homomorphism, has not been formulated in an 5-4 adaptive transform coding context. Other forms of short term spectrum parameterization are also worthy of study. In conclusion, future speech digitization developments can not afford to ignore the successes of adaptive transform coding techniques.

2) Vocoders[2]: A vocoder is a category of voice codec that analyzes and synthesizes the human voice signal for audio data compression, multiplexing, voice encryption, voice transformation, etc. Basically vocoder was designed to reduce the channel bandwidth in telecommunication. In the channel vocoder algorithm, among the two components of an analytic signal, considering only the amplitude component and simply ignoring the phase component tends to result in an unclear voice; on methods for rectifying this, see phase vocoder. In the encoder, the input is passed through a multiband filter, then each band is passed through an envelope follower, and the control signals from the envelope followers are transmitted to the decoder. The decoder applies these (amplitude) control signals to corresponding filters for re-synthesis. Since these control signals change only slowly compared to the original speech waveform, the bandwidth required to transmit speech can be reduced. This allows more speech channels to share a single communication channel, such as a radio channel or a submarine cable (i.e. multiplexing).

i) Linear Predictive Coders: Linear predictive coding (LPC) is a tool used mostly in audio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model. It is one of the most powerful speech analysis techniques, and one of the most useful methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters. LPC starts with the assumption that a speech signal is produced by a buzzer at the end of a tube (voiced sounds), with occasional added hissing and popping sounds (sibilants and plosive sounds). Although apparently crude, this model is actually a close approximation of the reality of speech production. The glottis (the space between the vocal folds) produces the buzz, which is characterized by its intensity (loudness) and frequency (pitch). The vocal tract (the throat and mouth) forms the tube, which is characterized by its resonances, which give rise to formants, or enhanced frequency bands in the sound produced. Hisses and pops are generated by the action of the tongue, lips and throat during sibilants and plosives. LPC analyzes the speech signal by estimating the formants, removing their effects from the speech signal, and estimating the intensity and frequency of the remaining buzz. The process of removing the formants is called inverse filtering, and the remaining signal after the subtraction of the filtered modeled signal is called the residue. The numbers which describe the intensity and frequency of the buzz, the formants, and the residue signal, can be stored or transmitted somewhere else. LPC synthesizes the speech signal by reversing the process: use the buzz parameters and the residue to create a source signal, use the formants to create a filter (which represents the tube), and run the source through the filter, resulting in speech. Because speech signals vary with time, this process is done on short chunks of the speech signal, which are called frames; generally 30 to 50 frames per second give intelligible speech with good compression.

iiFormant synthesis: Formant synthesis does not use human speech samples at runtime. Instead, the synthesized speech output is created using additive synthesis and an acoustic model (physical modelling synthesis). Parameters such as fundamental frequency, voicing, and noise levels are varied over time to create a waveform of artificial speech. This method is sometimes called *rules-based synthesis*; however, many concatenative systems also have rules-based components. Many systems based on formant synthesis technology generate artificial, robotic-sounding speech that would never be mistaken for human speech. However, maximum naturalness is not always the goal of a speech synthesis system, and formant synthesis systems have advantages over concatenative systems. Formant-synthesized speech can be reliably intelligible, even at very high speeds, avoiding the acoustic glitches that commonly plague concatenative systems. High-speed synthesized speech is used by the visually impaired to quickly navigate computers using a screen reader. Formant synthesizers are usually smaller programs than concatenative systems because they do not have a database of speech samples. They can therefore be used in embedded systems, where memory and microprocessor power are especially limited. Because formant-based systems have complete control of all aspects of the output speech, a wide variety of prosodies and intonations can be output, conveying not just questions and statements, but a variety of emotions and tones of voice.

V. CONCLUSION

As the data grows day by day, the short and compression communication is required, Speech compression techniques plays a vital role in it. In this paper all the possible speech

compression techniques are discussed that can be used to compress the data before transmission of data so that it can consume less bandwidth.

REFERENCES

[1] //www.springer.com/cda/content/document/cda_downloaddocument/9781447149040-c2.pdf?SGWID=0-0-45-1369003-p174740098.

[2] B.Juang, Hyun Bae, ECE 8873 Data Compression & Modeling, Georgia Institute of Technology , 2004

[3] Jerry D. Gibson ,"Speech Coding Methods, Standards, and Applications",, Department of Electrical & Computer Engineering University of California, Santa Barbara, CA 93106-6065

[4] https://www.techopedia.com/definition/5877/adaptive-differential-pulse-code-modulation-adpcm

[5] http://www-mobile.ecs.soton.ac.uk/speech_codecs/waveform.html

[6] Time Compressed Speech definition. (http://psychologydictionary.org/timecompressedspeech/).