International Journal of Computer Theory and Engineering, Vol. 2, No. 4, August, 2010
1793-8201

Exploring the Discrete Wavelet Transform as a Tool for Hindi Speech Recognition

Shivesh Ranjan

***Abstract*—In this paper, we propose a new scheme for recognition of isolated words in Hindi Language speech, based on the Discrete Wavelet Transform. We first compute the Discrete Wavelet Transform coefficients of the speech signal. Then, Linear Predictive Coding Coefficients of the Discrete Wavelet Transform coefficients are calculated. Our scheme then uses K Means Algorithm on the obtained Linear Predictive Coding Coefficients to form a Vector Quantized codebook. Recognition of a spoken Hindi word is carried out by first calculating its Discrete Wavelet Transform Coefficients, followed by Linear Predictive Coding Coefficient calculation of these Coefficients, and then deciding in favor of the Hindi word whose corresponding centroid (in the Vector Quantized codebook) gives a minimum squared Euclidean distance error with respect to the word under test.**

***Index Terms*—discrete wavelet transform; linear predicitive coding; vector quantization; hindi ;speech recognition.**

I. Introduction

Hindi is the most widely spoken language in India, therefore, a speech recognition scheme for Hindi is expected to be of widespread use in diverse fields like railway ticket reservations, cellular phone based banking services, air-ticket reservations etc.

Our approach is primarily concerned with exploring a new feature extraction method using the Discrete Wavelet Transform (DWT) and the Linear Predictive Coding (LPC) coefficients calculation. We have avoided Hidden Markov Models (HMMs) [2,4] in our scheme as our main focus is confined to showing, how the features derived by applying the DWT ,can be used to recognize Hindi Speech. Earlier works on Hindi Speech recognition using wavelets [3] have employed linear prediction on the DWT coefficients too, but our approach does not involve calculation of linear prediction coefficients separately for the approximation and detail coefficients. In our scheme, we find the LPC coefficients of the DWT coefficients in a manner very similar to that used in finding the LPC coefficients of an actual speech signal [4].The use of DWT for speech recognition has also been investigated in [8].

To demonstrate our scheme for Hindi recognition, we first constructed a data base of 10 Hindi words (the numbers 1 through 10 in Hindi) sampled at 8KHz.Ten samples of each word were taken, Thus, a 100 words database was constructed.DWT and LPC analysis was carried out on each of the words, followed by K-Means Algorithm [1,4] to form a 10 entries VQ codebook. A different 100 samples set (of the same ten words) was taken and recognition was then attempted for each of the words in the test sample. Thus, a total of 100 recognition attempts were made. All the Hindi speech samples: both for forming the database for constructing the VQ codebook, and the 100 samples of the ten Hindi words to be tested, were taken from the same speaker (An adult male native speaker.)

Shivesh Ranjan is with Electronics and Communication Engineering, Manipal Institute of Technology, Manipal 576104, India(email: srjn1@yahoo.com).

II. The Discrete Wavelet Transform

The DWT can be used for Multi Resolution Analysis (MRA) [5,6],where a given signal is decomposed into what are known as the approximation and detail coefficients . A given function f(t) satisfying certain conditions [5], can be expressed through the following representation

$$f(t)=\sum_{j=1}^{L}\sum_{K=-\infty}^{\infty} d(j,k)\varphi(2^{-j}t-K)+\sum_{K=-\infty}^{\infty} a(L,K)\theta\left(2^{-L}t-K\right)$$

Where φ(t) is the mother wavelet and θ (t) is the scaling function. a(L,k) is called the approximation coefficient at scale L and d(j,K) is called the detail coefficient at scale j.The approximation and detail coefficients can be expressed as

$$a(L,K)=\frac{1}{\sqrt{2^{L}}}\int_{-\infty}^{\infty} f(t)\,\theta(2^{-L}t-K)dt$$

$$d(j,K)=\frac{1}{\sqrt{2^{j}}}\int_{-\infty}^{\infty} f(t)\varphi(2^{-j}t-K)dt$$

Based on the choice of the mother wavelet φ(t) and scaling function θ(t), different families of wavelets can be constructed[5,6,9,10].We used three distinct families of DWTs namely: the Daubechies wavelets (db), the Discrete Meyers wavelets (dmey) and the Coiflets (coif) in our recognition scheme.

TABLE I.

