We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,900
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Discrete Wavelet Transform & Linear Prediction Coding Based Method for Speech Recognition via Neural Network

K. Daqrouq¹, A.R. Al-Qawasmi¹, K.Y. Al Azzawi² and T. Abu Hilal³

¹Philadelphia University/Communications and Electronics Department
²University of Technology
³Dhofar University
¹Jordan
²Iraq
³Oman

1. Introduction

In the proposed work, the techniques of wavelet transform (WT) and neural network were introduced for speech based text-independent speaker identification and Arabic vowel recognition. The linear prediction coding coefficients (LPCC) of discrete wavelet transform (DWT) upon level 3 features extraction method was developed. Feature vector fed to probabilistic neural networks (PNN) for classification. The functions of features extraction and classification are performed using the wavelet transform and neural networks (DWTPNN) expert system. The declared results show that the proposed method can make a powerful analysis with average identification rates reached 93%. Two published methods were investigated for comparison. The best recognition rate selection obtained was for framed DWT. Discrete wavelet transform was studied to improve the system robustness against the noise of 0dB. Our investigation of speaker-independent Arabic vowels classifier system performance is performed via several experiments depending on vowel type. The declared results show that the proposed method can make an effectual analysis with identification rates may reach 93%.

In general, a speaker identification system can be implemented by observing the voiced/unvoiced components or through analyzing the energy distribution of utterances. A number of digital signal processing algorithms, such as LPC technique (Adami & Barone, 2001; Tajima, Port, & Dalby, 1997), Mel frequency cepstral coefficients (MFCCs) (Mashao & Skosan, 2006; Sroka & Braida, 2005; Kanedera, Arai, Hermansky & Pavel, 1999; Daqrouq & Al-Faouri, 2010), DWT (Fonseca, Guido, Scalaßara, Maciel, & Pereira, 2007) and wavelet packet transform (WPT) (Lung, 2006; Zhang & Jiao, 2004) are extensively utilized. In the beginning of 1990s, Mel frequency cepstral technique became the most widely used technique for recognition purposes due to its aptitude to represent the speech spectrum in a compacted form (Sarikaya & ansen, 2000). Actually, MFCCs simulate the model of humans’ auditory perception and have been proven to be very effective in automatic speech recognition system and modeling the individual frequency components of speech signals.
ESI has been under research by a large number of researchers for about four decades (Reynolds, Quatieri, & Dunn, 2000). From a commercial point of view, ESI is a technology with potentially large market due to the applications of frequently ranges from automation of operator- helped service to speech-to-text aiding system for hearing impaired individuals (Reynolds et al., 2000).

Artificial neural network performance is depending mainly on the size and quality of training samples (Visser, Otsuka, & Lee, 2003). When the number of training data is small, not representative of the possibility space, standard neural network results are poor (Kosko & Bart, 1992). Incorporation of neural fuzzy or wavelet techniques can improve performance in this case, particularly, by input matrix dimensionality decreasing (Nava & Taylor, 1996). Artificial neural networks (ANN) are known to be excellent classifiers, but their performance can be prevented by the size and quality of the training set. Fuzzy theory has been used successfully in many applications (Cowdy & Tufekci, 2000). This applications show that fuzzy theory can be used to improve neural network performance.

In this study, authors improve effective feature extraction method for text-independent system, taking in consideration that the size of ANN input is very crucial issue. This affects quality of the training set. For this reason, the presented features extraction method offers a reduction of dimensionality of features comparing with conventional methods. LPCC of DWT in conjunction is utilized. For classification of features extraction coefficients, PNN is proposed.

In this paper, an expert system for speaker identification was proposed for the investigation of the speech signals using pattern identification. The speaker identification performance of this method demonstrated on the total 59 individual speakers (39 male speakers and 20 female speakers). LPCC in conjunction with DWT upon level seven features extraction method were developed. For performing the classification process PNN was investigated. The function of feature extraction and classification is performed using the DWPN expert system. The declared results show that the proposed method can make an effectual analysis. The average identification rates were 94.89, better than other methods published before. It was found that the recognition rates enhanced upon increasing the number of feature sets (by higher DWT levels). Nevertheless, the improvement implies a tradeoff between the recognition rate and extracting time. The proposed method can offer a significant computational advantage by reducing the dimensionality of the WT coefficients by means of LPCC. DWT approximation Sub-signal via several levels instead of original imposter had good performance on real noise facing, particularly upon level 3 and 4.

