Nahrain University, College of Engineering Journal (NUCEJ) Vol.14 No.1, 2011 pp.34-45

Spoken Word Recognition Using Slantlet Transform and Dynamic Time Warping

Dr. Sadiq J. Abou-Loukh
University of Baghdad, College of Engineering, Electrical Eng. Dept

Samah Mutasher Gatea
University of Baghdad, College of Engineering, Electrical Eng. Dept

Abstract:

Speech recognition system has been widely used by many researchers using different methods to fulfill a fast and accurate system. Speech signal recognition is a typical classification problem, which generally includes two main parts: feature extraction and classification. In this work, three feature extraction methods, namely SLT, DWT Db1 and DWT Db4, were compared. The dynamic time warping (DTW) algorithm is used for recognition. Twenty three Arabic words were recorded fifteen different times in a studio by one speaker to form a database. The proposed system was evaluated using this database. The result shows recognition accuracy of 93.04%, 92.17% and 94.78% using DWT Db1, DWT Db4 and SLT respectively.

Keywords: Speech Signal Recognition, Slantlet Transform, Dynamic Time Warping, Discrete Wavelet Transform.

1. Introduction

Speech signals are composed of a sequence of sounds. These sounds and the transitions between them serve as a symbolic representation of information. The arrangement of these sounds (symbols) is governed by the rules of language. The study of these rules and their implications in human communication is the domain of linguistic, and the study and classification of the sounds of speech is called phonetics. Speech can be presented in terms of its message content or information. An alternative way of characterizing speech is in terms of the signal carrying the message information, i.e., the acoustic waveform (Rabiner, 1978). Speech is one of the most important tools for communication between human and his environment, therefore manufacturing of Automatic System Recognition (ASR) is desired for him all the time. The task of a speech recognizer is to determine automatically the spoken words, regardless of the variability introduced by speaker identity, manner of speaking, and environmental conditions (Abbas, 2005 and Pour, 2009).
The general block diagram of the speech recognition system is presented in **Fig. 1**

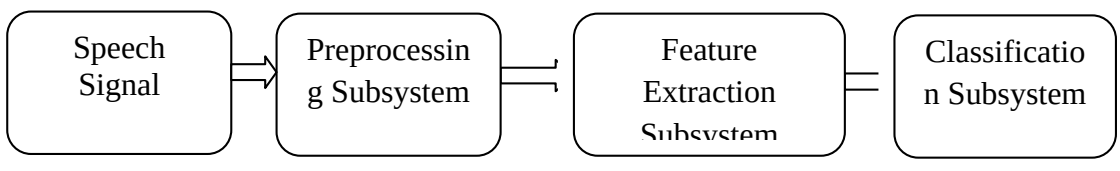


Fig. 1: General block diagram of the speech recognition system

Basic speech recognition system includes three stages, namely the pre-processing stage, the feature extraction stage, and the classification stage.
The preprocessing stage includes sampling, framing, and windowing processes:

a. Sampling: the speech signal is sampled with suitable sampling rate to convert it from analog to digital.

b. Framing and windowing: due to physical constraints, the vocal tract shape generally changes fairly slowly with time and tends to be fairly constant over short intervals (around 10–20 ms). A reasonable approximation is therefore to analyze

the speech signal into a sequence of frames**,** where each frame is represented by a single feature vector describing the average spectrum for a short time interval. Prior to any frequency analysis, each section of signal is multiplied by a tapered window (usually a Hamming window). This type of windowing is necessary to reduce any discontinuities at the edges of the selected region, which would otherwise cause problems for the subsequent frequency analysis by introducing spurious high-frequency components into the spectrum. Equation (1) defines the window function (Amin, 2008). The window mostly used is the hamming window which is given in eq. (2) (Lipeika, 2002)

$$X_1(n) = x_1(n)w(n) \qquad 0 \le n \le N-1 \qquad (1)$$

$$w(n) = 0.54 - 0.46\cos\frac{2\pi n}{N-1} \qquad 0 \le n \le N-1 \qquad (2)$$

Where:
$x_1(n)$ is the frame of speech signal and w (n) is window function

In this paper a slantlet based approach was used to extract the features from speech signal. Dynamic time warping (DTW) algorithm was used to recognize the applied input speech.

2. Discrete Wavelet Transform

The discrete wavelet transform -DWT- employs two sets of functions, called scaling functions and wavelet functions, which are associated with low pass and high pass filters, respectively (Arivazhagan, 2007).
The discrete wavelet transform(DWT) is defined by the following equation (Burrus, 1998).

$$g(t) = \sum_k c_{j0}(k)2^{j0/2}\varphi(2^{j0}t-k) + \sum_k\sum_{j=j0}^{\infty} d_j(k)2^{j/2}\psi(2^j t-k) \qquad (3)$$

Where:
$\varphi(t)$'s are the scaling function , $\psi(t)$'s are the wavelet function, k is the time translation index and j is the scale parameter.

The DWT is a linear transformation that operates on a data vector whose length is an integer power of two, transforming into a numerically different vector of the same length. It is a tool that separates data into different frequency components, and then studies each component with resolution matched to its scale. DWT is computed with a cascade of filtering followed by a factor 2 down sampling. (Kociolek, 2001).