NUMBER	HINDI WORD	SYMBOL USED IN THE PAPER
1	“ek”	one
2	“do”	two
3	“teen”	three
4	“char”	four
5	“paanch”	five
6	“chhae”	six
7	“saat”	seven
8	“aath”	eight
9	“nau”	nine
10	“dus”	ten

III. Speech Database Construction And DWT Coeficients Computaion

A. Constrction of Database

An adult male, native speaker of Hindi was asked to utter the Hindi words (1 through 10, see Table-1), and his voice was sampled at 8KHz.The speech signal of each word was then isolated from silence. The samples were then stored in ascending order: first, the ten samples corresponding to word one ("ek") were stored, then the ten samples of two and so on.

B. Caculating the DWT approximation and detail coefficients

Each of the 100 speech samples were then decomposed into approximation and detail coefficients using DWT. Five different sets of decomposition were carried out on each of the 100 speech samples, using 5 different DWTs (of 3 different wavelet families). Of the five different decomposition sets, three sets of decompositions were carried out using the Daubechies wavelets as they have been reported to be highly successful in speech compression schemes using wavelets [7].A single decomposition of the 100 samples was performed using each of the remaining two wavelet families: Coiflets and discrete Meyer wavelets.Fig.1 shows this decomposition process. The DWTs and their symbols used in this paper are

- Daubechies Wavelets [9]
  Daubechies8, 3-Level decomposition (db8, Lev3)
  Daubechiesb8, 5-Level decomposition (db8, Lev5)
  Daubechies10, 5-Level decomposition (db10, Lev5)
- Coiflets (coif) [9]
  Coiflets5, 5-Level decomposition (coif5, Lev5)
- Discrete Meyer Wavelets [10] : Discrete Meyer, 5-Level decomposition (demy Lev5)

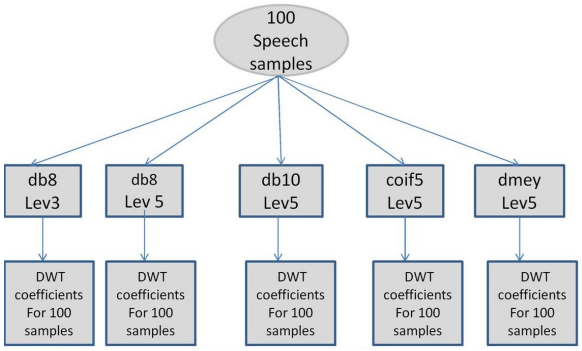

Figure 1. Decomposition of speech signal using DWTs

IV. Formation of VQ Codebook and Testing

We obtained five sets of DWT coefficients from the previous step. Each of these sets had 100 entries. Each entry was actually the collection of DWT coefficients of the speech signal from which it was derived.

A. Computing LPC coefficients from the approximation and detail coefficients

To compute LPC coefficients from the DWT coefficients, we employed a method similar to that used for finding LPC coefficients of speech signals [4].

- The DWT coefficients of each speech signal were arranged in descending order, starting from the corresponding highest level approximation coefficients, followed the same level's detail coefficients ,followed subsequently by lower levels detail coefficients in descending order.
- Then, the DWT coefficients were framed into frames of 160 samples in length.
- Overlap between successive frames was kept at 80 samples
- Each frame was multiplied by a 160 point Hamming window. No pre-emphasis [4] was done unlike the speech signals.
- The 10th order LPC coefficients for each frame were found, and the whole process was carried on the first 9 frames of the DWT coefficients (for each speech signal .Thus, we obtained a total of 90 LPC coefficients, ten for each frame (of length 160 samples) for the DWT coefficients of a single speech signal .In effect, we used the first 900 terms from each of the 100 DWT coefficients.

At the end of this stage, we got five sets of LPC coefficients, each of which had 100 rows (corresponding to the 10 utterances of each word) and 90 columns (corresponding to the 90 LPC coefficients derived from the DWT coefficients of each speech signal). These sets were then used to construct their respective database, which was to be used in recognition (discussed later).

B. Using K-Means algorithm to form a VQ codebook

We chose to use K-means Algorithm [1,4] to perform clustering of the LPC coefficients (computed in the previous stage) into ten clusters, in order to form the VQ Codebook. The algorithm clustered the points in the 100 by 90 LPC coefficients matrix into ten clusters, and returned the index and cluster centroid location for each of the 100 entries in the matrix. Since our recognition scheme relied on the K-Means Algorithm for recognition and, we did not use HMMs in the later stages for recognition, we proposed the following algorithm to form a VQ codebook

As the order of entries in the 100 by 90 LPC coefficients matrix was known (the first 10 entries corresponded to the word one, the next ten for two and so on), we used this information, to our advantage in forming a VQ code book

- Starting from the first ten indices, returned by the K-Means algorithm, we choose the index appearing the largest number of times in the group (of ten) as the index of the group. The corresponding centroid was designated the centroid of the group.
- The same process was repeated for the next ten indices. We continued this till all 10 groups (i.e. a total of 100 entries) were assigned to ten different groups. This simple algorithm would have failed, if a certain index were in majority in more than one group, because then, the assigning of index to both the groups would have become ambiguous. However, such a situation was not observed. So, unique indices were assigned to each of the ten groups (group1 through group10). In fact, we did have a conflict resolution scheme to form a VQ table, for the simplified case, when a given index appeared in majority in two groups. However, for the more complex case of more than two, the simple conflict resolution scheme that we present (later), will not work. However, such an ambiguous case is least expected to occur as we did not encounter any.

