International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 02 | Feb 2021 www.irjet.net p-ISSN: 2395-0072

ANALYSIS OF CEPSTRAL AND LINEAR PREDICTION IN VARIOUS SPEECH SIGNAL

Vinodh Kumar.M[1], Umamaheshwaran.G[2]

[1]PG Scholar, Applied Electronics, Department of ECE, PMC Tech, Hosur, Tamilnadu, India
[2]Assistant Professor, Department of ECE, PMC Tech, Hosur, Tamilnadu, India

---

Abstract - *In this project analyzing of Voiced, Unvoiced and silence regions of speech from their time domain and frequency domain representation. The deconvolution Cepstral homomorphism processing is frequently used. It relates to the flow of air that comes from lungs during speech generation. For unvoiced sounds like in "s" as well as "f", vocal coeds are relaxed and the glottis is opened. For voiced sounds like, "a", "e". The vocal cord actually vibrates which in turn gives the pitch value. The shaping of the spectrum is done by using filter. This helps in creating different sounds and relates towards vocal tract organs. A speech recognition system tries to reduce the influence on the source (the system must provide same "answer" to get a high pitch woman's voice and for to secure a low pitch males voice), and specify the filter. The problem in source e(n) along with filter impulse answer h(n) are convoluted. Mel-scale is used for auditory modeling. Two famous tests that generated the bark and Mel scale machines were taken from the literature and its outcomes are reported here to characterize human auditory system. The specific LPC is approximately such as sufficient statistics of haphazard samples throughout statics. The LPC coefficients are actually the very least squares estimators from the regression coefficient if the minimum variance linear estimators from the regression coefficients. In LP research the redundancy within the speech signal is used.*

Key Words: Mel-scale, Linear Prediction, Cepstrum, Liftering, Voiced speech, unvoiced speech, Silence speech,

1. INTRODUCTION

Speech is an acoustic signal produced from a speech production system. From our understanding of signals and systems, the system characteristic depends on the design of the system. For the case of linear time invariant system, this is completely characterized in terms its impulse response. However, the nature of response depends on the type of input excitation to the system. For instance, we have impulse response, step response, sinusoidal response and so on for a given system. Each of these output responses are used to understand the behavior of the system under different conditions. A similar phenomenon happens in the production of speech also. Based on the input excitation phenomenon, the speech production can be broadly categorized into three activities. The first case where the input excitation is nearly periodic in nature, the second case where the input excitation is random noise-like in nature and third case where there is no excitation to the system. Accordingly, the speech signal can be broadly categorized into three regions.

1.1 Voiced Speech

If the input excitation is nearly periodic impulse sequence, then the corresponding speech looks visually nearly periodic and is termed as voiced speech. During the production of voiced speech, the air exhaling out of lungs through the trachea is interrupted periodically by the vibrating vocal folds. Due to this, the glottal wave is generated that excites the speech production system resulting in the voiced speech.

1.2 Unvoiced Speech

If the excitation is random noise-like, then the resulting speech will also be random noise-like without any periodic nature and is termed as Unvoiced Speech During the production of unvoiced speech, the air exhaling out of lungs through the trachea is not interrupted by the vibrating vocal folds. However, starting from glottis, somewhere along the length of vocal tract, total or partial closure occurs which results in obstructing air flow completely or narrowly.

1.3 Silence Region

The speech production process involves generating Voiced and unvoiced speech in succession, separated by is called Silence region.

2. RELATED WORK

The input excitation, then the system component can be separated/ constructed by exciting the system with the inputs and collecting its responses. This is what is done in same channel estimation problems. In the second case, if we knew the system response, then the input excitation can be recovered the inverse filter theory concept. For instance, Linear Prediction (LP) analysis of speech to recover excitation. There is yet another type of deconvolution, where the assumption is both input excitations as well as system responses are unknown. Speech is composed of excitation source and vocal tract system components. In order to analyze and model the excitation and system components of the speech

Independently and also use that in various speech processing applications, these two components have to be separated from the speech. The objective of cepstral analysis is to separate the speech into its source and system components without any a priori knowledge about source and / or system. The redundancy in the speech signal is exploited in the LP analysis. The prediction of current sample as a linear combination of past *p* samples form the basis of linear prediction analysis where *p* is the order of prediction. The cepstrum computation discussed so far is known as the real cepstrum.

