

Automatic Speech Recognition Features Extraction Techniques: A Multi-criteria Comparison

Maria Labied, Abdessamad Belangour

Laboratory of Information Technology and Modeling LTIM
Hassan II University, Ben M'sik Faculty of Sciences
Casablanca, Morocco

Abstract—Features extraction is an important step in Automatic Speech Recognition, which consists of determining the audio signal components that are useful for identifying linguistic content while removing background noise and irrelevant information. The main objective of features extraction is to identify the discriminative and robust features in the acoustic data. The derived feature vector should possess the characteristics of low dimensionality, long-time stability, non-sensitivity to noise, and no correlation with other features, which makes the application of a robust feature extraction technique a significant challenge for Automatic Speech Recognition. Many comparative studies have been carried out to compare different speech recognition feature extraction techniques, but none of them have evaluated the criteria to be considered when applying a feature extraction technique. The objective of this work is to answer some of the questions that may arise when considering which feature extraction techniques to apply, through a multi-criteria comparison of different features extraction techniques using the Weighted Scoring Method.

Keywords—Automatic speech recognition; feature extraction; comparative study; MFCC; PCA; LPC; DWT; WSM

I. INTRODUCTION

Features extraction is a fundamental step in the Automatic Speech Recognition (ASR) process, in which relevant data are extracted from a speech. After pre-processing a speech signal (noise reduction, endpoint identification, pre-emphasis, framing, and normalization), the feature extraction stage retains a set of predefined features from the processed speech, using extraction techniques such as Mel-Frequency Cepstral Coefficients (MFCCs), Discrete Wavelet Transforms (DWTs), Linear Predictive Coding (LPC) and other techniques that will be explored in greater depth in this paper focusing on the advantages and disadvantages of each one.

The content of this paper is structured as follows. In Section 2 we review related work that has been done to compare existing features extraction techniques. In Section 3 we describe the different features extraction techniques. Then, in Section 4, we present the main advantages and disadvantages of each extraction method. In Section 5 we provide a multi-criteria comparison of the different methods based on the Weighted Scoring Method (WSM). Finally, we end with a conclusion.

II. RELATED WORK

Several works have been conducted to compare ASR features extraction techniques [1][2][3][4][5][6]. Most of this research has been focused on the advantages and disadvantages of each extraction method. Nevertheless, it is relevant to illustrate the importance of each one depending on the criteria that represent the key elements when deciding on which method to use for feature extraction in a speech recognition system.

In the research [3] most commonly used feature extraction techniques have been discussed, like LPC, MFCC, Zero Crossings With Peak Amplitudes (ZCPA), DTW, and Relative Spectral Processing (RASTA). In this work, the limitations of each the advantage of each technique have been addressed. Also, it was mentioned that most research used only a single feature extraction technique and that is important to think about using hybrid techniques that combine between two or more than one feature extraction technique. In the same scope, another research [4] has been established a comparison of various feature extraction techniques (MFCC, LPC, DWT, PLP...) considering the specific advantages and the shortcomings of each.

An analysis of different feature extraction techniques has been investigated in the work [5], for isolated words speech in a clean and noisy environment for feature extraction techniques like PLP, RASTA PLP, LPCC, and MFCC. This analysis has been based on a comparison of the obtained accuracy, using each technique in both noisy and clean environments. Another work [6] has studied the performance of commonly used feature extraction techniques (MFCC, LPC, and PLP) for speech recognition. Through illustrating their benefits and drawbacks. This paper highlights the importance of hybrid feature extraction techniques to benefit from the advantage of multiple techniques at the same time.

III. SPEECH FEATURES EXTRACTION TECHNIQUES

The features extraction methods are used to remove irrelevant information from a speech signal. Depending on the type of feature to be extracted, feature extraction methods can be classified into two main categories: Spectral feature analysis methods, which use the spectral representation of the speech signal. And temporal feature analysis methods, which use the original form of a signal.

The well-known feature extraction method in the field of ASR is the Mel-frequency cepstral coefficient (MFCC). In addition to this technique, there are other extraction methods for ASR, such as the Discrete wavelet transform (DWT), Wavelet packet transforms (WPT), Relative Spectral-Perceptual Linear Prediction (RASTA-PLP), Linear predictive coding (LPC), and others. We present in-depth each of these methods in the following sections.

