ISSN 2229-5518

Wavelet Energy based Satistical Learning Approaches to Vocoid Consonant Recognition

1 T. M. Thasleema, 2 B Sujith & 3 N K Narayanan
1,2 Department of Computer Science, Central University of Kerala, Kasaragod
3 Department of Information Technology, Kannur University, Kannur
Kerla, India
1 thasnitm1@gmail.com, 2 sujith.pallikkara@gmail.com, 3 nknarayanan@gmail.com

Abstract—State – of – the – art Automatic Speech Recognition (ASR) employs rigorous experimental evaluations on large, standard corpora from the real world. In recent years ASR and Machine Learning (ML) algorithms have had a great deal of influences on each other and feature selections can be considered as an essential task in ML. Compared with traditional basic speech feature extraction techniques, Wavelet Transform (WT) are highly capable of interpreting information content of the signal. This paper focuses on the recognition of Malayalam Vocoid Consonant (VC) speech units, a unique characteristic of the Malayalam language, using WT based Wavelet Energy (WE) parameters to capture the acoustical properties of each speech units. In the classification stage ML based on statistical approaches using with k – Nearest Neighbor (k – NN) is implemented. From the experimental results it is reported that k-NN algorithm can be perform well with wavelet family db5 compared with others in speaker independent environment.

Keywords – Autoamtic speech recognition, k- Nearest Neighor , Machine learning, Speaker independent environment, Wavelet Energy paramers, Wavelt Transform, Vocoid consonant recognition.

1 INTRODUCTION

State – of – art Automatic Speech Recognition (ASR) system has been implemented mostly by diligent empirical evaluations performed on collection of standard corpora from real world. In recent years ASR and Machine Learning (ML) communities have had larger impact on each other. However with prevailing techniques for ML, even if more computational and data resources are used in developing an ASR system, accuracy improvements is slowing down. New methods from ML promise tremendous improvements of ASR technology. This paper discusses the application of two different ML techniques based on statistical learning approaches applied for the recognition of Vocoid Consonats (VC) sounds in Malayalam in a speaker independent environment.

A unique characteristic of Malayalam include the existence of Vocoid Consonant (VC) which is derived from the basic Consonant – Vowel (CV) units. A Vocoid Consonant is a special consonant letter that represents a pure consonant independently, without help of a virama . A virama is a diacritic attached to a consonant letter to show that the consonant is not followed by an inherent vowel. Anusvara and visarga fit this definition but are not usually included. ISCII and Unicode 5.0 treat a Vocoid Consonant as a glyph variant of a base consonant letter. In Unicode 5.1 and later, however, Vocoid Consonant letters are treated as independent characters, encoded atomically.

For the present work all the experiments are carried out using 5 Vocoid Consonant in Malayalam and are tabulated in Table 1.

TABLE 1

VOCOID CONSONANT UNITS IN MALAYALAM

Sound/IPA	Unicode name	Bacic CV Sound
ൻ /N/	Chillu N	na ന
ൽ /l/	Chillu L	la ല
ൺ /NN/	Chillu NN	ṇa ണ
ർ /RR/	Chillu RR	ra ര
ൾ /LL/	Chillu LL	ḷa ള

There have been a lots of popular attempts carried out towards ASR which kept the research in this area vibrant [1][2][3][4]. Since human speech is highly dynamic in nature, in order to achieve a reliable representation of the speech signal in the time – frequency plane a multi resolution approach is needed. Wavelet Transform (WT) is a tool for Multi Resolution Analysis (MRA) which can be used to efficiently represent the speech signal in the time – frequency

---

- *Thasleema is currently working as an Assistant Professor, Department of Computer Science, Central University of Kerala, Kerala, India. Her specialization include automatic speech recognition, pattern recognition etc E-mail: thasnitm1@hotmail.com*
- *Sujith is currently working Department of Computer Science, Central University of Kerala, Kerala, India. E-mail: sujith.pallikkara@gmail.com*
- *Narayanan is a currently professor, Department of Information Technology, Kannur University, Kannur, Kerala, India. His area of research include digital signal processing, speech processing, image processing etc*

ISSN 2229-5518

plane. There have been lots of works reported in the literature using WT for the feature extraction process [5][6][7].