As the real cepstrum is computed from the log magnitude spectrum, the phase part is ignored. This will not enable the reconstruction of the sequence from the cepstrum. However the reconstruction can be done by preserving the Fourier phase and use it for reconstruction from the real cepstrum.

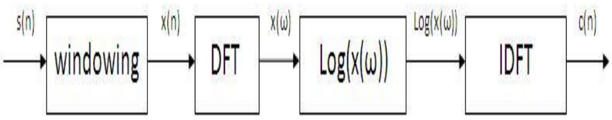

Fig – 1 Block diagram representing complex cepstrum

3. CONCLUSIONS

The objective of this experiment is to separate the excitation and vocal tract components of a given speech signal by Cepstral analysis. The first step is to convert the speech into short-term segments of size 15-20 ms. Here the frame size is fixed to 20 ms. Then each frame is multiplied by a hamming window. Then Cepstral representation of short-term speech is computed by finding the IDFT of the log magnitude spectrum. The objective of this experiment to estimate the LP coefficients of order P by autocorrelation method. The first step is to perform the autocorrelation analysis of speech frames having length of 15-20ms after multiplying it with a hamming window.

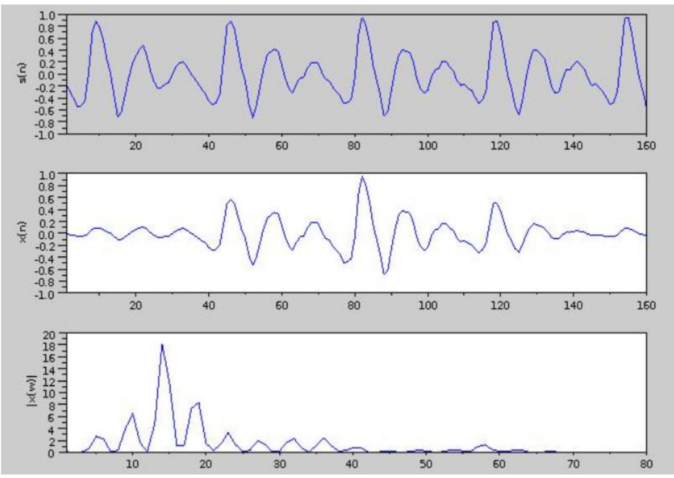

Fig - 2 Voiced speech cepstrum

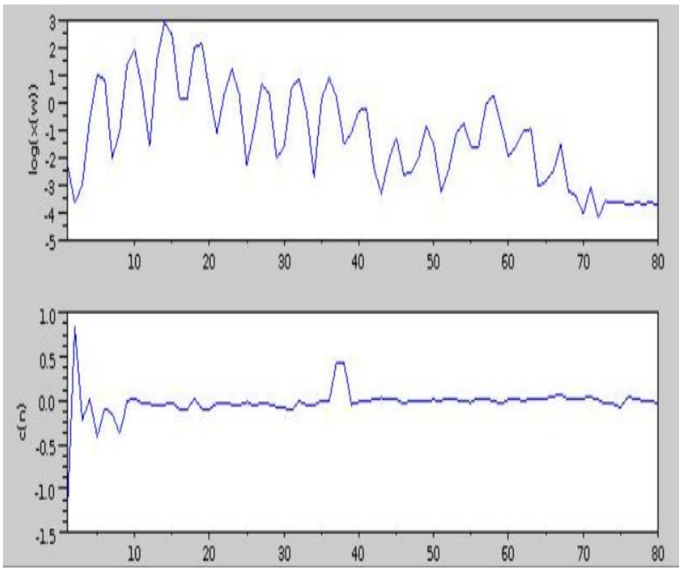

Fig -3 20 ms voiced speech segment

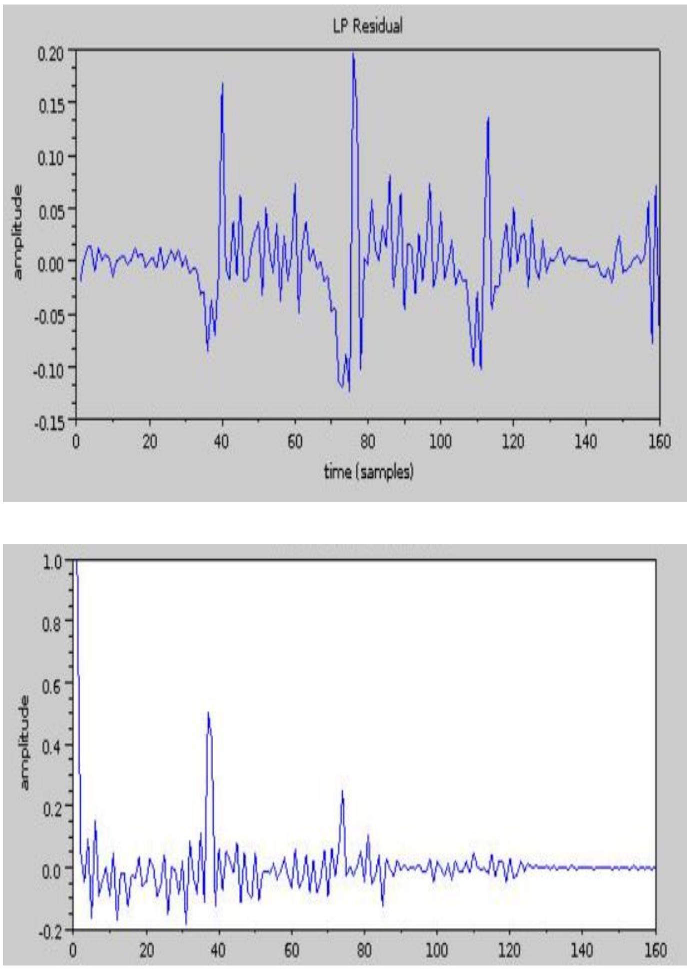