The DWT analyzes the signal at different frequency bands with different resolutions by decomposing the signal into a coarse approximation and detail information. The decomposition of the signal into different frequency bands is simply obtained by successive high-pass and low-pass filtering of the time domain signal. The original signal x[n] is first passed through a half band high-pass filter g[n] and a low-pass filter h[n] as shown in **Fig. 2.**

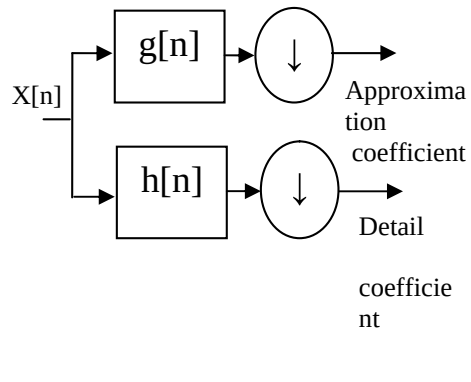


Fig. 2: Approximation and detail coefficients using DWT

After each filtering, half of the samples/data can be segregated (or eliminated) as per the Nyquist's rule. Since the high-pass filtered signal has now the highest frequency of π/2 radians instead of π, the signal can therefore be down sampled by 2. This constitutes one level of decomposition. Output from the high pass filter is downloaded as the level 1 detail D1, and the output from the low pass filter becomes the level 1 Approximation, A1. Starting a fresh with A1, the process can be successively repeated as per requirements (Sahu, 2007). This filtering and eliminating can be repeated on the scaling coefficients to give the two scale structure Repeating this on the scaling coefficients is called iterating the filter bank. Iterating the filter bank again gives the three-scale structure as shown in **Fig.3**(Burrus,1998):

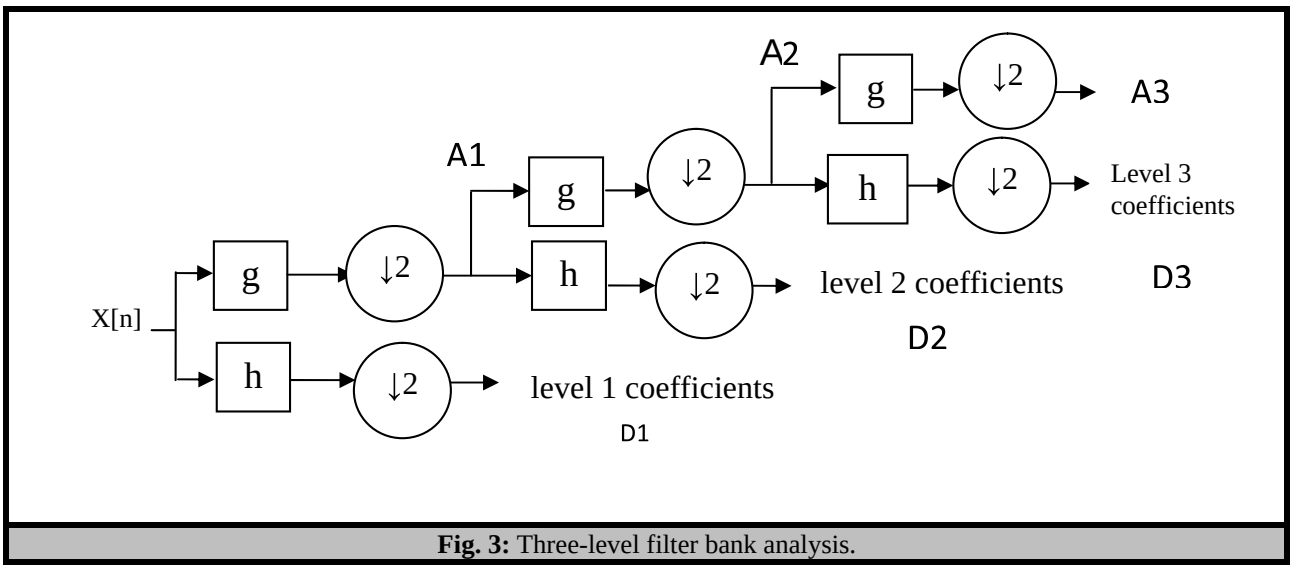
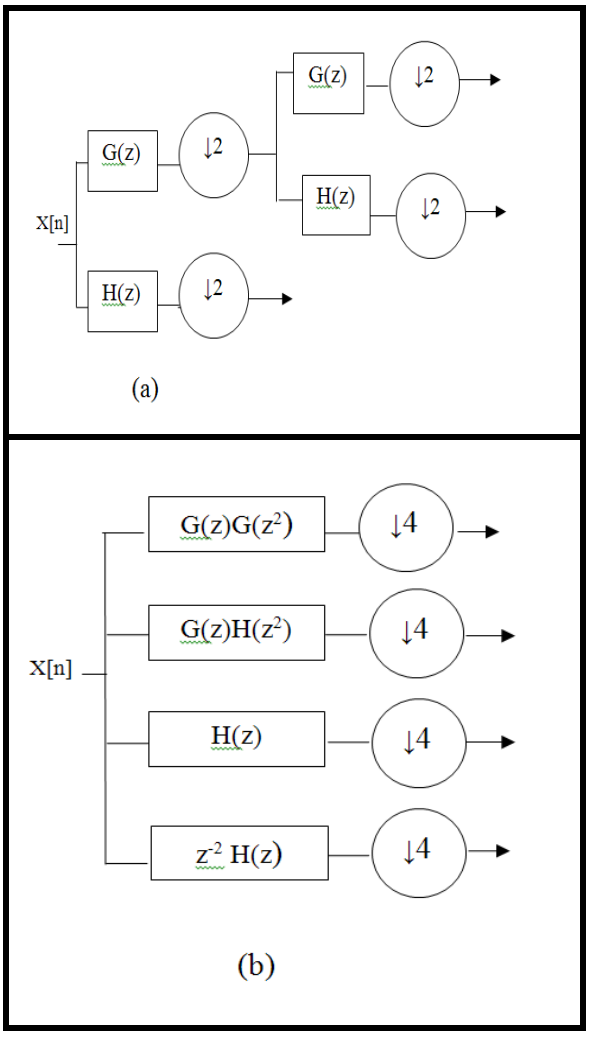


Fig. 3: Three-level filter bank analysis.