Group1 corresponded to the first ten entries of the LPC

coefficients matrix, which were actually the LPC coefficients derived from the first ten original speech signals' DWT coefficients .So we used the index of group 1 as the index for the Hindi word one (i.e. "ek") and the corresponding centroid was identified as the centroid of the cluster in which LPC coefficients derived from the word one lay. Similarly, we obtained the indices and centroids for each of the remaining nine Hindi words (two through ten).This information was then used to form a VQ table with the corresponding Hindi word as its index and the related centroid as its content.

C. *Conflict Resolution Scheme to form a VQ Codebook*

Consider the case when the same index appears in majority in more than one group. As mentioned previously, it becomes difficult to assign unique indices to the different groups in such a case. To overcome this, we propose the following simple approach.

- Case 1: The same index appears in majority in two, but its distribution is unequal. In other words, the number of times the index (which is in majority), appears in the two groups is not equal. For this particular case, we can resolve the ambiguity by simply assigning the index to that particular group, in which the index appears the more number of times.
- Case 2: The same index appears in majority in two groups, and the distribution in both of them is equal. In this case, the assigning of the index to a particular group is arbitrary. Once an index is assigned to a particular group, the other group is searched for the index that appears the second largest number of times. This index is then assigned as the index of the group.

We were also tempted to test our approach for forming the VQ code book from the DWT coefficients themselves, rather than using the algorithm on the LPC coefficients derived from them. To this end, we attempted to run our codebook formation algorithm on the DWT coefficients derived from the speech signals.

Much to our disappointment, we found that the algorithm failed completely on the DWT coefficients. In fact, all the DWT coefficients of the 100 samples failed to get grouped in ten clusters with ten different indices in majority. Surprisingly, all of them had a few indices in majority in all the ten groups, making the assignment of a particular index to a group, virtually impossible. This ruled out any possibility of using the DWT coefficients directly, since our approach relied on obtaining a ten entry VQ code book. Table 2 shows the number of distinct indices that were in majority in the ten groups, according to our simple rule of assigning an index to a group, as discussed previously, ruling out any possibility of performing recognition without finding the LPC coefficients of the DWT coefficients. To sum up, we needed ten distinct indices to recognize ten different words, while this approach assigned just a single index to all the ten groups! So, it was rejected.

TABLE II.

DWT	db8 Lev3	db8 Lev5	db10 Lev5	coif5 Lev5	dmey Lev5
Number Of distinct indices assigned to the 10 groups	1	1	1	1	1

D. *Testing isolated words using the centrods and indices of the VQ Codebook*

To test a given Hindi word, we first found its DWT coefficients, then, the LPC coefficients of the DWT coefficients were found. The LPC coefficients were then matched with each entry in the VQ table and a decision was made in favor of the index, the content (i.e. the centroid) of which gave minimum squared Euclidean Distance with respect to the word under test.

We tested 10 different samples of each Hindi word. Thus, a total of 100 samples were tested for each of the five DWT types. Fig. 2 shows the overall recognition scheme employed in our approach, taking db8 Lev3 (Daubechies 8, 3 Level) decomposition as an example.

Similarly, recognition was carried out for each of the four remaining DWT decompositions based approaches, and the performance in each case was noted.