2. Discrete Wavelet Transform

The DWT indicates an arbitrary square integrable function as a superposition of a family of basis functions called wavelet functions. A family of wavelet basis functions can be produced by translating and dilating the mother wavelet related to the family (Mallat, 1989). The DWT coefficients can be generated by taking the inner product between the input signal and the wavelet functions. Since the basis functions (wavelet functions) are translated and dilated versions of each other, a simpler algorithm, known as Mallat's pyramid tree algorithm, has been proposed in (Mallat, 1989).

The DWT can be treated as the multiresolution decomposition of a sequence. It takes a length \( N \) sequence \( a(n) \) as input and produces a length \( N \) sequence as the output. The
output has values at the highest resolution (level 1) and $N/4$ values at the next resolution (level 2), and so on. Let $N = 2^m$, and let the number of frequencies, or resolutions, be $m$, we are bearing in mind $m = \log N$ octaves [18]. So that, the frequency index $k$ varies as $1, 2, \ldots, m$ corresponding to the scales $2^1, 2^2, \ldots, 2^m$. In As described by Mallat pyramid algorithm (Fig. 1), the DWT coefficients of the previous stage are expressed as follows (Souani et al., 2000):

$$W_L(n,k) = \sum_i W_L(i,k-1)h(i-2n),$$  \hspace{0.5cm} (1a)

$$W_H(n,k) = \sum_i W_L(i,k-1)g(i-2n),$$  \hspace{0.5cm} (1b)

Where $W_L(p,q)$ is $p$th scaling coefficient at the $q$th stage, $W_H(p,q)$ is the $p$th wavelet coefficient at the $q$th stage, and $h(n), g(n)$ are the dilation coefficients relating to the scaling and wavelet functions, respectively.

For computing the DWT coefficients of the discrete-time data (signal), it is assumed that the input data represents the DWT coefficients of a high resolution stage. Equations (1a) and (1b) may be used for obtaining DWT coefficients of subsequent stages. In practice, this decomposition is used only for a few stages. We note that the dilation coefficients $h(n)$ stand for a low-pass filter, where the corresponding $g(n)$ stands for a high-pass filter. In order that, DWT takes out information from the signal at different scales. The first level of wavelet decomposition extracts the details of the signal (high frequency parts), while the second and all subsequent wavelet decompositions take out progressively coarser information (lower frequency parts). Each step of retransforming the low-pass output is called dilation. A schematic of three stages DWT decomposition is shown in Fig. 1. $H$ presents the High pass filter and $L$ denotes the low pass filter. At the output of each filter the result is down sampled (decimated) by taking one coefficient and leave other (Souani et al., 2000).

So as to reconstruct the original data, the DWT coefficients are up sampled (insertion of a zero between two samples) and passed through another set of low- and high-pass filters, which are expressed as

$$W_L(n,k) = \sum_p W_L(p,k+1)h'(n-2p) + \sum_l W_H(l,k+1)g'(n-2l),$$  \hspace{0.5cm} (2)

where $h'(n)$ and $g'(n)$ are the low- and the high-pass synthesis filter, respectively. It is observed from Eq. (2) that the $k$th level DWT coefficients may be obtained from $(k+1)$th level DWT coefficients. Efficiently supported wavelets are generally used in various applications.

In the last decade, there has been a huge increase in the applications of wavelets in various scientific disciplines. Typical applications of wavelets include signal processing, image processing, security systems, numerical analysis, statistics, biomedicine, etc. Wavelet transform tenders a wide variety of useful features, on the contrary to other transforms, such as Fourier transform or cosine transform. Some of these are as follows:

- Adaptive time-frequency windows,
- Lower aliasing distortion for signal processing applications,
- Computational complexity of $O(N)$, where $N$ is the length of data;
- Inherent scalability;
- Efficient Very Low Scale Integration implementation

Fig. 1. a. DWT-tree by Mallat's Algorithm; b. IDWT by Mallat's Algorithm

3. The use of DWT for feature extraction

Before the stage of features extraction, the speech data are processed by a silence removing algorithm followed by the application of a pre-processed by applying the normalization on speech signals to make the signals comparable regardless of differences in magnitude. In this study three feature extraction methods based on discrete wavelet transform are discussed in the following part of the paper.