3. Slantlet Transform

The slantlet transform (SLT) is based on an improved version of the usual DWT, where the support of the discrete-time basis functions is reduced (Chatterjee, 2009). The slantlet transform is an orthogonal discrete wavelet transform with two zero moments and with improved time localization, the basis of the slantlet is based on a filter bank structure where different filters are used for each scale. Consider a usual two-scale iterated DWT filter bank shown in **Fig. 4** (a) and its equivalent form **(Fig. 4** (b)). The slantlet filter bank is based on the structure of the equivalent form shown in **Fig. 4** (b), but it is occupied by different filters that are not products. With this extra degree of freedom obtained by giving up the product form, filters of shorter length are designed satisfying orthogonality and zero moment conditions (Selesnick, 1999). For two-channel case the Daubechies filter is the shortest filter which makes the filter bank orthogonal and has K zero moments. For K=2 zero moments the iterated filters of **Fig. 4** (b) are of lengths 10 and 4 but the slantlet filter bank with K=2 zero moments shown in **Fig. 4** (c) has filter lengths 8 and 4. Thus the two-scale slantlet filter bank has a filter length which is two samples less than that of a two-scale iterated Daubechies-2 filter bank. This difference grows with the increased number of stages. Some characteristic features of the slantlet filter bank are orthogonal, having two zero moments and has octave-band characteristic. Each filter bank has a scale dilation factor of two and provides a multiresolution decomposition. The slantlet filters are piecewise linear. Even though there is no tree structure for slantlet it can be efficiently implemented like an iterated DWT filter bank. Therefore, computational complexities of the slantlet are of the same order as that of the DWT, but slantlet transform gives better performance in denoising and compression of the signals (Selesnick, 1999).



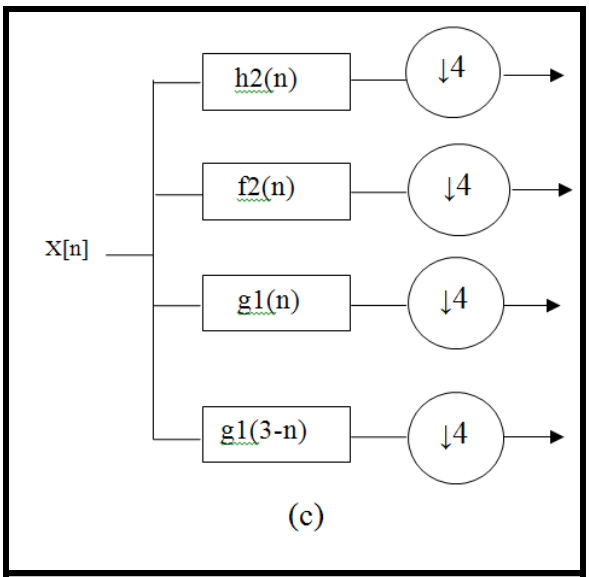


Fig. 4: (a) Two-scale iterated filter bank DWT
(b) Equivalent form using the D
(c) Two-scale filter bank using SLT

3.1 Derivations of Slantlet Filters Coefficients

The filters that construct the slantlet filter bank are g_i (n), f_i (n), and h_i (n). The L-scale filter bank has 2L channels. The low-pass filter is to be called h_L (n). The filter adjacent to the low-pass channel is to be called f_L (n). Both h_L (n) and f_L (n) are to be followed by downsampling by 2^L. The remaining 2L-2 channels are filtered by g_i (n) and its shifted time-reverse for i=1,…, L-1. Each is to be followed by downsampling by 2^{i+1} (Selesnick, 1999).
The sought filter $g_i(n)$ is described by four parameters and can be written as

$$g_i(n)=\begin{cases} a_0+a_1 n & for\ n=0,..,2^i-1 \\ b_0+b_1(n-2^i) & for\ n=2^i,..,2^{i+1}-1 \end{cases} \quad (4)$$

To obtain g_i (n) such that the sought L-scale filter bank is orthogonal with 2 zero moments, requires obtaining parameters a_0, a_1, b_0, b_1 such that

$$m = 2^i$$
$$s_1 = 6\sqrt{m/((m^2-1)(4m^2-1))}$$
$$t_1 = 2\sqrt{3/(m.(m^2-1))}$$
$$s_0 = -s_1.(m-1)/2$$
$$t_0 = ((m+1).s_1/3 - mt_1)(m-1)/(2m)$$
$$a_0 = (s_0+t_0)/2$$
$$b_0 = (s_0-t_0)/2$$
$$a_1 = (s_1+t_1)/2$$
$$b_1 = (s_1-t_1)/2$$

The same approach work for f_i (n) and h_i (n).

$$h_i(n)=\begin{cases} a_0+a_1 n & for\ n=0,..,2^i-1 \\ b_0+b_1(n-2^i) & for\ n=2^i,..,2^{i+1}-1 \end{cases} \quad (5)$$

$$f_i(n)=\begin{cases} c_0+c_1 n & for\ n=0,...,2^i-1 \\ d_0+d_1(n-2^i) & for\ n=2^i,...,2^{i+1}-1 \end{cases} \quad (6)$$

Where

$$m = 2^i$$
$$u = 1/\sqrt{m}$$
$$v = \sqrt{(2m^2+1)/3}$$
$$a_0 = u.(v+1)/(2m)$$
$$b_0 = u.(2m-v-1)/(2m)$$
$$a_1 = u/m$$
$$b_1 = -a1$$
$$q = \sqrt{3/(m.(m^2-1))}/m$$
$$c_1 = -q.(v-m)$$
$$d_1 = -q.(v+m)$$
$$d_0 = d_1.(v+1-2m)/2$$
$$c_0 = c_1.(v+1)/2$$

4. Dynamic Time Warping Algorithm

The definition of dynamic time warping (DTW) is based on the notion of warping path (Chang, 2008). The DTW algorithm which is based on dynamic programming finds an optimal match between two sequences of feature vectors by allowing for stretching and compression of sections of the sequences (Shanker, 2007).