The present research work is motivated by the knowledge that a little attempts were rendered for the automatic speech recognition of CV speech unit in English, Hindi, Tamil, Bengali, Marathi Chinese etc. Recently very few research attempts were reported so far in the area of Malayalam vowel and consonant-vowel recognition [8][9][10]. But no works have been reported in the literature on VC speech unit recognition in Malayalam, which is the principal language of South Indian state of Kerala. So more basic research works are essential in the area of Malayalam VC speech unit recognition, since these sounds are important to structure the language and to make the word or sentence meaningful

The objective of the present work is to investigate on the use of MRA and statistical learning approaches such as k – Nearest Neighbor (k – NN) for the classification of Malayalam VC speech unit waveforms using WT based Wavelet Energy (WE) parameters. The rest of the paper organized as follows. Section II of this paper explains an overview on feature extraction using WT and and section II.A and II.B gives detailed description on Discrete Wavelet Transform (DWT) and feature extraction using WE parameters respectively. Section III explains a brief description on classification based on k-NN approaches. Section IV presents the simulation experiments and results obtained using k – NN approaches and finally section V gives conclusion and direction for future work.

2 Feature Extraction Using Wavelet Transform

Certain ideas of wavelet theory appeared quite a long time ago [11]. Over the last decades wavelet analysis has turned to be a standard technique in the areas of geophysics, meteorology, audio signal processing and image compression 12][13]. Wavelet transform can be defined as the transformation of the signal under analysis into another representation which presents the signal in a more useful form [14]. Mathematically a wavelet can be denoted as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) \quad (1)$$

Based on the definition of wavelet, wavelet transform of a signal f(t) can be mathematically represented as

$$W_{(a,b)} = \int_{-\infty}^{\infty} f(t)\psi_{a,b}\,dt \quad (2)$$

$$W_{(a,b)} = \int_{t=-\infty}^{\infty} f(t)\frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right)$$

ie (3)

Wavelet transforms are proved to be a very important and useful tool for signal and image processing [15]. Several applications of wavelet transform are already proposed in the literature[16][17]. Coifman and Maggioni had proposed wavelets based on diffusion operators in their work [18]. Hammond et al introduced wavelets on graph via spectral graph theory [19]. Kadambe has made a study on pitch detection algorithm for speech signal using wavelet transform [20]. O Farook et al had proposed in his research work the use of wavelet based feature extraction technique for Hindi phoneme recognition and proved that the proposed technique achieves a better performance over MFCC based features [21].

2.1 Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) has been treated as a Natural Wavelet Transform (NWT) for discrete time signals by different authors [22][23]. For computing the wavelet coefficients several discrete algorithms have been established [24]. Daubechies and some others had invented DWT peculiarly designed for analyzing finite set of observations over the set of scales using dyadic discretization [25][26]. As Daubechies mentioned in his work DWT can be interpreted as a discretization of Continuous Wavelet Transform (CWT) through sampling specific wavelet parameters.

$$DWT_{j,k} = CWT\left\{f(t); a = 2^j, b = k.2^j\right\}, \; j,k \in Z \quad (4)$$

Then wavelet in eqn (1) can be written as

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}(t-k)) \quad (5)$$

and eqn(3) becomes

$$W_{j,k}(t) = \sum_j \sum_k f(k)2^{-j/2}\psi(2^{-j}t-k) \quad (6)$$

The DWT which is based on subband coding is a fast computation of wavelet transform. The theory of sub band coding is introduced by Croisier et al for devising a technique to decompose discrete time signals [27]. Burt defined a similar technique to subband coding and named it as pyramidical coding which is also known as Multi Resolution Analysis (MRA) [28]. Later Vetterly had developed some improvements to the subband coding schemes, removing the existing redundancy in the pyramidical coding scheme [29].

