International Journal of Innovations in Engineering and Technology (IJIET)
http://dx.doi.org/10.21172/ijiet.261.01

Analyzing the Effectiveness of Speech Feature Extraction Methods: A Comparative Study

Monali Rajendra Baviskar
Department of Computer Engineering
Government College of Engineering, Aurangabad, Maharashtra, India

Dr. Praveen Chhatanrao Shetiye
Associate Professor, Computer Science and Engineering,
MCA department, Government College of Engineering, Aurangabad, Maharashtra, India.

Abstract- Feature recognition techniques are used in speech recognition systems to identify and extract relevant information from speech signals. In the case of disordered speech, such as that produced by individuals with speech disorders, these techniques can be particularly challenging due to the variability and unpredictability of the speech patterns. To improve the accuracy of speech recognition for disordered speech, various feature recognition techniques have been proposed, including the use of spectral features, pitch-based features, and prosodic features. These techniques take into account the unique characteristics of disordered speech and can improve the performance of speech recognition systems. There are several techniques used for speech feature extraction, some of the most commonly used include Mel-Frequency Cepstral Coefficients (MFCCs), Linear Predictive Coding (LPC), Perceptual Linear Predictive (PLP), Line Spectral Frequencies (LSF). Overall, the development of feature recognition techniques in disordered speech recognition systems has the potential to significantly improve the quality of life for individuals with speech disorders by enabling them to more easily communicate with electronic devices.

Keywords – MFCC, LPC, LSF, PLP

I. MEL-FREQUENCY CEPSTRAL COEFFICIENTS (MFCCS)

Mel-Frequency Cepstral Coefficients (MFCCs) are a feature representation commonly used in speech and music processing. They summarize the spectral envelope (shape of sound frequencies) of a sound clip into a compact set of coefficients, which can then be used for tasks such as speaker recognition, music genre classification, and speech recognition.

MFCCs are the most commonly used feature extraction technique in speech recognition systems. They are calculated by applying a Mel-scale filter bank to the log power spectrum of the speech signal, followed by discrete cosine transform (DCT). MFCCs capture the spectral envelope of the speech signal, which is particularly important for recognizing phonemes and words.[1][2]

In simple terms, MFCCs can be thought of as a way to extract the unique characteristics of a sound and represent them as numerical values, allowing computers to analyze and recognize different sounds.

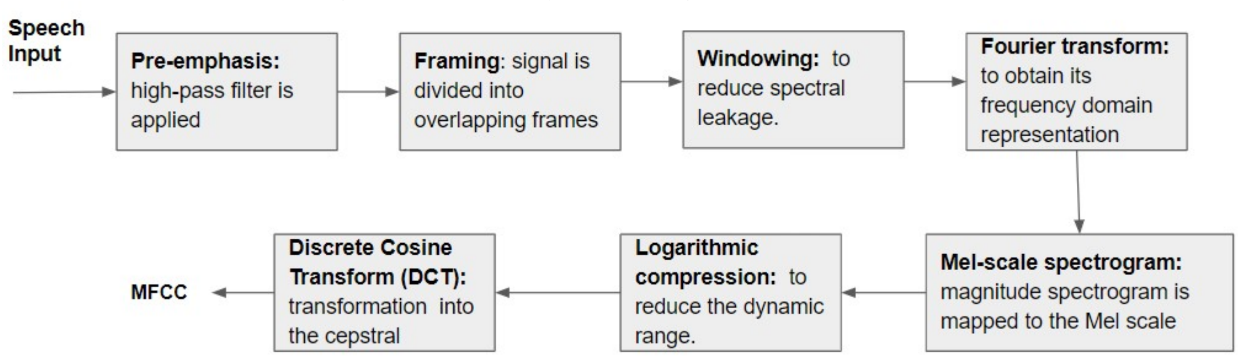

Figure 1. Processing steps involved in computing MFCC

http://dx.doi.org/10.21172/ijiet.261.01

Features:
1. Human auditory system: MFCCs are designed to mimic the human auditory system's response to sound, making them a good representation of speech for computer processing.
2. Compact representation: MFCCs provide a compact representation of the spectral envelope of a sound, reducing the amount of data that needs to be processed.
3. Robustness to noise: MFCCs are robust to additive noise, meaning that they can still accurately represent speech even in noisy environments.
4. Invariant to changes in pitch: MFCCs are designed to be invariant to changes in pitch, making them a good representation for speech recognition across different speakers with different pitch levels.
5. Good performance: MFCCs have been extensively tested and shown to have good performance on a variety of speech recognition tasks.

II. LINEAR PREDICTIVE CODING (LPC)

Linear Predictive Coding (LPC) is a mathematical method used to analyze and represent speech signals. It is widely used in speech and audio processing for tasks such as speech synthesis, speech coding, and speech recognition.