A. Spectral Feature Analysis Methods

1) *Mel-Frequency Cepstral Coefficients (MFCC)*: Several researchers chose MFCC as their feature extraction method. Since the mid-1980s, MFCCs are the most widely used feature extraction method in the field of ASR. Most of the works concerning Moroccan Darija Speech recognition have used the MFCC as a feature extraction method [7], [8]. The main purpose of this feature extraction method is to mimic the human ear. The MFCC is calculated by first splitting the speech signal into alternating frames with a length of 25 or 30 milliseconds and a 10-millisecond overlap between consecutive frames. The discrete Fourier transform (DFT) is computed on each windowed frame after each frame is multiplied with a Hamming window function.

MFCC is well-known and commonly used in the field of speech recognition, but they do have some drawbacks. The key disadvantage of MFCC is its poor robustness to noise signals, as noise signals change all MFCCs if at least one frequency band is skewed. Various normalization techniques are used for enhancing the robustness of MFCC to noise-corrupted speech signals, in both training and testing conditions. These include features statistics normalization techniques such as mean and variance normalization (MVN), histogram equalization (HEQ), and cepstral mean normalization (CMN). Another important issue with MFCCs is that these are derived only from the power spectrum of a speech signal, ignoring the spectrum phase. However, the provided information by this phase is also useful for speech perception. This problem is tackled by performing speech enhancement before starting features extraction.

2) *Principal Component Analysis (PCA)*: Determining a linear combination that can be used to represent the original speech signal is the main role of PCA in the feature extraction stage. PCA is mainly used for dimensionality reduction and features de-correlation. It is the most used method to increase the robustness of the speech recognition systems in a noisy environment. The research presented in [9], states that the PCA analysis is required when the speech signal is corrupted by noises. Another research confirms that the usage of PCA had given further reduction in the error rates [10]. In the research [11] the combination of PCA with MFCC had increased the recognition rates obtained with noisy speech signals from 63.9% to 75.0%. However, less accuracy is obtained using PCA for spontaneous and continuous speech recognition [12].

3) *Linear Predictive Coding (LPC)*: The LPC is the most important method for extracting features [13] and the most

used in several works [14][15]. Unlike MFCC, this method imitates the basic structure of a vocal tract when a sound is produced. LPC analysis is carried out by generating frames for the input speech signal, then performing windowing on each frame to reduce the discontinuities at the beginning or the end of a frame. Finally, the inter-frame autocorrelation is calculated. LPC method recognition quality is affected by noises, several works proposed new approaches to enhance the performance of this method in a noisy environment [16]. Multiple works have used LPC in combination with DWT [17], by using DWT to decompose the input speech signal and LPC to model each sub-band. The results obtained confirm that this method outperforms by 10% the MFCC method.

4) *Linear Predictive Cepstral Coefficient (LPCC)*: The LPCC is considered as an extension of the LPC method [8]. After performing the LPC analysis, cepstral analysis is carried out to obtain the corresponding cepstral coefficient. Many researchers studied the performance of both LPCC and MFCC. The results obtained in [18] show that MFCC and LPCC share the same results. Another research [19] compared the two methods confirm that LPCC was 5.5% faster and 10% more efficient than MFCC.

5) *Perceptual Linear Prediction (PLP)*: The PLP method is mainly used to remove unwanted information from a speech signal and improves the speech recognition rate. The PLP analysis consists of two important stages. First by approaching the auditory system spectrum by the model of all poles, then calculating the auditory spectrum [20]. The results of PLP analysis and LPC are identical, with the exception that the order of the PLP analysis model is half of the LPC model. This model allows for storage saving of automatic speech recognition storage and also provides good ASR performance.