2.2 Wavelet Energy based Subband Features

In this paper a sub-band feature extraction technique based on wavelet transform is proposed. The features were based on the energy in the frequency sub-bands obtained by the discrete wavelet decomposition. First the speech signal is decomposed using discrete wavelet transform, and total energy of the wavelet coefficients in each frequency sub-band is calculated. This is normalized by dividing total energy by the number of wavelet coefficients in the corresponding sub-band. Since the number of coefficients in a sub-band depends on the bandwidth obtained after decomposition, the above normalization procedure results into a non-uniform weighting of the energies. To see the effect of mother wavelet on feature extraction, Daubechies wavelets different mother wavelets are used.

- Calculate the energy of the wavelet coefficients in each sub-band.
- If cj;k is the jth wavelet coefficient in the kth sub-band then the total energy (Ep) in the sub-band p is given by:

$$E_p = \sum_{j=1}^{N} (C_{jp})^2 \text{ , j=1,2,……L} \tag{7}$$

$$F_p = \frac{E_p}{N_p} \text{ , p=1,2,…..L} \tag{8}$$

where Np is the number of wavelet coefficients in the pth sub-band and L is the number of sub-bands. The calculated energy is then divided by the number of wavelet coefficients in the corresponding sub-band, thereby giving average energy per wavelet coefficients per sub-band Fp.

To evaluate noise robust features, first calculate the average sub-band energy µ by using the equation given below.

$$\mu = \frac{1}{L} \sum_{j=1}^{N} F_p \text{ , p=1,2……..L} \tag{9}$$

The final features FFp are calculated by using the equation below:

$$FF_p = F_p - 0.5\mu \text{ , p=1,2…….L} \tag{10}$$

This gives L sub-band energy based features.

The variance based feature (VF) is extracted as follows

$$\mu = \frac{1}{L} \sum_{j=1}^{N} (F_p - \mu)^2 \text{ , p=1,2…….L} \tag{11}$$

This gives a total of L+1 features for each phoneme.
The method is repeated for different mother wavelets and the performance of the classifiers are evaluated in using these features.

3 Classification using k – NN

Pattern classification using distance function is an earliest concept in pattern recognition [30][31]. Here the proximity of an unknown pattern to a class serves as a measure of its classifications. k – NN is a well known non – parametric classifier, where a posteriori probability is estimated from the frequency of the nearest neighbors of the unknown pattern [32]. For classifying each incoming pattern k – NN requires an appropriate value of k. A newly introduced pattern is then classified to the group where the majority of k nearest neighbor belongs [33]. Hand proposed an effective trial and error approach for identifying the value of k that incur highest recognition accuracy. Various pattern recognition studies with highest performance accuracy are also reported based on these classification techniques. Consider the cases of m classes ci , i = 1,2,…….m, and a set of N samples pattern yi , i = 1,2,…..N whose classification is priory known. Let x denote an arbitrary incoming pattern. The nearest neighbor classification approach classifies x in the pattern class of its nearest neighbor in the set yi.

Ie If $||x-y_j||^2$ = min $||x-y_j||^2$, where 1≤ i ≤ N then x in c_j

This is 1 – NN rule since it employs only one nearest neighbour to x for classification. This can be extended by considering k – Nearest Neighbours to x and using a majority – rule type classifier. The simulation experiments and the results obtained using k-NN algorithm based on statistical approaches are explained in the next section.

4 Simulation Experiments and Results

All the simulation experiments are corroborated using Maalyalam VC speech unit database of 5 VC speech units uttered by both male and female 300 natively Malayalam speakers of age between 21 – 26 years. For the experimental study we have used 8 kHz sampled VC speech units which is low pass filtered to band limit 4 kHz. In the classification stage we have divided the dataset into training and testing sets which contains first 150 speakers speech samples for training and the last 150 speakers speech samples for testing. The average recognition accuracies obtained for Malayalam Vocoid Consonant speech unit database containing 5 VC speech units using WE parameters for different wavelet families based on k-NN are tabulated in table 2. The average recognition accuracies for each VC speech unit using WE parameters and k-NN algorithm are tabulated table 3. The recognition accuracies obtained using WE with k – NN for different values of k are given in table 4.

TABLE 2
RECOGNITION ACCURACY USING WE PARAMETER FOR DIFFERENT WAVELET FAMILIES.