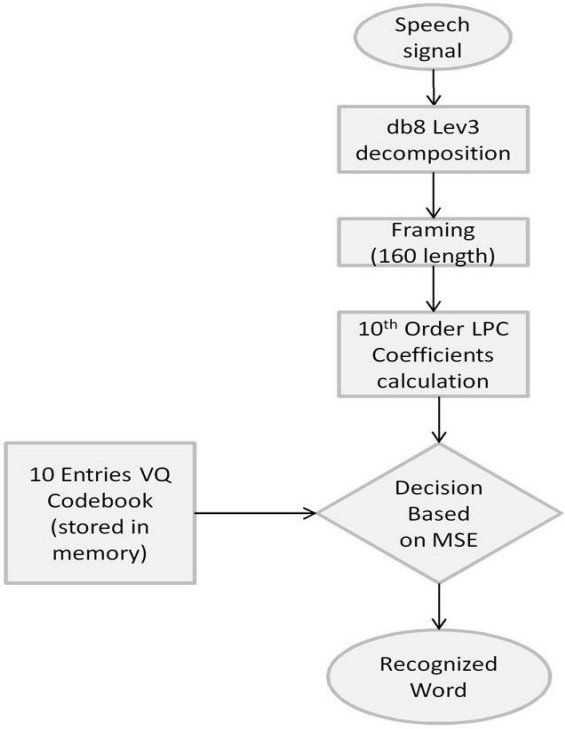

Figure 2. Recognition using db8 Lev3 decomposition

E. *Effect of number of terms of the DWT coefficients (used in LPC coefficients) on recognition.*

We were also interested in observing if varying the number of terms (of the DWT coefficients) that were used in the LPC coefficient calculation, had any effect on the overall recognition process.

For this, we used the same procedure, but with the difference that instead of finding the 90 LPC coefficients from the first 800 samples of each of the DWT coefficients, we utilized the first 1600 terms for the LPC coefficients.

In this case, for each speech signal, we obtained a 190 element row vector (19 frames) after the LPC coefficients calculation stage. Thus, for a total of 100 speech signals, we obtained a matrix of 100 by 190 entries. Everything else remained similar to what we have already discussed for the case when we took the first 800 samples only (as in IV.B).

V. Results

TABLE III.

Hindi Word	db8 Lev3	db8 Lev5	db10 Lev5	coif5 Lev 5	dmey Lev5
One (“ek”)	90	70	50	70	70
Two(“do”)	90	100	100	70	100
Three(“teen”)	100	80	90	90	60
Four(“char”)	50	100	80	100	90
Five(“panch”	80	100	100	100	90
Six(“chhae”)	30	60	70	70	70
Seven(“saat”)	70	90	90	100	80
Eight(“aath”)	50	60	30	20	60
Nine (“nau”)	80	90	90	90	50
Ten(“dus”)	90	100	90	50	100

Table 3 shows the success percentage of each of the five types of DWT based approaches, when we used the first 800 samples to form the 90 LPC coefficients vector (for each word).

Table 4 shows the results when we took the first 1600 samples, and formed a 190 elements row vector of LPC coefficients, for the particular cases of db8 Lev3 (Daubechies8, 3 level) and db8, Lev 5 based decompositions.

We would like to emphasize that other DWTs are also expected to give different results when the number of samples vary; we chose these two, just to examine the nature of the effect.

TABLE IV.

Hindi Word	db8 Lev3 (1600 Terms)	db8 Lev5 (1600 Terms)
One (“ek”)	90	70
Two(“do”)	100	100
Three(“teen”)	100	90
Four(“char”)	100	100
Five(“panch”	100	100
Six(“chhae”)	90	90
Seven(“saat”)	90	90
Eight(“aath”)	80	90
Nine (“nau”)	100	80
Ten(“dus”)	100	100

Fig.3 shows the average success percentage of each of the five DWTs based recognition scheme.Fig.4 shows the average success percentage of recognition of each of the ten words. Note that both of these correspond to the original case, when we had taken 800 terms of the DWT coefficients to find the 10^{th} order LPC coefficients.

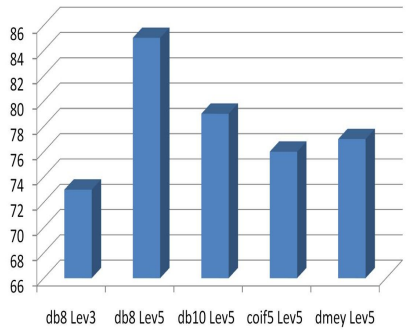

Figure 3. Suceess of different DWTs

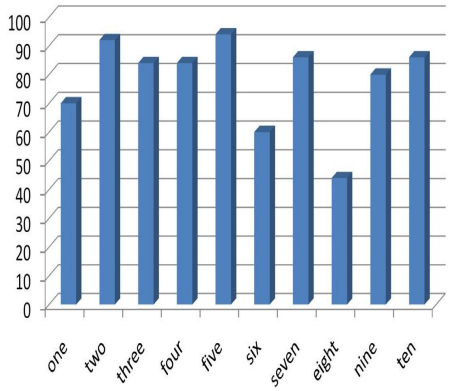

Figure 4. Success of Individual Words

To, appreciate the effect of increased number of samples on the recognition of words ,Fig. 5 shows the percentage increase in performance for each of the two cases in which, double the original number of samples were used (i.e. 1600 samples).