B. Temporal Feature Analysis Methods

1) *Discrete Wavelet Transforms (DWT)*: In addition to the frequency information, the temporal information in speech signals is also important for speech recognition applications [21][22]. Due to the non-stationary nature of speech signals, DWT obtains temporal information by re-scaling, shifting, and analyzing the mother wavelet. In this manner, the input speech signal is analyzed at various frequencies and resolutions. Since a speech signal is analyzed at decreasing frequency resolution for increasing frequencies, the DWT provides an appropriate model for the human auditory system and it was used in various researches at the feature extraction stage [23][24]. In comparison to MFCC, the DWT offers better frequency resolution at lower frequencies. As previously mentioned, MFCC is not robust for noise-corrupted speech signals. Because of their ability to provide localized time and frequency information, DWT was effectively used for de-noising tasks.

Several researchers considered combining the DWT and the MFCC to gain the benefits of both methods. This combination is known as Mel-Frequency Discrete Wavelet Coefficients (MFDWC) and is produced by applying the DWT

to the Mel-Scaled log filter bank energies of a speech frame. In the works [25][26] the MFDWC method was used and for both clean and noisy environments, the results showed that MFDWC achieved higher accuracy as compared to MFCC and wavelet transforms alone.

2) *Wavelet Packet Transforms (WPT)*: The wavelet packet transform is an extension of the standard wavelet decomposition that provides extra signal processing options. When compared to the wavelet transform, it better represents high-frequency information. The main difference between wavelet transforms and wavelet packet transform is that the latter split details as well as approximations.

WPTs are similar to DWTs, except that both the approximation and detail coefficients are more decomposed. The research [1] compared WPT's performance to that of DWT for the task of ASR, the results showed that WPT-based methods performed better as than WPT's.

3) *Relative Spectra-Perceptual Linear Prediction (RASTA-PLP)*: The RASTA-PLP analysis involves combining the RASTA technique with the PLP method to improve the robustness of the PLP features. This method is based on the fact that the temporal properties of a speech signal environment differ from those of the speech signal. Thus, by using band-pass filtering on each frequency sub-band of a speech signal, the effects of channel mismatch between the training and testing environments are reduced and the short-term noises are smoothed [27]. The work done in [28] confirms the robustness of RASTA-PLP for noisy environments. Another work [29] have compared the RASTA-PLP with LPC and MFCC feature extractions techniques, the obtained results shown That the RASTA-PLP performs better than MFCC and LPC for noisy speech signal with an accuracy of 73% while 60% accuracy had been

obtained using MFCC and 53% of accuracy obtained using LPC. Furthermore, RASTA-PLP performs much better when it is combined with the WPT method.

IV. COMPARISON OF FEATURE EXTRACTION TECHNIQUES

There are several criteria to consider when chosen a feature extraction technique, such as the accuracy of recognition in a noisy speech environment, computations costs, storage space, temporal information of speech signals, and others.

When it comes to noisy environments RASTA-PLP outperforms MFCC, PLP, and LPC features extractions methods. The MFCC is suitable for a clean speech and performs better for an isolated speech environment, while it is low robust to noise and not suitable for a continuous speech environment since the MFCC frame may contain information of more than one phoneme. For more robustness either MFCC, LPC or PLP may be combined with other feature techniques such as DWT or WPT for enhancing systems robustness when it is needed to use such feature extraction techniques in a noisy environment [30]. The temporal information of a speech signal is as significant as its frequency information. The DWT and WPT outperform the well-known MFCC for such issues. Thanks to these two methods a better accuracy is achieved for phonemes recognition. Also, available memory space is an important criterion for choosing the feature extraction method that will achieve good accuracy with a limited feature vector size. DWT may be a good option when only small storage is available. While MFCC requires more storage space. For this reason most of the time, MFCC is used in combination with other features extraction techniques to reduce the dimensionality of extracted features and to obtain good accuracy, vector quantization (VQ), PCA, or LDA [31][32].

The profits and constraints of the above discussed features extraction techniques are illustrated in Table I.