Suppose there are two numerical sequences (a_1,a_2, ..., a_n) and (b_1, b_2, ..., b_m). As we can see, the length of the two sequences can be different. The algorithm starts with local distances calculation between the elements of the two sequences using different types of distances. The most frequent used method for distance calculation is the absolute distance between the values of the two elements (Euclidian distance). Those results stored in a matrix of distances with n lines and m columns of general term (Furtunà, 2008):

$$d_{ij} = |a_i - b_j|, \quad i = 1 \to n, \quad j = 1 \to m \qquad (7)$$

Starting with local distances matrix, then the minimal distance matrix between sequences is determined using a dynamic programming algorithm and the following optimization criterion:

$$Y_{ij} = d_{ij} + \min\left(Y_{i-1,j-1}, Y_{i-1,j}, Y_{i,j-1}\right) \qquad (8)$$

where Y_{ij} is the minimal distance between the subsequences (a_1,a_2, ..., a_i) and (b_1, b_2, ..., b_j).
A warping path is a path through minimal distance matrix from Y_{11} element to Y_{nm} element consisting of those Y_{ij} elements that have formed the Y_{nm} distance.
The global warp cost of the two sequences is defined as shown below (Furtunà, 2008):

$$GC = \frac{1}{p}\sum_{i=1}^{p} w_i \qquad (9)$$

where w_i are those elements that belong to warping path, and p is the number of them.

5. Proposed Speech Recognition System

The general block diagram for the proposed model of speech recognition system is shown in **Fig. 5:**

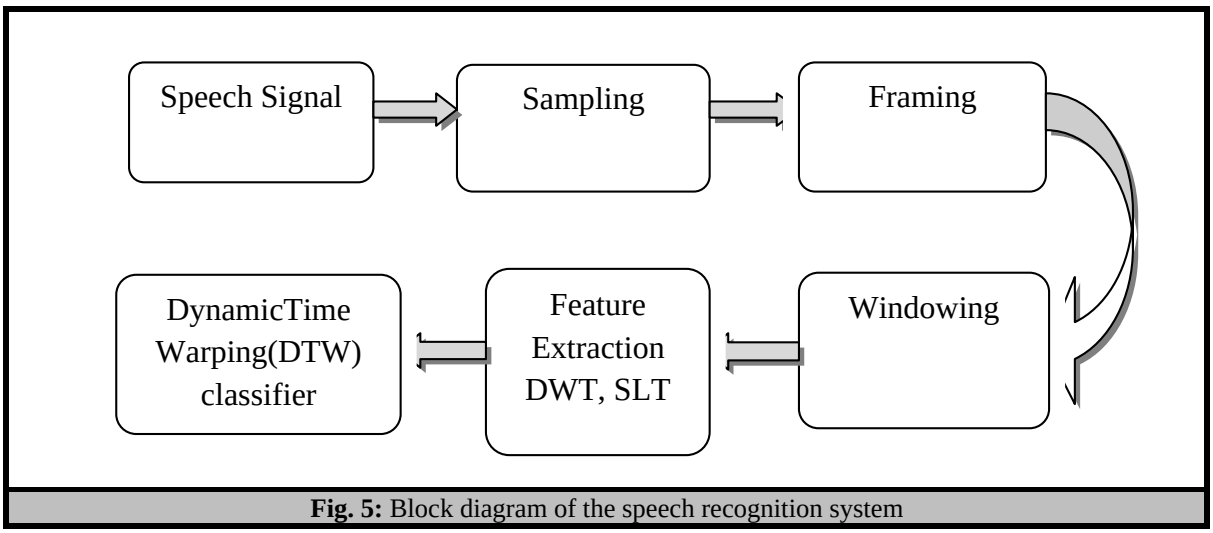


Fig. 5: Block diagram of the speech recognition system

5.1 Speech Signal

The speech signal was recorded in a nearly noise free environment by microphone in a studio, this database consists of twenty three Arabic words. These words are spoken by one speaker, and this speaker utters each word by different fifteen versions. It is clear that the length of the words is different, these versions for the same word vary in length. The total number of words in the database is 345 utterances for one speaker and this database was used for training and evaluation of the proposed algorithm.

The twenty three Arabic words were shown in **Table 1**:

Table 1: The words and their corresponding number of versions

Word number	Word name	Total number of versions used	Base dataset	Test dataset
1	اشارة	15	10	5
2	لندن	15	10	5
3	افتح	15	10	5
4	الخير	15	10	5
5	تصميم	15	10	5
6	ثقيل	15	10	5
7	خاص	15	10	5
8	دوران	15	10	5
9	رازق	15	10	5
10	رحمن	15	10	5
11	زيارة	15	10	5
12	صباح	15	10	5
13	صديق	15	10	5
14	عمودي	15	10	5
15	كامل	15	10	5
16	محمد	15	10	5
17	معلومات	15	10	5
18	نظام	15	10	5
19	وفاء	15	10	5
20	ياسين	15	10	5
21	يمين	15	10	5
22	مساء	15	10	5
23	زهرة	15	10	5

5.2 Sampling

The speech signals are sampled to convert it from analogue to digital signal. The sampling rate has been down sampled from 44 KHz to 8 KHz.

5.3 Framing

In this stage the continuous speech signal is blocked in frames of N samples. Since we deal with speech signal, which is non stationary signal (vary with time), the framing process is essential to deal with frames not with whole signal. After this stage the speech signal has many frames and the number of frames depends on the number of samples for each word. The number of samples for each frame is 256 samples.