Linear Predictive Coding (LPC): LPC is a technique for modeling the vocal tract and predicting the speech signal based on the model parameters. The parameters are then used as features for speech recognition. LPC has been widely used for speech recognition due to its computational efficiency and robustness against noise. [3][4]

The basic idea behind LPC is to model a speech signal as a weighted sum of past samples, with the weights representing the vocal tract characteristics of the speaker. The vocal tract characteristics are captured using a set of LPC coefficients, which can be estimated from the speech signal using linear regression.

Once the LPC coefficients are obtained, they can be used to synthesize a speech signal that is similar to the original, but with a reduced number of samples. This makes LPC a useful tool for speech coding and compression.

In speech recognition, LPC coefficients can be used as features to represent the spectral envelope of the speech signal. By modeling the vocal tract characteristics of the speaker, LPC features can provide a more compact and robust representation of speech compared to raw speech samples, improving the accuracy of speech recognition systems.

Overall, LPC is a powerful tool for analyzing and representing speech signals, and it continues to be widely used in various speech and audio processing applications.

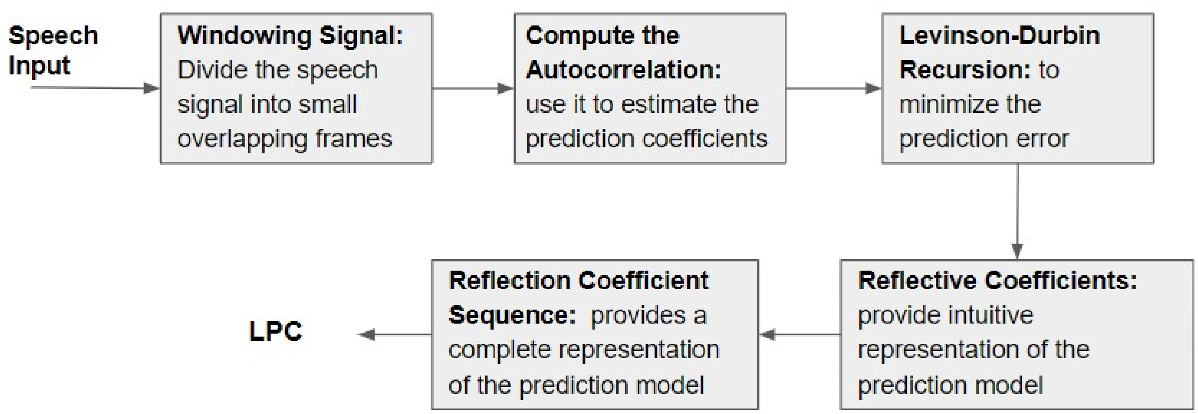

Figure 2. Processing steps involved in computing LPC

Features:
1. Predictive Modeling: LPC models the vocal tract as a linear filter and predicts the current sample based on past samples.
2. Filter Coefficients: LPC estimates the filter coefficients that characterize the vocal tract and represent the speech signal.

3. Compact Representation: LPC provides a compact representation of the speech signal by representing it as a small number of filter coefficients.
4. Noise Reduction: LPC can be used to reduce noise in speech signals by predicting and removing the noise component.
5. Coding Efficiency: LPC provides efficient coding of speech signals by taking advantage of the redundancy in speech signals.
6. Widely Used: LPC is widely used in speech coding, speech synthesis, speech recognition, and speech analysis applications.
7. Versatility: LPC can be adapted to various applications by choosing the appropriate model order and analysis window size.

III. PERCEPTUAL LINEAR PREDICTIVE (PLP)

Perceptual Linear Predictive (PLP) is a method for processing and analyzing speech signals that aims to model the way the human auditory system processes speech.

PLP is an extension of LPC that incorporates human perceptual principles and has been found to perform better than LPC in speech recognition tasks. PLP features are based on the spectral envelope of the speech signal, as in MFCCs, but use a more sophisticated filter bank and DCT to obtain the features.[5][6]

PLP works by first converting the speech signal into a representation that corresponds to the auditory spectral envelope. This is done by applying a series of processing steps, such as transforming the signal into the frequency domain, and computing a smoothed estimate of the spectral envelope.

Next, PLP applies a linear predictive coding (LPC) technique to model the spectral envelope as a filter. This involves estimating a set of filter coefficients that can be used to predict the spectral envelope from past samples.

The resulting filter coefficients and other PLP features can be used for various speech processing tasks, such as speech coding, speech recognition, and speech analysis. The goal of PLP is to provide a representation of speech that is more representative of how the human auditory system perceives speech, compared to traditional LPC.