TABLE I. FEATURES EXTRACTION TECHNIQUES COMPARISON

Feature extraction	Pros	Cons
MFCC	- High recognition accuracy [27] - Good discrimination and low coefficients correlation [27]	- Inaccurate recognition in noisy speech [27] -high dimensional features vectors [31]
PCA	- Robustness to noises [9] - Reduce the feature vector's size while retaining important information [15]	- Expensive in terms of computing for high-dimensional data [27]
LPC	- Computation speed - Robust for extracting features from speech signals with a low bit rate [33]	- Highly correlated feature coefficients [27] - Unable to distinguish words with similar phonemes
LPCC	- Decorrelate feature coefficients by the cepstral analysis - Robust than LPC [27]	- Unable to analyze local events accurately
PLP	- Low-dimensional feature vector [27] - Reduce the gap between voiced and unvoiced speech	- Altered Spectral balance [27]
DWT	- Denoising speech signal [34] - Compressing speech signal without significant loss of its quality [27]	- Inflexible [27]
RASTA-PLP	- Robustness - Excludes variations between cepstral components and speech signal [27]	- Low performance for noiseless speech [21]

V. MULTI-CRITERIA COMPARISON

After describing the advantages and disadvantages of each feature extraction method in this section we present a multi-criteria comparison of these methods. In this comparison, we used the Weighted Scoring Method (WSM), which is known as the simple additive weighting method, which involves adding up the criteria values for each alternative and applying the individual criteria weights [35]. To apply this method, we went through the steps below:

- Criteria selection.
- Assigning weights to criteria based on their importance.
- Creating a matrix containing weights for each criterion.
- Calculation of weights scores.

A. Comparison Criteria

The choice of comparison criteria was based on the common characteristic shared between previously cited feature extraction methods. In the following we present the most important criteria to be considered when choosing a speech feature extraction technique:

- C1=Robustness to noises: This criterion involves if we can use a feature extraction technique when a speech signal is corrupted by noises.
- C2=Memory storage: This criterion concerns the size of storage space required for the spectral analysis of a speech while maintaining important information from a speech.
- C3=Dimensionality reduction: This criterion indicates the ability of a feature extraction technique to reduce the dimensionality of extracted features while obtaining a good accuracy.
- C4=Computational Complexity: This criterion covers the computational costs of a feature extraction method in terms of time and speed.
- C5=Computational Speed: This criterion covers the computational costs of a feature extraction method in terms of time and speed.
- C6=Temporal information within speech: This criterion highlights the implication of temporary information of a speech by the feature extraction method.
- C7=Suitability for continuous speech: This criterion considers the performance of a feature extraction method in the context of a continuous speech.
- C8=Suitability for spontaneous speech: This criterion considers the performance of a feature extraction method in the context of a spontaneous or a real-time speech.
- C9=Suitability for isolated words speech: This criterion points out the performance of feature extraction when dealing with isolated words speech.

- C10=Reinforcing recognition rate: This criterion indicates whether the application of a feature extraction method improves the speech recognition rate.

B. Application of WSM

The application of the WSM method consists of determining the multi-criteria matrix where the columns represent the feature extraction methods and the rows represent the criteria with their corresponding weights. The score attributed to each criterion has been induced from the comparison detailed in the previous sections.

The scoring is based on five levels where each one is defined as follows:

- Score "1": A poor nor lower performance is obtained using a method.
- Score "2": Inflexibility and lack of efficiency using a method.
- Score "3": A good option, but there are some limitations to use a method.
- Score "4": Significant results are obtained, more flexibility for foreign contexts is needed.
- Score "5": Approved efficiency, all requirements are met by using a method.

In Table II, the resulting WSM Matrix is presented according to the score assigned to each criterion.

TABLE II. WSM MATRIX

	MFCC	PCA	LPC	LPCC	PLP	DWT	RASTA-PLP
C1	1	5	1	1	4	5	5
C2	2	4	3	3	5	4	5
C3	2	5	4	4	5	5	4
C4	5	2	3	4	4	2	4
C5	3	4	5	5	4	4	5
C6	3	4	1	1	3	5	3
C7	2	3	2	3	4	4	4
C8	2	2	2	3	3	3	4
C9	5	4	4	4	4	2	3
C10	4	5	3	4	4	5	5

VI. DISCUSSION

According to the multi-criteria comparison of different feature extraction techniques, each feature extraction technique performs better for a particular criterion, as shown in the WSM results in Matrix presented in Table III, we can see that MFCC is less complex and effective for extracting features from an isolated words speech, but it lacks robustness to noises, which make it non preferred for other speech types (spontaneous, continuous).