5.4 Hamming Windowing

Each frame of the word was multiplied by the hamming window; the advantage of this multiplication is to minimize the signal discontinuities at the beginning and the end of each frame.

5.5 Feature Extraction

The feature extraction stage was used two methods, namely, DWT and SLT coefficients extractions.

5.5.1 DWT Coefficients Extraction

The DWT was applied on each frame. In this work 1 to 8 DWT level were applied. The 3-level wavelet decomposition structure is shown in **Fig. 3**, where the result is four different subsets, three subsets for the details (the wavelet function coefficients) and the subset four is the approximation subset (the scaling function coefficients).

Most of the energy of the speech signal lies in the lower frequency bands, the other sub-bands contain most detail information of the signal and they are discarded, since the frequency band covered by these levels contains much noise and is less necessary for representing the approximate shape of the speech signal. Hence take the approximation coefficients and discard the details coefficients. In this work the wavelet Daubechies 1 type and 4 type were used.

5.5.2 SLT Coefficients Extraction

The SLT was found using the 8-scale filter bank, this structure will be obtained using 16 different filters and the result of using this transformation on the extracted frames (256 samples for each frame) will be 16 different subbands as shown in **Fig. 6**, L is the number of scales.

Where:

a1: is 1 sample which is the output of the low pass filter h8 after down sampling by $2^L = 256$, L=8.

d1: is 1 sample which is the output of the band pass filter f8 after down sampling by $2^L = 256$, L=8.

d2: is 1 sample which is the output of the band pass filter g7 after down sampling by $2^{i+1} = 2^{7+1} = 256$.

d3: is 1 sample which is the output of the band pass filter (shifted time reverse of g7) after down sampling by $2^{7+1} = 256$.

d4: are 2 samples which are the outputs of the band pass filter g6 after down sampling $2^{6+1} = 128$.

d5: are 2 samples which are the outputs of the band pass filter (shifted time reverse g6) after down sampling $2^{6+1} = 128$.

d6: are 4 samples which are the outputs of the band pass filter g5 after down sampling by $2^{5+1} = 64$.

d7: are 4 samples which are the outputs of the band pass filter (shifted time reverse of g5) after down sampling by $2^{5+1} = 64$.

d8: are 8 samples which are the outputs of the band pass filter g4 after down sampling by $2^{4+1} = 32$.

d9: are 8 samples which are the outputs of the band pass filter (shifted time reverse of g4) after down sampling by 2^{4+1} =32.
d10: are 16 samples which are the outputs of the band pass filter g3 after down sampling by 2^{3+1} =16
d11: are 16 samples which are the outputs of the band pass filter (shifted time reverse of g3) after down sampling 2^{3+1} = 16
d12: are 32 samples which are the outputs of the band pass filter g2 after down sampling by 2^{2+1} = 8.
d13: are 32 samples which are the outputs of the band pass filter (shifted time reverse of g2) after down sampling 2^{2+1} = 8.
d14: are 64 samples which are the outputs of the band pass filter g1 after down sampling by 2^{1+1} = 4.
d15: are 64 samples which are the outputs of the band pass filter (shifted time reverse of g1) after down sampling by 2^{1+1} = 4.
All these filters are taken for extracted features for each frame.

5.6 Dynamic Time Warping Algorithm

After feature extraction phase, each word version represented by one feature vector, these feature vectors are different in length. Dynamic time warping DTW classifier is suitable for data with different length.
The DTW algorithm for classification of words is explained as follows:

a) The database was divided as a base and a test datasets. The base dataset consists of 10 versions of each word while the test dataset consists of 5 versions of each word, the total samples in basing dataset were 230 (10*23) and in testing dataset were 115 (5*23).

b) The distance values between each version from test with each version from base were calculated for all words, i.e., calculate the distance between v11 and each version from 230 versions in base and this calculation is repeated for each version from 115 versions in test. The distance calculation is as follows:
(i) Find the warping path between two versions by using the DTW algorithm.
(ii) Calculate the global warp cost of the two versions which is defined in eq. (9).

For example consider two words w1, and w2.
Assume that the first word has three versions
v1= [1 2 -3 4], v2= [1 -2 3 -4], v3= [2 -1 1]
The second word has three versions too
v4= [-1 3 -4 5 2], v5= [-1 3 -4 6], v6= [-2 -1 3 4 5 6]
Let us choose (v1, v2, v4, v5) as a base and (v3, v6) a test.

a) Apply DTW on v3 with each version in the base (v1, v2, v4, v5)
(i) Apply the algorithm of DTW for v3 and v1.