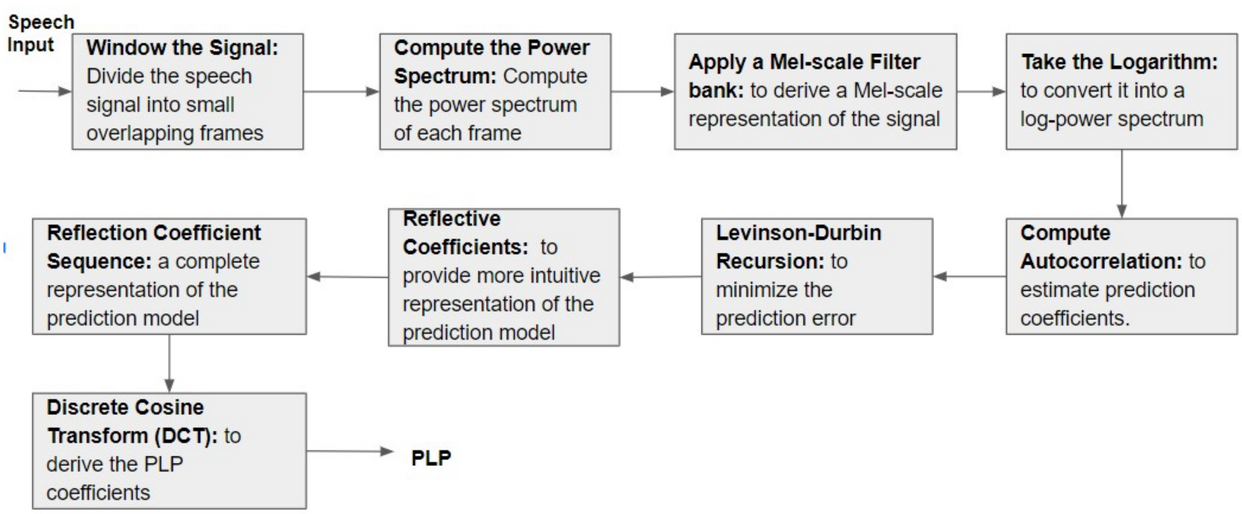

Figure 3. Processing steps involved in computing PLP

Features:

1. Auditory Modeling: PLP models the human auditory system by converting the speech signal into a representation that corresponds to the auditory spectral envelope.
2. Linear Predictive Coding (LPC): PLP applies LPC to model the auditory spectral envelope as a filter, providing a compact representation of the speech signal.

http://dx.doi.org/10.21172/ijiet.261.01

3. Smoothed Spectral Envelope: PLP computes a smoothed estimate of the spectral envelope, reducing spectral fluctuations and emphasizing spectral structures that are important for perception.
4. Robustness: PLP is designed to be robust to various types of distortions, such as noise and interference, making it well-suited for practical speech processing applications.
5. Perceptually Motivated Features: PLP provides a set of features that are motivated by the human auditory system, making it well-suited for speech processing tasks that aim to model perceptual properties of speech.
6. Widely Used: PLP is widely used in speech processing, speech recognition, and speech analysis applications, due to its effectiveness and simplicity.
7. Adaptability: PLP can be adapted to various applications by choosing the appropriate processing steps, such as the type of window function used, and the number of filter coefficients.

IV. LINE SPECTRAL FREQUENCIES (LSF)

Line Spectral Frequencies (LSF) is a speech feature extraction technique used in the field of speech processing. LSFs are obtained by representing the spectral envelope of speech signals in terms of the line spectral frequencies of the vocal tract resonances, also known as formants.[7][8][9][10]

The LSFs are obtained using a mathematical technique called line spectral pairs (LSPs), which involves transforming the filter coefficients of the speech signal into a set of line spectral frequencies. The LSFs provide a compact and robust representation of the spectral envelope of speech signals and capture important information about the formants of the speech.

One advantage of LSFs over other speech feature extraction techniques, such as Mel-Frequency Cepstral Coefficients (MFCCs) and Linear Predictive Coding (LPC), is that they are less sensitive to quantization error and provide improved robustness in noisy environments.

LSFs have been used in a variety of speech processing applications, including speech recognition, speaker identification, and emotion recognition. They have also been used in speech coding, which involves the efficient representation of speech signals for communication purposes.

In summary, LSFs are a useful and widely used speech feature extraction technique that provide a compact and robust representation of the spectral envelope of speech signals and capture important information about the formants of the speech.

