Effectiveness of Linear Predictive Coding in Telephony based applications of Speech Recognition

Jagdev Singh Kaleka
Senior Lecturer, Govt. Polytechnic for Women,
Deptt. Of Technical Education and Industrial Training, Govt. of Punjab

Introduction

Real time speech recognition is an active area of speech research. Human perception of speech is, in part, reliant on detection of formant frequencies and their transitions times. Subjects utilize formant transition length to categorize sounds. Speech recognition systems can be classified according to the type of speech, the size of the vocabulary, the basic units and the speaker dependence. The position of a speech-recognition system in these dimensions determines which algorithm can or has to be used. Speaker independence is the typical aspect of a speech-recognition application where the rate of success is independent of the user.
The demand for telecommunications applications of automatic speech recognition [1] has exploded in recent years. This area seems a natural candidate for speech recognition systems, since it embraces a tremendous variety of applications that rely entirely on audio signals and serial interfaces. However, the telecommunication environment strains the capabilities of current technology, given its broad range of uncontrollable variables, from speaker characteristics to telephone handsets and line quality.

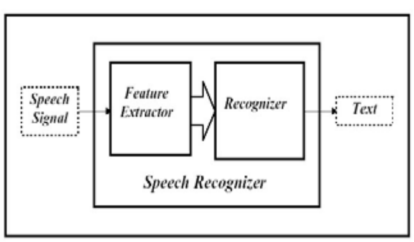

Fig1: Speech Recognition Process

Speech coding refers to the process of reducing the bit rate of digital speech representations for transmission or storage, while maintaining a speech quality that is acceptable for the application. LPC (Linear Predictive Coding) is a well-known technique for speech analysis-synthesis at low bit rates. The analysis consists in finding a time-based series of n-pole IIR filters whose coefficients better adapt to the formants of a speech signal. These computations produce a residual signal that is the exact complement to the information kept in the coefficients, if we wish to recover the original. The model of the human vocal tract mechanism assumed by LPC presupposes that speech can be reduced to a succession of voiced or unvoiced sounds.
Speech recognition research and development has several goals [2]. Simplifying the interface between user and machine is one major goal. Machine speech recognition and understanding has the potential to greatly simplify the way people work with machines.

Speech Recognition System

The architecture of a speech recognition [3-4] engine can be broken down into a feature extraction module, a pattern-matching algorithm, and a hypothesis block. The feature extraction block transforms the input speech into a set of spectral components. The pattern-matching block uses the spectral patterns and compares them with some known patterns.

Feature extraction is the process by which a small amount of data is taken from the voice signal. This data should vary significantly from speaker to speaker and not be affected by noise or communication channel variation. The standard method for deriving this data is *linear predictive coding*, or LPC, which models the vocal tract with a finite-order all-pole transfer function. The coefficients of this model accurately indicate the instantaneous configuration of the vocal tract. These can then be used to identify the type of sound being produced, for speech recognition, or in this case, to identify the speaker based on small variations in the parameters.

Speech quality as produced by a speech coder is a function of bit rate, complexity, delay and bandwidth. Thus when considering speech coders it is important to review all these attributes. LPC makes coding at low bit rates possible.

Speech Recognition Techniques

The complexity and the success rate probability of any Speech Recognition System [2] depends on number of speakers, vocabulary size, language complexity, and environment conditions.

There are three major types of speech recognition techniques. First, the ***acoustic-phonetic approach*** assumes that the phonetic units are broadly characterized by a set of features, such as formant frequency, voiced/unvoiced, and pitch. These features are extracted from the speech signal and are used to segment and label the speech.

Second, the ***pattern recognition approach*** requires no explicit knowledge of speech. This approach has two steps - namely, training of speech patterns based on some generic spectral parameter set and recognition of patterns via pattern comparison. The popular pattern recognition techniques include template matching, Hidden Markov Modeling, and artificial neural network (ANN).

Third, the ***artificial intelligence approach*** attempts to mechanize the recognition procedure according to the way a person applies its intelligence in visualizing, analyzing, and finally making a decision on the measured acoustic features.

Applications

Speech recognition has proven its value in a number of industries from travel and banking to telecommunications and automobiles. Natural speech now acts as a substitute for pointing, clicking, and keyboard entries. Speech Recognition simplifies the interface between the user and the machine and allows hands free and quick voice access to information. Voice Recognition over telephone is one of the major applications of the technology. Some telephony-based applications include:

- Home banking applications
- Voice requests
- Shopping programs
- Directory assistance
- Voice activated calling
- Call routing
- Queries on voice prompts
- Advanced-messaging systems will forward voice calls, email, and faxes after receiving voice commands.

Linear Predictive Coding

Speech compression is required in long-distance communication, high-quality speech storage, and message encryption. Utilizing speech compression makes it possible for more users to share the available system. Linear Predictive Coding [5-11] is one possible technique of analyzing and synthesizing human speech. This method is used to successfully estimate basic speech parameters like pitch, formants and spectra.