3.1 DWT method with LPC

For an orthogonal wavelet function, a library of DWT bases is generated. Each of these bases offers a particular way of coding signals, preserving global energy and reconstructing exact features. The DWT is used to extract additional features to guarantee higher recognition rate. In this study, DWT is applied at the stage of feature extraction, but these data are not proper for classifier due to a great amount of data length. Thus, we have to seek for a better representation for the speaker features. Previous studies proposed that the use of LPC of DWT as features in recognition tasks is competent. (Adami & Barone, 2001; Tajima, Port, & Dalby, 1997) Suggested a method to calculate the LPC orders of wavelet transform for speaker recognition.
In this method the LPC is obtained from DWT Sub signals. The DWT at level three is generated and then 30 LPC orders are obtained for each sub signals to be combined in one feature vector. The main advantage of such sophisticated feature method is to extract different LPC impact based on multi resolution of DWT capability. LPC orders sequence will contain distinguishable information as well as wavelet transform. Fig.3 shows LPC orders calculated for DWT at depth 3 for three different utterances for the same person. We may notice that the feature vector extracted by DWT and LPC is appropriate for speaker recognition.

3.2 DWT method with entropy

As seen in above studies, the entropy of the specific sub-band signal may be employed as features for recognition tasks. This is possible because each Arabic vowel has distinct energy (see Fig.3). In this paper, the entropy obtained from the DWT will be employed for speaker recognition. The features extraction method can be explained as follows:
- Decomposing the speech signal by wavelet packet transform at level 7, with Daubechies type (db2).
- Calculating three entropy types for all 256 nodes at depth 7 for wavelet packet using the following equations:
  
  Shannon entropy:
  \[
  E_1(s) = -\sum s_i^2 \log(s_i^2)
  \]  

  Log energy entropy:

Fig. 3. LPC orders calculated for DWT at depth 3 for three different utterances for the same person
\[ E_1(s) = \sum_i \log(s_i^2) \]  \hspace{1cm} (4)

Sure entropy:

\[ |s_i| \leq p \Rightarrow E(s) = \sum_i \min(s_i^2, p^2) \]  \hspace{1cm} (5)

Where \( s \) is the signal, \( s_i \) are the DWT coefficients and \( p \) is a positive threshold. Entropy is a common concept in many fields, mainly in signal processing. Classical entropy-based criterion describes information-related properties for a precise representation of a given signal. Entropy is commonly used in image processing; it possesses information about the concentration of the image. On the other hand, a method for measuring the entropy appears as a supreme tool for quantifying the ordering of non-stationary signals. Fig. 3 shows the three entropies calculated for DWT at depth 3 for three different utterances for the same person. We may notice that the feature vector extracted by DWT and entropy is appropriate for speaker recognition. This conclusion has been obtained by interpretation the following criterion: the feature vector extracted should possess the following properties: 1) Vary widely from class to class. 2) Stable over a long period of time. 3) Should not have correlation with other features (see Fig. 4).

![Entropy graph](Fig. 4. Entropy calculated for DWT at depth 3 for three different utterances for the same person)

4. Proposed probabilistic neural networks algorithm

We create a probabilistic neural network algorithm for classification problem (see Fig. 5 and Fig. 6):

\[ \text{Net} = \text{PNN}(P, T, \text{SPREAD}), \]

where \( P \) is a \( 4 \times 2^{q+1} \times 24 \) matrix of 24 input vowel feature vectors for net training, of \( 2^{q+1} \) (minus 2, repeated original node) WP nodes number;

www.intechopen.com
Discrete Wavelet Transform & Linear Prediction Coding Based Method for Speech Recognition via Neural Network

\[
P = \begin{bmatrix}
WR_{r_1} & WR_{r_2} & \cdots & WR_{r_{24}} \\
WR_{s_1} & WR_{s_2} & \cdots & WR_{s_{24}} \\
\vdots & \vdots & \ddots & \vdots \\
WR_{4r_{24}r_1} & WR_{4r_{24}s_2} & \cdots & WR_{4r_{24}r_{24}} 
\end{bmatrix},
\]

(6)

\[T = [1, 2, 3, \ldots, 24], (5)\]

is the target class vector 

and SPREAD is spread of radial basis functions. We employ a SPREAD value of 1 because that is a typical distance between the input vectors. If SPREAD is near zero the network acts as a nearest neighbor classifier. As SPREAD becomes larger the designed, network will take into account several nearby design vectors.