SL. No	Wavelet Family	Average Recognition
		k- NN
1	Db2	68.6
2	Db4	74.2
3	Db5	**78.6**
4	Coif1	72.1
5	Coif2	72.6
6	Sym4	73.3
7	Sym6	69.3

TABLE 3
RECOGNITION ACCURACY OBTAINED FOR EACH VC UNIT IN DB5

SL. No	Speech unit	Average Recognition Using k-NN
1	ൻ	90.5
2	ൽ	85.6
3	ൺ	78.9

ISSN 2229-5518

4	ര	76.6
5	ശ	78.4

TABLE 4
RECOGNITION ACCURACY USING WE AND K - NN FOR DIFFERENT K AND DB5.

SL. No	K value	Average Recognition
1	2	**78.6**
2	3	75.7
3	4	60
4	5	45.3
5	6	28.6
6	7	23.3

5 CONCLUSION

This paper encapsulates the recognition of Vocoid Consonant speech units in Malayalam in a speaker independent environment. The simulation experiments conducted using statistical approaches based Machine Learning (ML) techniques are discussed. To evaluate the performance of the classifiers a Multi Resolution Approaches using Wavelet Transform (WT) is used. To capture the acoustical properties of each speech sound Wavelet Energy (WE) parameters are extracted. The recognition accuracies are calculated using k – Nearest Neighbor (k – NN) and it is observed that k-NN approaches can perform well with WE parameters in the speaker independent environment. Classification of Malayalam consonants inherent with all the short vowels and the recognition of spoken words including vocoid consonants are some of our future research direction.

REFERENCES

[1] Anil K Jain, Robert P W Duin and Jianchang Mao, "Statistical Pattern Recognition : A Review", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 22(1), Jan 2000.

[2] Forgie J W and Forgie C D, "Results obtained from a Vowel Recognition Computer Program", Journal of Acoustical Society of America, Vol. 31, pp. 1480 – 1489, 1959.

[3] Reddy D R, "An approach to Computer speech recognition by Direct Analysis of the speech wave", Computer Science Dept., Stanford University Technical Report No. C549, 1966.

[4] Gold B and Morgan N, "Speech and Audio Signal Processing", New York: John Wiley & Sons Inc., 2000.

[5] H. Bourlard and S. Dupont, "Subband-based speech recognition," in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSPSignal Processing (ICASSP '97), vol. 2, pp. 1251–1254, Munich, Germany, April 1997.

[6] M. Gupta and A. Gilbert, "Robust speech recognition using wavelet coefficient features," in Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU '01), pp. 445–448, Madonna di Campiglio, Trento, Italy, December 2001.

[7] R. Sarikaya, B. L. Pellom, and J. H. L. Hansen, "Wavelet packet transform features with application to speaker Identification," in Proceedings of the 3rd IEEE Nordic Signal Processing Symposium(NORSIG '98), pp. 81–84, Vigsø, Denmark, June 1998.

[8] T. M. Thasleema, P. Prajith & N. K. Narayanan, "Time domain Non-Linear Feature Parameter for Consonant Classification", International Journal of Speech Technology, Vol. 15(2), pp. 227 – 239, 2012

[9] N. K. Narayanan, T. M. Thasleema & P. Prajith, "Reconstructed State Space Model for Recogniton of Consonant – Vowel (CV) Utternaces using Support Vector Machines", International Journal of Artificial Intelligence & Application", Vol. 3(2), pp. 101 – 119, 2012.

[10] T. M. Thasleema & N. K. Narayanan, "Consoant Classification using Decision Directed Acyclic Graph Support Vector Machine Algorithm", International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 6(1), pp. 59 – 74, 2013.

[11] A Grossman, J Morlet & P Gaoupillaud, "Cycle octave and related transforms in seismic signal Analysis", Geoexploration, Vol. 23, pp. 85-102, 1984.

[12] C torrence and G P Compo, "A practical guide to Wavelet Analysis", Bull. Amer. Meteorological. Soc., Vol. 79(1), pp. 61 – 78 , 1998.