TABLE III. WSM RESULTS

	Weights	MFCC	PCA	LPC	LPCC	PLP	DWT	RASTA-PLP
C1	10	10	50	10	10	40	50	50
C2	10	20	40	30	30	50	40	50
C3	10	20	50	40	40	50	50	40
C4	10	50	20	30	40	40	20	40
C5	15	45	60	75	75	60	60	75
C6	5	15	20	5	5	15	25	15
C7	10	20	30	20	30	40	40	40
C8	5	10	10	10	15	15	15	20
C9	15	75	60	60	60	60	30	45
C10	10	40	50	30	40	40	50	50
Scores	100	305	390	310	345	410	380	425

VII. CONCLUSION

In this work, we presented a multi-criteria comparison of commonly used feature extraction techniques in ASR. The choice of feature extraction techniques is a crucial step in the speech recognition process since the wisely we choose the extraction technique, the more accurate results we get.

This comparison revealed that each extraction method has reliability and performance issues. Also, the results showed the importance of applying hybrid feature extraction techniques since each of the presented extraction techniques complements the work of another.

The main objective of this multi-criteria comparison is to help researchers to select the feature extraction method according to the criteria that matter most for speech recognition.

REFERENCES

- [1] M. Iosif, G. Todor, S. Mihalis, and F. Nikos, "Comparison of Speech Features on the Speech Recognition Task," *J. Comput. Sci.*, vol. 3, no. 8, pp. 608–616, 2007.
- [2] A. N. Mishra, M. C. Shrotriya, and S. N. Sharan, "Comparative Wavelet , PLP and LPC Speech Recognition Techniques on the Hindi Speech Digits Database," vol. 7546, pp. 1–6, 2010, doi: 10.1117/12.856318.
- [3] P. K. Kurzekar, R. R. Deshmukh, V. B. Waghmare, and P. P. Shrishrimal, "A Comparative Study of Feature Extraction Techniques for Speech Recognition System," *Int. J. Innov. Res. Sci. Eng. Technol.*, vol. 03, no. 12, pp. 18006–18016, Dec. 2014, doi: 10.15680/IJRSET.2014.0312034.
- [4] G. Kalyani and S. Chaudhari, "A Review of Various Techniques Related to Feature Extraction and Classification for Speech Signal Analysis," in *Lecture Notes in Electrical Engineering*, vol. 601, 2020, pp. 696–702.
- [5] G. Kaur, M. Srivastava, and A. Kumar, "Analysis of feature extraction methods for speaker dependent speech recognition," *Int. J. Eng. Technol. Innov.*, vol. 7, no. 2, pp. 78–88, 2017.
- [6] U. Sharma, S. Maheshkar, and A. N. Mishra, "Study of robust feature extraction techniques for speech recognition system," in *2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*, Feb. 2015, pp. 654–658, doi: 10.1109/ABLAZE.2015.7154944.
- [7] B. Mouaz, B. H. Abderrahim, and E. Abdelmajid, "Speech recognition of Moroccan dialect using hidden Markov models," in *Procedia Computer Science*, Jan. 2019, vol. 151, pp. 985–991, doi: 10.1016/j.procs.2019.04.138.
- [8] A. Ezzine, H. Satori, M. Hamidi, and K. Satori, "Moroccan Dialect Speech Recognition System Based on CMU SphinxTools," in *2020 International Conference on Intelligent Systems and Computer Vision (ISCV)*, Jun. 2020, pp. 1–5, doi: 10.1109/ISCV49265.2020.9204250.
- [9] H. Veisi and H. Sameti, "The integration of principal component analysis and cepstral mean subtraction in parallel model combination for robust speech recognition," *Digit. Signal Process.*, vol. 21, no. 1, pp. 36–53, Jan. 2011, doi: 10.1016/j.dsp.2010.07.004.
- [10] J.-Y. Lee and J. Hung, "Exploiting principal component analysis in modulation spectrum enhancement for robust speech recognition," in *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, Jul. 2011, pp. 1947–1951, doi: 10.1109/FSKD.2011.6019893.
- [11] T. Takiguchi and Y. Ariki, "PCA-Based Speech Enhancement for Distorted Speech Recognition," *J. Multimed.*, vol. 2, no. 5, Sep. 2007, doi: 10.4304/jmm.2.5.13-18.
- [12] S. N. Sawant and M. S. Kumbhar, "Real time Sign Language Recognition using PCA," in *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, May 2014, pp. 1412–1415, doi: 10.1109/ICACCCT.2014.7019333.
- [13] M. Narayanbhai Raja, M. Priyanka Richhpal Jangid, and S. M. Gulhane, "LINEAR PREDICTIVE CODING," *Int. J. Eng. Sci.*, pp. 373–379, Apr. 2015, [Online]. Available: <http://www.ijesrt.com>.