Fig. 5. Structure of the original probabilistic neural network

www.intechopen.com
5. Results and discussion

5.1 Speaker identification by DWTLPC

A testing database was produced from Arabic language. The recording environment is a normal office environment through PC-sound card, with frequency 4 KHz and sampling frequency 16 KHz. These utterances are Arabic spoken words. Total 47 individual speakers (19 to 40 years old) who are 31 individual male and 16 individual female spoken these Arabic words for training and testing phases. The total number of tokens considered for training and testing was 653.

It were performed experiments using total 653 the Arabic utterances of total 47 individual speakers (31 male speakers and 16 female speakers). For each of these speakers, up to 15 speech signals were used. 6 of these signals were used for training and from 4 to 9 of these signals (depend of recordings signals for each speaker) were used for testing the expert system (Fig.6). In this experiment, 93.26% correct classification was obtained by means of DWTLPC among the 47 different speaker signal classes. Testing results are tabulated in Tab.1. It, clearly, indicates the usefulness and the trustworthiness of the proposed approach for extracting features from speech signals gender identification system.
<table>
<thead>
<tr>
<th>Speaker</th>
<th>Number of Signals</th>
<th>Recognized Signals</th>
<th>Recognition Rate [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sp.1</td>
<td>9</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>Sp.2</td>
<td>9</td>
<td>8</td>
<td>88.88</td>
</tr>
<tr>
<td>Sp.3</td>
<td>9</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>Sp.4</td>
<td>9</td>
<td>8</td>
<td>88.88</td>
</tr>
<tr>
<td>Sp.5</td>
<td>9</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>Sp.6</td>
<td>9</td>
<td>6</td>
<td>66.66</td>
</tr>
<tr>
<td>Sp.7</td>
<td>9</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>Sp.8</td>
<td>9</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>Sp.9</td>
<td>9</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>Sp.10</td>
<td>9</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>Sp.11</td>
<td>9</td>
<td>8</td>
<td>88.88</td>
</tr>
<tr>
<td>Sp.12</td>
<td>9</td>
<td>6</td>
<td>66.66</td>
</tr>
<tr>
<td>Sp.13</td>
<td>9</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>Sp.14</td>
<td>9</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>Sp.15</td>
<td>9</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>Sp.16</td>
<td>9</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>Sp.17</td>
<td>8</td>
<td>7</td>
<td>87.5</td>
</tr>
<tr>
<td>Sp.18</td>
<td>8</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Sp.19</td>
<td>8</td>
<td>7</td>
<td>87.5</td>
</tr>
<tr>
<td>Sp.20</td>
<td>4</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>Sp.21</td>
<td>4</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>Sp.22</td>
<td>4</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>Sp.23</td>
<td>4</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>Sp.24</td>
<td>4</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>Sp.25</td>
<td>8</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Sp.26</td>
<td>8</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Sp.27</td>
<td>8</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Sp.28</td>
<td>8</td>
<td>5</td>
<td>62.5</td>
</tr>
<tr>
<td>Sp.29</td>
<td>8</td>
<td>7</td>
<td>87.5</td>
</tr>
<tr>
<td>Sp.30</td>
<td>8</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Sp.31</td>
<td>8</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Sp.32</td>
<td>8</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Sp.33</td>
<td>8</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Sp.34</td>
<td>8</td>
<td>7</td>
<td>87.5</td>
</tr>
<tr>
<td>Sp.35</td>
<td>8</td>
<td>7</td>
<td>87.5</td>
</tr>
<tr>
<td>Sp.36</td>
<td>8</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Sp.37</td>
<td>8</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Sp.38</td>
<td>8</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Sp.39</td>
<td>8</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Sp.40</td>
<td>8</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>Sp.41</td>
<td>8</td>
<td>7</td>
<td>87.5</td>
</tr>
<tr>
<td>Sp.42</td>
<td>8</td>
<td>7</td>
<td>87.5</td>
</tr>
</tbody>
</table>
Table 1. DWTLPC Identification Rate results

Table 2 shows the experimental results of different approaches used in the experimental investigation for comparison. Modified DWT with proposed feature extraction method (MDWT LPC), framing DWTLPC (FDWT LPC) illustrated in Fig.8, where LPC orders are obtained from six frames of each DWT sub signal and proposed method DWTLPC were investigated for comparison. The recognition rate of MDWT LPC reached the lowest value. The best recognition rate selection obtained was 93.53% for FDWT LPC.