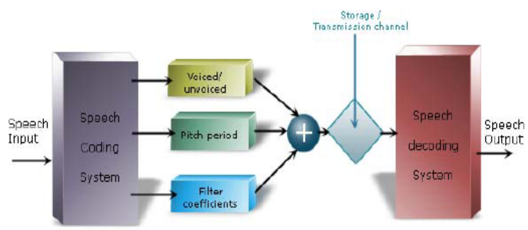

Fig2 : LPC Speech Analysis and Synthesis

The values of predictor coefficients that minimize error/residual signal are found by assigning the partial derivatives of signal with respect to coefficients to zeros. Thus, we get *p* equations with *p* unknown variables (p=number of coefficients). The Levinson-Durbin algorithm solves the Toeplitz matrix.

The principle behind the use of LPC [6,7] is to minimize the sum of the squared differences between the original speech signal and the estimated speech signal over a finite duration. This could be used to give a unique set of predictor coefficients. These predictor coefficients are normally estimated every frame, which is normally 20 ms long. The summation is computed starting at k=1 up to p, which will be 10 for the LPC-10 algorithm.

The LPC analysis [8,9] of each frame also involves the decision-making process of concluding if a sound is voiced or unvoiced. If a sound is decided to be voiced, an impulse train is used to represent it, with nonzero taps occurring every pitch period. A pitch-detecting algorithm is employed to determine to correct pitch period / frequency. We used the autocorrelation function to estimate the pitch period. However, if the frame is unvoiced, then white noise is used to represent it and a pitch period of T=0 is transmitted. Therefore, either white noise or impulse train becomes the excitation of the LPC synthesis filter. It is important to re-emphasize that the pitch, gain and coefficient parameters will be varying with time from one frame to another.

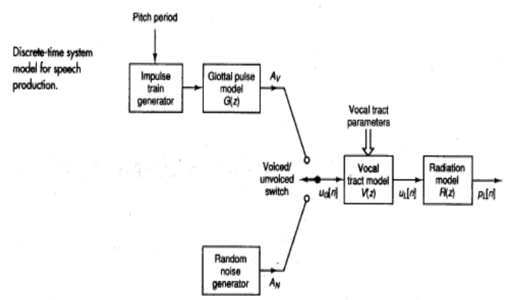

Fig3 : Speech Production

The number of bits transmitted every frame is given by following

1 bit	voiced/unvoiced
6 bits	pitch period (60 values)
10 bits	k_1 and k_2 (5 each)
10 bits	k_3 and k_4 (5 each)
16 bits	k_5, k_6, k_7, k_8 (4 each)
3 bits	k_9
2 bits	k_{10}
5 bits	gain G
1 bit	synchronization
54 bits	TOTAL BITS PER FRAME

Using a segment size of 54 bits/segment and sampling frequency of 8kHz, the segment rate comes out be approx. 44.44 segments/sec if we take 180 samples/sec. The bit rate, which is given by segment size times the segment rate equals 2400 bits/sec which is a sufficient bit rate for telephone and communication based applications.

Thus, Linear Predictive Coding [10] is the most succesful and efficiently used technique for speech coding. It takes into account the redundancy in the speech signal and thus enables speech coding at low bit rate possible.

Conclusion

The most exciting, promising, and probably most cost-effective technology for the future and maybe even for the present is speech recognition, a technology that is on the leading edge of technological development today. Multimedia refers to having a variety of media presented either simultaneously or sequentially. Thus, speech coding for multimedia automatically implies that the speech coding bit-stream will be sharing the communication channel with other signals.

Significant advances have been made in the area of speech coding over the last 15 years and speech coding algorithms are now available which can produce communication quality speech at a bit-rate as low as 2.4 kbits/s [11]. These advances combined with current DSP hardware technology have made it possible to utilize speech coding in telecommunication applications which includes directory assistance, call routing, messaging, and digit/name dialing.

Reference

[1] Claudio Becchetti, "Speech Recognition Systems", Wiley, April 1999.

[2] LR Rabiner, MR Sambur, "Speaker Independent Recognition of Connected Digits", IEEE International Conference on Acoustics, Speech, and Signal Processing, April 1976.

[3] L. R. Rabiner, R. W. Schafer, "Digital processing of speech signals" Prentice-Hall, Sept. 1978.

[4] L. Rabiner and B. H. Juang, "Fundamentals of Speech Recognition", Prentice Hall, April 1993.

[5] Wai C. Chu, "Speech Coding Algorithms", Wiley, April 2003.

[6] Randy Goldberg, Lance Riek, "A Practical Handbook of Speech Coders", CRC Press, 2000.

[7] Jason Woodard, "Speech Coding", www-mobile.ecs.soton.ac.uk/speech_codecs/index.html.

[8] O. M. Ibarra, F. Curatelli, "A Brief Introduction to Speech Analysis and Recognition", May 2000, www.mor.itesm.mx.

[9] "Speech Coding with Linear Predictive Coding", www.dspexperts.com/projects.

[10] Richard V. Cox, Peter Kroon, "Low Bit-Rate Speech Coders for Multimedia Communication", IEEE Communication Magazine, Dec 1996.

[11] Christian Plessl, Simon Maurer, "Hardware/Software Codesign in speech compression applications", Feb 2000, www.tik.ee.ethz.ch.