IJETST- Volume||01||Issue||06||Pages 945-948||August||ISSN 2348-9480 2014

Open access Journal International Journal of Emerging Trends in Science and Technology

Speech Parameters Characterization Using Data Mining Techniques

Authors

Kavita, Ranjeeta Kaushik

Dept of CSE

NCCE

Israna, Panipat

Email: *kttdoll@gmail.com* , *er.ranjeeta@gmail.com*

ABSTRACT- *Speech has many parameters such as pitch, energy, Noise, Change in speaking rate, Change in articulation. Speech is also of two type i.e. voiced speech and unvoiced speech. Voiced speech has high frequency as compared to the unvoiced speech. Many algorithms can be used to extract features of the speech. Here we use LPC (Linear predictive coding) method to extract the features of speech. The aim of this report is to show the difference between the feature of the child speech and adult speech. We also synthesize the speech by removing the unvoiced speech from the original speech by removing the low frequency speech from high frequency speech. This research is focused on implementing and studying Linear Predictive Coding (LPC) algorithm.*

INTRODUCTION

The studies of how human speech sound are produced and how they can be used in the form of a language in an establish scientific discipline. There is a lot of research work has been going on in the industry and in educational institutes on how efficiently can we encode and transmit speech signal from the transmitter and then retrieve it back to the receiver with high accuracy.[1]

Human speech produced due to many factor such as pitch, noise, energy etc. due to these factor we can classify them as voiced speech and unvoiced speech.

BASIC PHYSICAL PRINCIPLES OF AUDIO

Speech has been produced when the air from within the lungs are expelled into the trachea, through the vocal tract, and then forced out through the mouth to generate speech. Here, lungs can be act as sound generator/source and the vocal tract act as a filter which produce the various type of the sounds that made up the speech.[1]

FUNDAMENTAL COMPONENT OF SPEECH

The vocal tract consists of the nose, the throat, the tongue,and the mouth. It is important to know some key definitions to understand how the vocal tract or vocal folds turn air from lungs into sound.

Phoneme:

A smallest unit in speech where substitution of one unit for another might makes a distinction of meaning. For example, in English the words “to” and “do” differ in the initial phoneme; and “dole” and “doll” differ in the middle [2].

There are two types of phonemes: voiced and unvoiced sounds, these sounds are considered in analyzing and synthesizing speech signal by linear predictive coding technique.

Voiced Sound:

Voiced sounds are usually vowels that usually have high level of average energy and very distinct formant and resonant frequencies.

Voiced sounds generates by Air from within the lungs by forcing the vocal cords. Because of vibration of vocal folds, seemingly periodic patterns in the form of series of air pulses are produced that are called glottal pulses.

Glottal pulses excite a vocal tract cavity and produce a vowel sound.[7]

Unvoiced Sound:

Unvoiced sounds are usually consonants having comparatively less energy at higher frequencies in comparison of the voiced sounds.
The source of unvoiced sound generation is turbulent flow of Air from vocal folds. During all this process, the vocal cords did not vibrate, whereas they will stay open until sound is produced.
In unvoiced speech there is no vibration in the vocal cords, no glottal pulses and, in unvoiced speech pitch is an unimportant attribute because unvoiced sounds are not periodic.

Properties of speech:

- Fricatives (s, sh, f, th) – These are produced when the vocal tract is closed at some location and air has been forced through that constriction or closed.
- Plosives (p, k, t) – These produced when the end of the vocal tract is closed or constricted momentarily while the air pressure built up and then the pressure is suddenly released.
- In English, there are about 40 phonemes (sound elements - 16 vowels, 24 consonants).
- In normal speech, 10 to 15 phonemes are spoken in one second. [7]

Methods used in this explain below:

LPC(linear predictive coding):
Linear Predictive Coding or "LPC" is analytical/synthesis method, introduced in the sixties for predicting a present sample of speech based on several previous samples .
\
LPC is an efficient way of getting synthesized speech signal. Due to the speed of the analysis algorithm and low bandwidth required for the encoded signals this method has good efficiency.[5]

FUNDAMENTAL OF LINEAR PREDICTIVE CODING

There are two ways of measuring/estimating the spectrum/spectral envelope, of a sound.

First way is through Fast Fourier Transform (FFT), which measures the spectrum of a sound by sampling amplitude values at equally spaced frequency points in the given range. It provides an accurate estimation of the spectrum.

Other method is to use Linear Predictive Coding, or LPC. It measures the overall spectral envelope to create a linear image of the sounds' spectrum. Both methods have their strengths, and weaknesses, but LPC method is particularly effective in manipulating the speech.