Other extraction methods like PCA, DWT, and RASTA-PLP are effective in reducing noise, which is an important factor in building a robust ASR system. Temporal information within a speech is less considered by most feature extraction methods. Also, memory storage optimization is an important issue to be filled by a feature extraction method.

From these results, we can deduce that none of the presented methods meet the flexibility and robustness requirements of ASR. The multi-criteria spider graph shown in Fig. 1 illustrates that there is no complete extraction method that respond to each criterion. However, we emphasize the importance of combining multiple feature extraction techniques to benefit from the effectiveness of each.

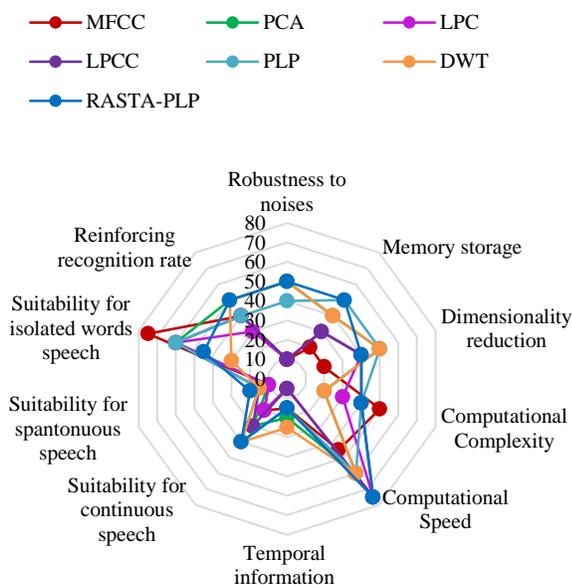


Fig. 1. Multi-criteria Spider Chart.