Fig – 4 The Speech and residual of 20 ms voiced speech segment

REFERENCES

[1] vlab.amrita.edu/Cepstral analysis of speech

[2] vlab.amrita.edu/Linear prediction of speech

[3] Anantha Krishna, K.Shama, and U.C.Niranjan(2004) "k Means nearest Neighbor Classifier for voice pathology," proceedings of IEEE, India Annual Conference (INDICON'04) pp.232-234.

[4] Barkana.B.D, Atal.B (2009) "A pattern recognition approach to voiced-unvoiced silence classification with applications of speech recognition'', IEEE Trans on ASSP, vol.ASSP, pp.201-212.

[5] Childers.D.G, M.Hahn and J.N.Larar, (1989) "Silent and voiced classification of speech, "IEEE Trans. On ASSP, vol.37 No.11,pp.1771-1774.

[6] Davis, S.B. Acoustic characteristics of Normal and pathological voices. Speech and language: Advances in Basic Research and practice, vol.1, pp.271-335.

[7] L.E.McBride, Jr.and K.S.Narendra, ''Optimization of time-varying systems", IEEE Trans.Automat,vol.10, pp.289-294.

[8] Paliwal K.K., Klejin W.B. Quantization of LPC parameter //Speech coding and Synthesis.pp.443-446.

[9] Qi.Y and B.R Hunt, (1993) "voiced-unvoiced-silence classification of speech using Hybrid Features and Network classifier'', IEEE Trans, Speech Audio Processing, vol.1.2, pp.250-255.