LPC generally deals with modeling and FFT makes the spectrum estimation.
LPC is one of the useful methods for encoding the good quality speech at low bit rate and one of the most powerful techniques for speech analysis. Accurate estimation of speech parameters has been done, and it is relatively efficient for computation
LPC digitally encodes the analog signals using a

single or multi level sampling system, in which at each sample time the value of the signal is predicted to be a linear function of the past values of the quantized signal.
Linear Predictive Coding relates to adaptive predictive coding (APC) because both use adaptive predictors.

APC uses less prediction coefficients to permit use of a lower information bit rate than LPC, whereas LPC uses more prediction coefficients and thus LPC requires a more complex processor.
When the predictor filter has been adjusted to predict the input, it can do so from the immediate preceding samples. The spectral peaks caused by the resonance of speech production will have to be removed.

The difference between the input speech and the predictor output (known as residual) will have roughly flat spectrum. For the same reason, the complete filtering process is sometimes referred to as inverse filtering.

LPC synthesizes the speech signal by reversing all process: i.e to create a source signal use the residue, formants are used to create an all-pole filter (which represent the tube), and then run the source through filter, which resulting in speech. The speech signal varies with time, so this process is done on the short chunks of the speech signal known as frames.

Normally 30-50 frames per second give intelligible speech with the good compression.[6]

LPC MODEL

5.1 Analysis / Encoding- Which is done by transmitter.

5.2 Synthesis/Decoding- which is done by receiver.

Applications:

LPC is generally used for speech analysis and resynthesis. It is used as a form of voice compression by phone companies, for example in the GSM standard.

It is also used for secure wireless, where voice must be, dizitied, encrypt and sent over a narrow voice channel; an early example of this is the US government's Navajo I.

LPC synthesis can be used to construct vocoders where musical instruments are used as excitation signal to the time-varying filter estimated from a singer's speech.

This is somewhat popular in electronic music. Paul Lansky made the well-known computer music piece not just more idle chatter using linear predictive coding. A 10th-order LPC was used in the popular 1980s Speak & Spell educational toy.

Waveform ROM in some digital sample-based music synthesizers made by Yamaha Corporation may be compressed using the LPC algorithm.

LPC predictors are used in Shorten, MPEG-4 ALS, FLAC, SILK audio codec, and other lossless audio codec's.

LPC is receiving some attention as a tool for use in the tonal analysis of violins and other stringed musical instrument.[10]

CONCLUSION

After thorough study of the LPC algorithms, how Linear Predictive Coding is used as an analysis/synthesis technique has been understood. Linear Predictive Coding is a lossy speech compression technique that attempt to model human production of the sound instead of transmitting a estimate of audio speech signal.

Due to lower bit rate, LPC is ideal for use in secure control systems because secure control systems are more concerned about the content and the meaning of the speech, rather than the quality of the speech.

REFERENCES

1. J.D. Markel and A.H. Gray (Jr.), "Linear Prediction of Speech", Springer-Verlag Berlin Heidelberg New York, New York, 1976, ISBN: 0-387-07563-1, pp 1-15.
2. John Holmes and Wendy Holmes, "Speech Synthesis and Recognition", Second Edition, Taylor & Francis Group, New York, 2001, ISBN: 0-7484-0856-8, Ch. 1,4 and 6.
3. Richard L. Klevans and Robert D. Rodman, "Voice Recognition", Artech House, Inc., MA, 1997, ISBN: 0-890006-927-1, pp 15-31
4. Jerry D. Gibson, Dave Lindbergh ,Toby Berger, Tom Lookabaugh, and Richard L. Baker, "Digital Compression for Multimedia- Principles and Standards", Morgan Kaufmann Publishers Inc., CA, 1998, ISBN: 1-55860-369-7, pp 1-4 & Ch. 6.
5. Uzdy, Z. , "Human speaker recognition performance of LPC voice processors", IEEE Transactions on Acoustics, Speech and Signal processing, Vol. 38, Issue 12, 1998
6. Yedlapalli, S.S., "Transforming Real Linear Prediction Coefficients to Line Spectral Representations With a Real FFT", IEEE Transactions on speech and audio processing, Vol. 13, Issue 6, 2005 f
7. http://www.national.com/ds/LM/LM741.pdf D. Roccheso, "Sound Processing"
8. http://www.faqs.org/docs/sp/sp-96.html (Date accessed: 04/24/2010, 05/02/2010)
9. Jeremy Bradbury, Linear Predictive Coding", Dec. 2000 http://my.fit.edu/~vkepuska/ece5525/lpc_paper.pdf