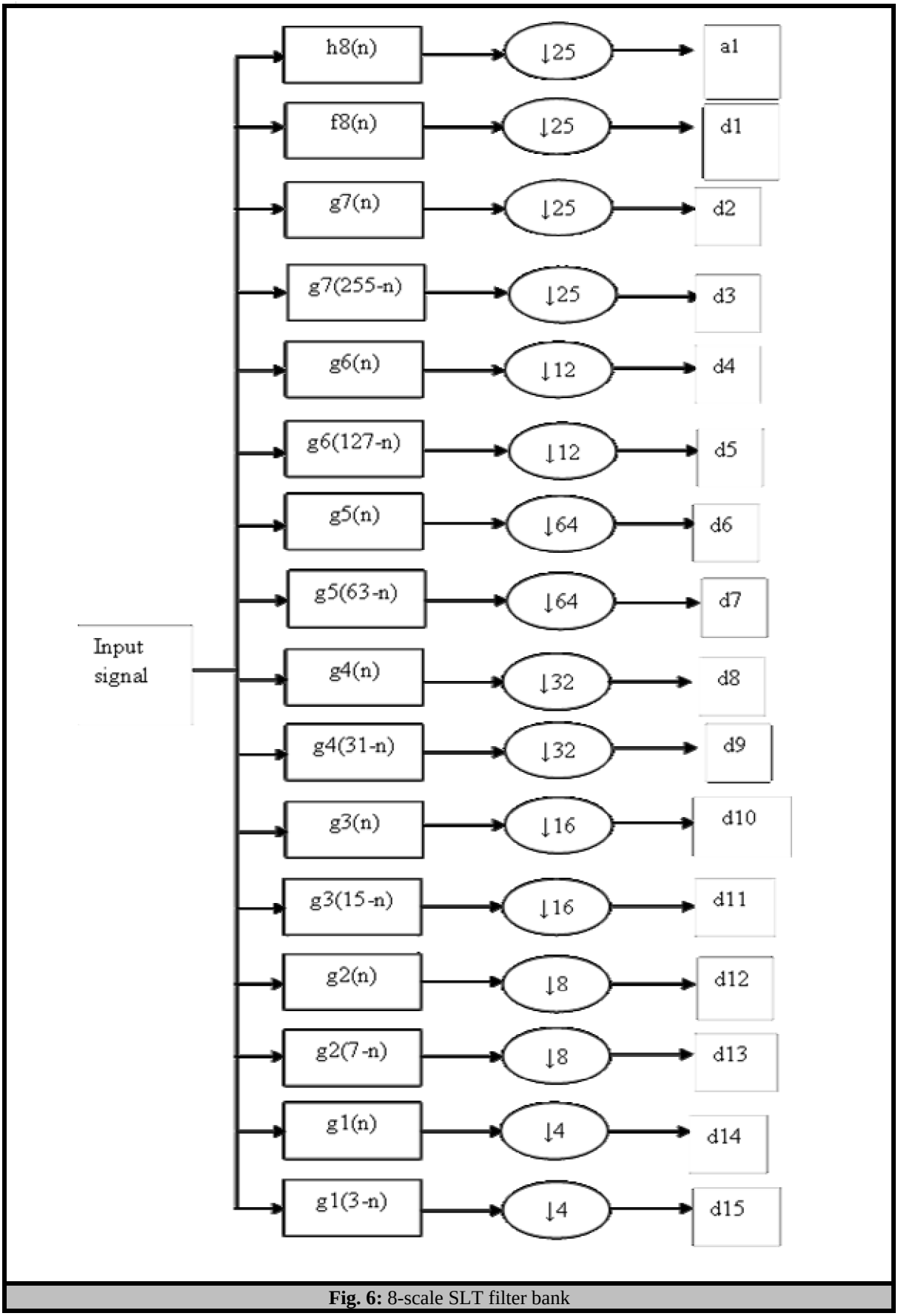
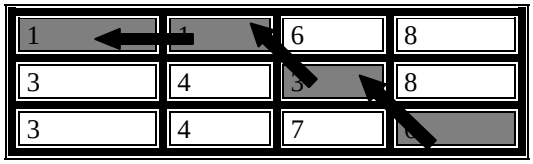


Fig. 6: 8-scale SLT filter bank

(1) Find the warping path between v3 and v1, the distance matrix is:

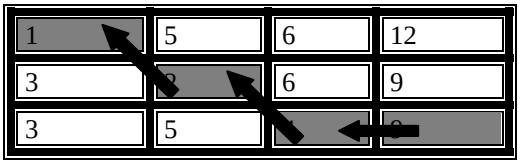


The warping path is (6, 3, 1, 1)

(2) Calculate the global warp cost (eq.9)

$$GC = \frac{1}{4}\sum_{i=1}^{4} w_i = \frac{(6+3+1+1)}{4} = 2.75$$

(ii) Apply the above algorithm on v3 with v2.
(1) The distance matrix is:

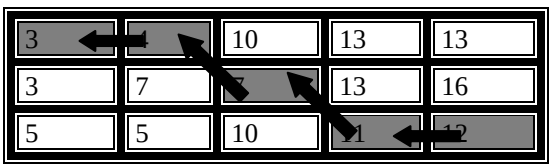


The warping path is (9, 4, 2, 1)

(2) The global warp cost is

$$GC = \frac{1}{4}\sum_{i=1}^{4} w_i = \frac{(9+4+2+1)}{4} = 4$$

(iii) Apply the same algorithm on v3 with v4.
(1) The warping path is shown below:

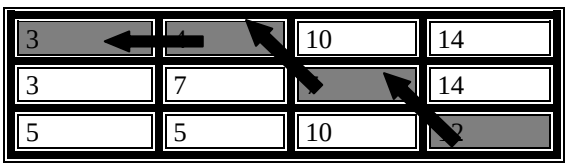


The warping path is (12, 11, 7, 4, 3)

(2) The global warp cost is

$$GC = \frac{1}{5}\sum_{i=1}^{5} w_i = \frac{(12+11+7+4+3)}{5} = 7.4$$

(iv) Take v3 with v5
(1) The warping path is shown below

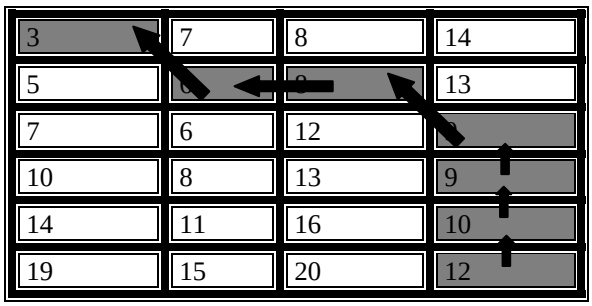


The warping path is (12, 7, 4, 3)

(2) The global warp cost is

$$GC = \frac{1}{4}\sum_{i=1}^{4} w_i = \frac{(12+7+4+3)}{4} = 6.5$$

b) Apply DTW on v6 with each version in the base (v1, v2, v4, v5)
(i) Apply the algorithm of DTW on v6 with v1.
(1) Find the warping path between v6 and v1