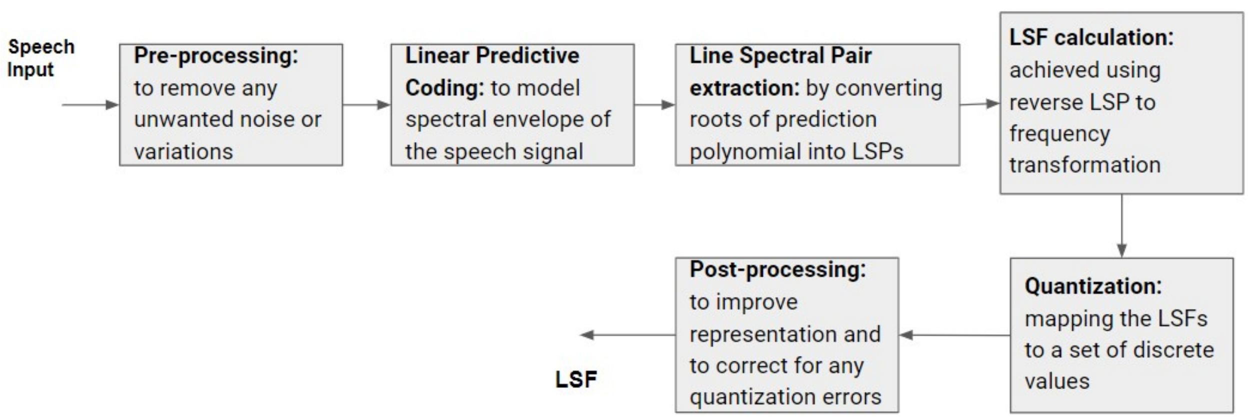

Figure 4. Processing steps involved in computing LSF

Features:

1. Robustness to noise: LPC coefficients are robust to noise and channel variability, as they provide a compact representation of the spectral envelope that is less sensitive to these factors compared to other feature extraction techniques.
2. Compact representation: LPC coefficients provide a compact representation of the speech signal, which makes them suitable for storage and transmission.

3. Ease of computation: LPC is a simple and computationally efficient technique, as it only requires the estimation of a small number of coefficients.
4. Flexibility: LPC can be adapted to different speech processing applications by varying the number of coefficients used and the way in which they are computed.
5. Widely used: LPC is a widely used speech feature extraction technique that has been successfully applied in a variety of speech processing applications, such as speech recognition, speaker identification, and emotion recognition.

V.CONCLUSION

In conclusion, speech feature extraction is a crucial step in speech processing and recognition. Various techniques have been developed and explored, including time-domain, frequency-domain, and cepstral analysis. Each technique has its advantages and disadvantages, and the choice of technique depends on the specific application requirements. Pre-processing steps such as filtering, and normalization can enhance the quality of the extracted features. Furthermore, machine learning algorithms, such as hidden Markov models and neural networks, can be used to classify and recognize speech based on the extracted features. Overall, selecting appropriate speech feature extraction techniques is essential for achieving accurate and efficient speech processing and recognition.

V. REFERENCE

[1] Davis and Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Transactions on Acoustics, Speech, and Signal Processing, 1980.

[2] R. A. Gopinath and R. C. Rout, “Speech signal processing and recognition,” International Journal of Applied Engineering Research, vol. 11, no. 6, pp. 3826–3830, 2016.

[3] J. D. Markel and A. H. Gray, "Linear Prediction of Speech," Springer, 1976.

[4] S. R. Mahanta and B. K. Kalita, “Speech recognition system: a review,” International Journal of Computer Applications, vol. 78, no. 13, pp. 12–16, 2013.

[5] H. Hermansky and N. Morgan, “Perceptual linear predictive (PLP) analysis of speech,” Journal of the Acoustical Society of America, vol. 87, no. 4, pp. 1738–1752, 1990.

[6] K. K. Paliwal and R. A. Gopinath, “Speech recognition using PLP and LPC analysis,” International Journal of Computer Science and Network Security, vol. 5, no. 3, pp. 153–159, 2005.

[7] Markel, J. D., & Gray, A. H. (1976). Linear Prediction of Speech. Springer US.

[8] Atal, B. S., & Hanauer, J. (1971). Speech Analysis and Synthesis by Linear Prediction of the Speech Wave. Bell System Technical Journal, 50(6), 1637-1652.

[9] Rabiner, L. R., & Schafer, R. W. (1978). Digital Processing of Speech Signals. Prentice-Hall.

[10] Balan, R., & Drygajlo, A. (2003). Linear Predictive Coding. Springer.

[11] Davis and Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," IEEE Transactions on Acoustics, Speech, and Signal Processing, 1980.

[12] R.J. Brugman and M.J. Murrell, "Mel-frequency cepstral coefficients for music modeling," Proceedings of the International Conference on Digital Audio Effects, 2002.

[13] P.N. Singh and R.N. De Alwis, "Mel-frequency cepstral coefficients for music genre classification," Proceedings of the International Conference on Advances in Computing, Communications, and Informatics, 2014.