[13] Hongyu Liao, Mrinal Kr and Bruce F Cockburn, "Efficient Architectures for 1 – D and 2 – D Lifting Based Wavelet Transform", IEEE Trans. on Signal Processing, Vol. 52(5), pp. 1315 – 1326, 2004.

[14] K P Soman and K I Ramachandran, "Insight into Wavelets, from Theory to Practice", Prentice Hall of India, 2005.

[15] S Mallat, "A wavelet Tour of Signal Processing, The Sparse Way", NewYork : Academic, 2009.

[16] S Mallat, " A Theory for Multi resolution Signal Decomposition : The Wavelet Representation", IEEE Trans. on Pattern Analysis and machine Intelligence, Vol. 11, pp. 674 – 693 , 1989.

[17] R Kronland – Martinet, J Morlet and A Grossman, " Analysis of Sound Patterns through Wavelet Transforms", Int. Journal of Pattern Recognition, Artificial Intelligence, Vol. 1(2), pp. 273 – 302 , 1987.

[18] R R Coifman and M Maggioni, "Diffusion Wavelets", Appl. Computat. Harmon. Anal, Vol. 21(1), pp. 53 – 94 , 2006.

[19] D K Hammond, P Vandergheynst and R Gribonval, "Wavelets on Graph via Spectral Graph Theory, Appl. Computat. armon. Anal, 2010.

[20] Kadambe S and Boudreaux – Bartels G F, "Application of the Wavelet Transform for Pitch Detection of Speech Signal", IEEE Trans. on Information Theory, Vol. 38, pp. 917 – 924 , 1992.

[21] O Farook, S Dutta and M C Shrotriya, "Wavelet Sub band Based Temporal Features for Robust Hindi Phoneme Recognition", Int. Journal of Wavelets, Multi resolution and Information Processing (IJWMIP), Vol. 8(6), pp. 847 – 859, 2010.

[22] M Vetterly and C Herley, "Wavelets and Filter banks : Theory and Design", IEEE Trans. on Signal Processing, Vol. 40(9), pp. 2207 – 2232, 1992.

[23] M J Shensa, "Affine Wavelets: Wedding the Atrous and mallat Algorithms", IEEE Trans. on Signal Processing, Vol 40, pp. 2464 – 2482, 1992.

[24] S Mallat, " Multi frequency Channel Decomposition of Images and Wavelet Models", IEEE. Trans on Acoustics, Speech and Signal Processing, Vol. 37, pp. 2091 – 2110 , 1989.

[25] I Daubechies, "Ten Lectures on Wavelets", Philadelphia, PA:Soc. Appl. Math, 1992.

[26] Ronald W Lindsay, Donald B Percival and D Andrew Rothrock, "The Discrete Wavelet Transform and the scale analysis of the surface properties of Sea Ice", IEEE Trans. on Geo Science and Remote Sensing, Vol. 34(3), pp. 771 – 787 , 1996.

[27] A Croisier, D Esteban and C Galand, "Perfect Channel Splitting by use of Interpolation/decimation/tree decomposition Techniques", in International

Conference on Information Science and Systems, pp. 443 – 446 , Patras, 1976.

[28] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. on Communication, vol. COM-31, pp. 532-540, 1983.

[29] M Vetterly, " Filter Banks Allowing Perfect Reconstruction", IEEE Trans. on Signal Processing, Vol. 10(3), pp. 219 – 244,1986.

[30] Tou J. T and Gonzalez R. C, "Pattern Recognition Principles", Addison – Wesley, London, 1974.

[31] Friedmen M and Kandel A, "Introduction to Pattern Recognition: Statistical, Structural, Neural and Fuzzy Logic Approach", World Scientific, 1999.

[32] Cover T M & Hart P E, "Nearest Neighbor Pattern Classification", IEEE trans. on InformationTheory, Vol. 13 (1), pp. 21 - 27 , 1967.

[33] Min-Chun Yu, " Multi – Criteria ABC analysis using artificial – intelligence based classification techniques", Elsevier – Expert Systems With Applications, Vol. 38, pp. 3416 – 3421, 2011.