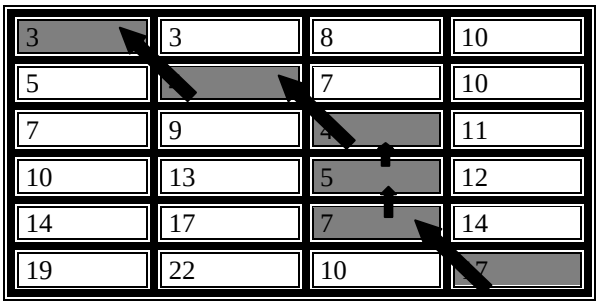


The warping path is (12, 10, 9, 9, 8, 6, 3)

(2) The global warp cost is

$$GC = \frac{1}{7}\sum_{i=1}^{7} w_i = \frac{(12+10+9+9+8+6+3)}{7} = 8.1$$

(ii) Apply the same algorithm on v6 with v2.
(1) Find the warping path between v6 and v2

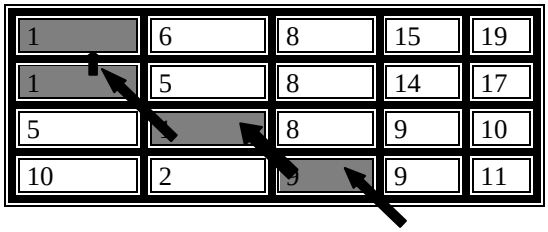


The warping path is (17, 7, 5, 4, 4, 3)

(2) The global warp cost is:

$$GC = \frac{1}{6}\sum_{i=1}^{6} w_i = \frac{(17+7+5+4+4+3)}{6} = 6.6$$

(iii) Apply the above algorithm on v6 with v4.
(1) Find the warping path between v6 and v4



16	4	11	9	12
23	7	14	10	13

The warping path is (13, 9, 9, 1, 1, 1)

(2) The global warp cost is

$$GC = \frac{1}{6}\sum_{i=1}^{6} w_i = \frac{(13+9+9+1+1+1)}{6} = 5.6$$

(iv) Apply the algorithm on v6 with v5.
(1) Find the warping path between v6 and v5

1	6	8	16
1	5	8	15
5	4	8	11
10	2	3	10
16	4	11	10
23	7	14	10

The warping path is (10, 10, 9, 1, 1, 1)

(2) The global warp cost is

$$GC = \frac{1}{6}\sum_{i=1}^{6} w_i = \frac{(10+10+9+1+1+1)}{6} = 5.3$$

The final cost or distance between each version from test and each version from base is shown in **Table 2**

Table 2: Final distance between each version from test and each version from base

		Word1	Word1	Word2	Word 2
		v1	v2	v4	v5
Word1	v3	2.75	4	7.4	6.5
Word2	v6	8.1	6.6	5.6	5.3

The distance between v3 and (v1, v2) is less than the distance between v3 and (v4, v5). DTW recognized the word one from word two because v1, v2 and v3 are the versions for the word one. The distance between v6 and (v4, v5) is less than the distance between v6 and (v1, v2). DTW recognized the word two from word one because v4, v5, and v6 are versions for the word two.
The test word belongs to the same class as the base word with minimum GC value.

Table 2 shows that, the distance between test word v3 and base word v1 is the minimum, and the distance between test word v6 and base word v5 is the minimum distance, hence the DTW algorithm recognized the word one from word two.
This is a simple example with simple database (four versions for the basing and two versions for the testing, and the data are not speech signals but are simple values to illustrate DTW classification algorithm.

6. Results

In this work, the proposed system has been applied on twenty three Arabic words. These words and their corresponding number of version are shown in **Table 1**.
The number of versions of each word has been divided into two parts:
a. One part of these versions used for the base of the DTW algorithm called "base versions", ten versions have been taken for each word.
b. The other part used for the test of the DTW algorithm called "test versions", five versions have been taken for each word, the test versions are tested on the DTW and their resultant error is used to give the measure of the generalization ability of this algorithm.
In this work three different transforms were used to extract the feature from the speech signal namely DWT Db1, DWT Db4 and SLT and the DTW algorithm was used for classification. To compare the performance of these transforms the accuracy or recognition rate of each one has been computed by the following equation:

$$\text{Accuracy} = \frac{\text{Total number of correct recognition}}{\text{Total number of testing versions}} * 100\% \qquad (10)$$

Table 3 shows the comparison between the accuracy of different recognition systems used in this work. **Fig. 7**: shows the accuracy of each recognition system.

Recognition system	Total number of testing versions	Total number of correct recognition	Accuracy
Db1 (1-level)	115	103	89.56%
Db1 (2-level)	115	104	90.43%
Db1 (3-level)	115	107	93.04%
Db1 (4-level)	115	102	88.69%
Db1 (5-level)	115	103	89.56%
Db4 (3-level)	115	106	92.17%
SLT	115	109	94.78%