- [14] F. Ishiyama, Y. Okugawa, and K. Takaya, "Linear predictive coding without Yule-Walker approximation for transient signal analysis: Application to switching noise," in 2017 IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA), Mar. 2017, pp. 44–48, doi: 10.1109/CSPA.2017.8064922.
- [15] L. Grama and C. Rusu, "Audio signal classification using Linear Predictive Coding and Random Forests," in 2017 International Conference on Speech Technology and Human-Computer Dialogue (SpED), Jul. 2017, pp. 1–9, doi: 10.1109/SPED.2017.7990431.
- [16] C. Demiroglu and T. P. Barnwell, "A Missing-Data Approach to Noise-Robust LPC Extraction for Voiced Speech Using Auxiliary Sensors," in Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005., vol. 1, pp. 773–776, doi: 10.1109/ICASSP.2005.1415228.
- [17] F. Z. Chelali and A. Djeradi, "Text dependant speaker recognition using MFCC, LPC and DWT," Int. J. Speech Technol., vol. 20, no. 3, pp. 725–740, 2017, doi: 10.1007/s10772-017-9441-1.
- [18] R. L. K. Venkateswarlu and R. V. Kumari, "Novel approach for speech recognition by using self-Organized maps," in 2011 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), Apr. 2011, pp. 215–222, doi: 10.1109/ETNCC.2011.5958519.
- [19] T. F. Li and S.-C. Chang, "Speech recognition of mandarin syllables using both linear predict coding cepstra and Mel frequency cepstra," in Proceedings of the 19th Conference on Computational Linguistics and Speech Processing, 2007, pp. 379–390, [Online]. Available: <https://www.aclweb.org/anthology/O07-2009/>.
- [20] I. Trabelsi and D. Ben Ayed, "On the use of different feature extraction methods for linear and non linear kernels," in 2012 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), Mar. 2012, pp. 797–802, doi: 10.1109/SETIT.2012.6482016.
- [21] M. A. Anusuya and S. K. Katti, "Comparison of Different Speech Feature Extraction Techniques with and without Wavelet Transform to Kannada Speech Recognition," Int. J. Comput. Appl., vol. 26, no. 4, pp. 19–24, 2011, doi: 10.5120/3092-4242.
- [22] H. Sak, A. Senior, K. Rao, and F. Beaufays, "Fast and accurate recurrent neural network acoustic models for speech recognition," Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH, vol. 2015-Janua, pp. 1468–1472, 2015.
- [23] S. L. Walker and S. Y. Foo, "Optimal Wavelets for Speech Signal Representations," Cybernetics, vol. 1, no. 4, pp. 44–46, 2003.
- [24] X. Tang, "Hybrid Hidden Markov Model and Artificial Neural Network for Automatic Speech Recognition," in 2009 Pacific-Asia Conference on Circuits, Communications and Systems, May 2009, pp. 682–685, doi: 10.1109/PACCS.2009.138.
- [25] Z. Tufekci, J. N. Gowdy, S. Gurbuz, and E. Patterson, "Applied mel-frequency discrete wavelet coefficients and parallel model compensation for noise-robust speech recognition," Speech Commun., vol. 48, pp. 1294–1307, 2006, doi: 10.1016/j.specom.2006.06.006.
- [26] A. Tavanaei, M. T. Manzuri, and H. Sameti, "Mel-Scaled Discrete Wavelet Transform and Dynamic Features for the Persian Phoneme Recognition," Int. Symp. Artif. Intell. signal Process., no. 4, pp. 138–140, 2011.
- [27] M. A. Katti and S. K. Anusuya, "Front end analysis of speech recognition : a review," Int. J. Speech Technol., vol. 14, no. 2, pp. 99–145, 2011, doi: 10.1007/s10772-010-9088-7.
- [28] S. Ganapathy, S. Thomas, and H. Hynek, "Modulation frequency features for phoneme recognition in noisy speech," J. Acoust. Soc. Am., vol. 125, no. 1, pp. EL8–EL12, 2009, doi: <https://doi.org/10.1121/1.3040022>.
- [29] H. Hermansky and N. Morgan, "RASTA processing of speech," IEEE Trans. Speech Audio Process., vol. 2, no. 4, pp. 578–589, 1994.
- [30] A. K. H. Al-Ali, D. Dean, B. Senadji, V. Chandran, and G. R. Naik, "Enhanced Forensic Speaker Verification Using a Combination of DWT and MFCC Feature Warping in the Presence of Noise and Reverberation Conditions," IEEE Access, vol. 5, pp. 15400–15413, 2017, doi: 10.1109/ACCESS.2017.2728801.
- [31] A. Winursito, R. Hidayat, and A. Bejo, "Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition," in 2018 International Conference on Information and Communications Technology (ICOIACT), Mar. 2018, pp. 379–383, doi: 10.1109/ICOIACT.2018.8350748.
- [32] A. E. Omer, "Joint MFCC-and-vector quantization based text-independent speaker recognition system," in 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE), Jan. 2017, pp. 1–6, doi: 10.1109/ICCCCEE.2017.7867612.
- [33] R. L. K. Venkateswarlu, R. V. Kumari, and G. V. Jayasri, "Speech recognition using Radial Basis Function neural network," in 2011 3rd International Conference on Electronics Computer Technology, Apr. 2011, pp. 441–445, doi: 10.1109/ICETECH.2011.5941788.
- [34] Z. Ping, T. Li-Zhen, and X. Dong-Feng, "Speech Recognition Algorithm of Parallel Subband HMM Based on Wavelet Analysis and Neural Network," Inf. Technol. J., vol. 8, no. 5, pp. 796–800, Jun. 2009, doi: 10.3923/itj.2009.796.800.
- [35] E. Mulliner, N. Malys, and V. Maliene, "Comparative analysis of MCDM methods for the assessment of sustainable housing affordability," Omega (United Kingdom), vol. 59, pp. 146–156, 2016, doi: 10.1016/j.omega.2015.05.013.