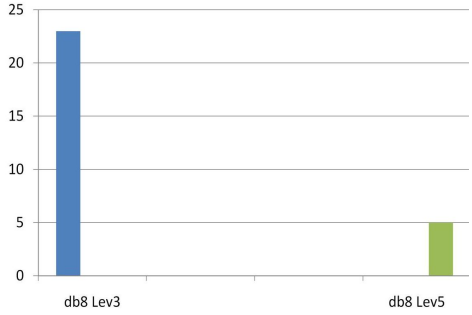

Figure 5. Percentage increase in performance

VI. Conclusion

As seen from the results, when working with the first 800 samples of the DWT coefficients to compute LPC coefficients, the Daubechies8, 5-level decomposition gave the highest percentage of success in the recognition of Hindi Speech. Clearly, it emerges as the candidate of choice for our DWT based speech recognition scheme. Daubechies10, 5-level decomposition and the Discrete Meyer wavelets gave comparable performance, while the Daubechies8, 3-level decomposition gave the poorest performance. The recognition of word eight (“aath”) had the poorest success percentage of getting recognized in this approach.

Doubling the number of samples had a very positive impact on the performance, in fact, the recognition by db8 Lev3 increased by an overwhelming 23 percent. However, it should be kept in mind that the price paid for this increase in performance, was an overall increase in computational complexity. Also, important is the fact that the recognition of the word eight also improves greatly, suggesting that the DWT coefficients in the first 800 samples of the word were relatively inefficient in recognition of this word.

This paper aimed at exploring the DWTs as a tool for recognition of Hindi Speech. But, our main focus in this paper was to identify the type of DWT, which would most likely give superior performance over other DWT types in speech recognition. The Recognition approach in this paper, after the feature extraction stage is clearly not very robust, as we have tried to keep our approach limited to identifying

the best possible wavelets family which can be used for DWT based Hindi Speech recognition.

It should also be observed that this approach can be used for recognition of speech in other languages as well. Any modification to our scheme to make it speaker independent will require taking a large number of utterance samples from speakers of different age groups, gender, accent etc. Then, the data (index and centroid) obtained after applying the K-Means algorithm on the LPC coefficients, can be used to train HMMs [2,4] and an HMM based speech recognition scheme can be employed[2].Such a scheme is expected to give good performance for speaker independent speech recognition.

REFERENCES

[1] J. MacQueen,"Some methods for classification and analysis of multivariate observations"Proc. Of Fifth Berkeley Symposium on Mathematical Statistics and Probability , June 21-July 18, 1965 and December 27, 1965-January 7, 1966,pp. 281-297

[2] B. H. Juang, L. R. Rabiner, S. E. Levinson and M. M. Sondhi, "Recent Develoments in the applicatiron of Hidden Markov Models to Speaker-independent Isolated Word Recognition", Proc. IEEE International Confrence on Accoustics,Speech and Signal Processing,March 1985,pp. 9-12

[3] Aditya Sharma,M. C. Shrotiya,Omar Farooq,Z. A. Abbas," Hybrid wavelet based LPC features for Hindi speech recognition", International Journal of Information and Communication Technology,2008,pp. 373-381

[4] Lawrence R. Rabiner, B. H. Juang,Fundamentals of Speech Recognition,2nd Indian Reprint, Pearson Education,Delhi,1993, pp. 133-167,357-422

[5] Gilbert Strang and Truong Nguen,Wavelets and Filter Banks.Wellesley-Cambridge Press,MA,1997,pp. 174-220,365-382

[6] Andrew K. Chan and Jaideva C. Goswami, Fundamentals of Wavelets,Wiley-India Edition,John Wiley & Sons Inc,New Delhi,1999, pp. 89-97

[7] Nikhil Rao, Speech Compression Using Wavelets, ELEC 4801 THESIS PROJECT.School of Information Technology and Electrical Engineering,The University of Queensland,October 2001

[8] Brian Gamulkiewicz and Michael Weeks," Wavelets based SpeechRecognition "Proc. IEEE Internatioanl Symposium on Micr-NanoMechatronics of Human Science,Dec.2003,pp. 678-681 Vol. 2,doi: 10.1109/MWSCAS.2003.1562377.

[9] Ingrid Daubechies,Ten Lectures on Wavelets,SIAM,1992,pp. 115-132,194-292,258-259

[10] Martin Vettereli and Jelena Kovacevic,Wavelets and Subband Coding,Prentice Hall,1995,pp. 233-238