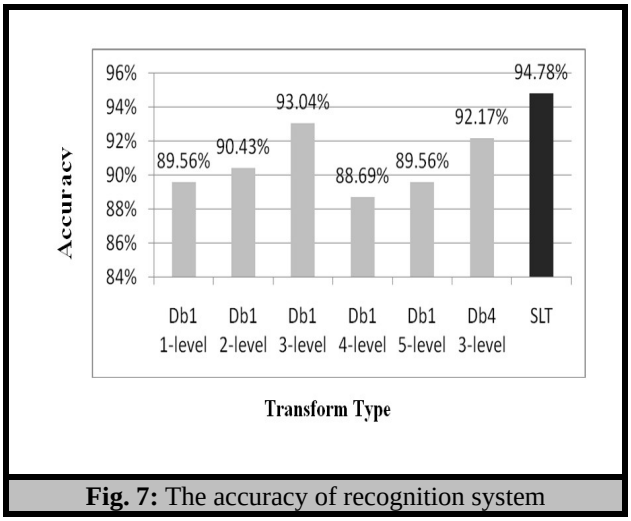


Fig. 7: The accuracy of recognition system

7. Conclusions

In this work three different types of transformation methods were used to extract features from the speech signal. The results show that the Db1 has the best result than Db4 and the 3level Db1 has the best result among the other levels. The number of scale in SLT depends on the number of samples in input signals and is govern by the relation ($L = \log_2 N$), N is number of samples in input signal and L is number of scale in SLT. In DWT, the number of decomposition level is not governing by above relation, but it can be chosen, and choosing the right decomposition level will affect recognition results. Since the digital filter is achieved by shifting and delaying operations and the slantlet transform has short filters with length approach to (2/3) of the length of iterated wavelet filter bank, therefore the SLT is faster than DWT. The SLT is an orthogonal transform and provides improved time localization than DWT, therefore it will improve the recognition results. The comparison of the accuracy of the SLT system with other systems used in this work gives a conclusion that the SLT gives improved accuracy for the speech signal recognition.

8. References

1. Abbas, T.M.J. **"Speech Recognition Using Features Combination"**, Ph.D. thesis, Iraqi Commission for Computers & Informatics -Informatics Institute for Postgraduate Studies, August, 2005.

2. Amin, T.B. & Mahmood, I. **"Speech Recognition Using Dynamic Time Warping"**, IEEE, 2nd International Conference on Advances in Space Technologies, Islamabad, Pakistan, vol. 2, pp.74-79, 29th-30th November, 2008.

3. Arivazhagan, S., Jebarani, W.S. & Kumaran, G. **"Performance Comparison of Discrete Wavelet Transform and Dual Tree Discrete Wavelet Transform for Automatic Airborne Target Detection"**, IEEE, International Conference on Computational Intelligence and Multimedia Applications, pp.495-500, 2007.

4. Burrus,C.S., Gopinath, R.A.& Guo, H. **"Introduction to Wavelets and Wavelet Transform"**, Prentice Hall, 1998.

5. Chang, P.C., Fan, C.Y., Lin, J.L. & Lin, J.J. **"Integrating a Piecewise Linear Representation Method with Dynamic Time warping system for Stock Trading Decision Making"**, IEEE Fourth International Conference on Natural Computation, pp.434-438, 2008.

6. Chatterjee, A., Maitra, M. & Goswami, S.K. **"Classification of overcurrent and inrush current for power system reliability using Slantlet transform and artificial neural network"**, Expert Systems with Applications, Elsevier, Vol. 36, pp.2391-2399, 2009.

7. Furtunà, T.F. **"Dynamic Programming Algorithms in Speech Recognition"**, Revista Informatica Economică. Vol. 46, No.2, Bucharest, pp.94-99, 2008.

8. Kociołek, M., Materka, A., Strzelecki, M. & Szczypiński, P. **"Discrete wavelet transform derived features for digital image texture analysis"**, Proc. of Interational Conference on Signals and Electronic Systems, Lodz, Poland, pp. 163-168, 18-21 September, 2001.

9. Lipeika, A., Lipeikiene, J. & Telksnys, L. **"Development of Isolated Word Speech Recognition System"**, Institute of Mathematics and Informatics, Vilnius, INFORMATICA, Vol. 13, No. 1, PP.37–46, 2002.

10. Pour, M.M. & Farokhi, F. **"An Advanced Method for Speech Recognition"**, World Academy of Science, Engineering and Technology, 49, pp.995-1000, 2009.

11. Rabiner, L.R. & Schafer, R.W. **"Digital Processing of Speech Signals"**, Englewood Cliffs, New Jersey: Prentice Hall, 1978.

12. Sahu, R. & Sanjeev, P.B. **"Co-integration of Stock Markets using Wavelet Theory and Data Mining"**, ABV-Indian Institute of Information Technology and Management, India, 2007.

13. Selesnick, I.W. **"The Slantlet Transform"**, IEEE Transaction on Signal Processing, Vol. 47, No. 5, pp. 1304-1313, MAY, 1999.

14. Shanker, A.P. & Rajagopalan, A.N. **"Off-line Signature Verification Using DTW"**, Pattern Recognition Letters, Vol. 28, Elsevier, pp.1407-1414, 2007

تمييز الكلمات بأستخدام تحويل المويل وطريقة ميلان الزمن الديناميكي

د. صادق جاسم أبو اللوخ سماح مطشر كاطع
جامعة بغداد، كلية الهندسة ، قسم الهندسة الكهربائية

الخلاصة:
استعمل نظام تمييز الكلام بصورة واسعة بواسطة عدد من الباحثين بطرائق مختلفة لتحقيق طريقة تمييزسريعة ودقيقة. أن تمييز اشارة الكلام تعتبر مشكلة تصنيف نوعية وهي تضم بصورة عامة جزئين اساسيين: استخلاص الميزات و التصنيف. تضمن هذا العمل مقارنة بين ثلاثة طرق لاستخلاص الخصائص وهي تحويل المويجة (Db1and Db4) وتحويل المويل (SLT). استخدمت طريقة ميلان الزمن الديناميكي (DTW) للتمييز. ثلاثة وعشرون كلمة عربية بخمسة عشرازمان مختلفة مسجلة في الاستوديو بواسطة متكلم واحد لتشكل قاعدة بيانات . النظام المقترح وجد باستخدام قاعدة البيانات هذه. النتيجة بينت أن دقة التمييز هي (٩٣.٠٤%، ٩٢.١٧% و ٩٤.٧٨%) باستخدام (Db1, Db4 and SLT) على التوالي.