Table 2. Comparison of different classification approaches

Fig. 7. Proposed system performance by using DWT approximation sub-signals (at level 1 to 4).

To improve the robustness of DWTLPC to additive white Gaussian noise (AWGN), same wavelet decomposition process was applied to DWT approximation sub-signal via several levels instead of original imposter (Daqrouq, 2011). Afterwards, the features extraction was applied to each of the obtained wavelet decomposition sub-signals (see Fig.7). After performing proposed classification mechanism for each sub-signal of distinct DWT level, we can notice that at level 3 and 4 the highest recognition rate was achieved (see Tab.4). In this experiment it was found that the recognition rates were not improved upon increasing the DWT level more than four.
Fig. 8. Feature extraction vectors for three signals of same speaker obtained by a. FDWTLPC. b. DWTLPC

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Number of Signals</th>
<th>Recognized Signals [100%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sp.1</td>
<td>24</td>
<td>0 0 25 40</td>
</tr>
<tr>
<td>Sp.2</td>
<td>24</td>
<td>0 22 87 10</td>
</tr>
<tr>
<td>Sp.3</td>
<td>24</td>
<td>0 25 25 25</td>
</tr>
<tr>
<td>Sp.4</td>
<td>24</td>
<td>0 0 50 100</td>
</tr>
<tr>
<td>Sp.5</td>
<td>24</td>
<td>10 55 60 40</td>
</tr>
<tr>
<td>Sp.6</td>
<td>24</td>
<td>25 12 25 100</td>
</tr>
</tbody>
</table>

Table 4. DWTLPC Identification Rate results through DWT with SNR= 0dB
5.2 Arabic vowel classification by using DWTLPC

In recent times, Arabic language became one of the most significant and broadly spoken languages in the world, with an expected number of 350 millions speakers distributed all over the world and mostly covering 22 Arabic countries. Arabic is Semitic language that characterizes by the existence of particular consonants like pharyngeal, glottal and emphatic consonants. Furthermore, it presents some phonetics and morpho-syntactic particularities. The morpho-syntactic structure built, around pattern roots (CVCVCV, CVCCVC, etc.) (Zitouni and Sarikaya, 2009). The Arabic alphabet consists of 28 letters that can be expanded to a set of 90 by additional shapes, marks, and vowels. The 28 letters represent the consonants and long vowels such as ʿ (pronounced as/a:/), َ (pronounced as/i:/), and ُ (pronounced as/u:/). The short vowels and certain other phonetic information such as consonant doubling (shadda) are not represented by letters directly, but by diacritics. A diacritic is a short stroke located above or below the consonant. Table 1 shows the complete set of Arabic diacritics. We split the Arabic diacritics into three sets: short vowels, doubled case endings, and syllabification marks. Short vowels are written as symbols either above or below the letter in text with diacritics, and dropped all together in text without diacritics. We get three short vowels: fatha: it represents the /a/ sound and is an oblique dash over a letter, damma: it represents the /u/ sound and has shape of a comma over a letter and kasra: it represents the /i/ sound and is an oblique dash under a letter as reported in Table 1.

In this work, speech signals were obtained via PC-sound card, with a sampling frequency of 16000 Hz. The Arabic vowels were recorded by 27 speakers: 5 females, along with 22 males. The recording process was provided in normal university office circumstances. Our study of speaker-independent Arabic vowels classifier system performance is performed via several experiments depending on vowel type. In the following three experiments the used feature extraction method is DWTLPC.

**Experimental-1**

We experimented 200 long Arabic vowels ُ (pronounced as/a:/) signals, 400 long Arabic vowels َ (pronounced as/i:/) signals and 90 long Arabic vowels ُ (pronounced as/u:/) signals. The results indicated that 96% were classified correctly for Arabic vowels ُ, 90% of the signals were classified correctly for Arabic vowel َ, and 94% of the signals were classified correctly for Arabic vowel ُ. Tab.5 shows the results of recognition rates.

<table>
<thead>
<tr>
<th>Long Vowels</th>
<th>Number of Signals</th>
<th>Accepted Signals</th>
<th>Not Recognized Signals</th>
<th>Recognition Rate [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long A ُ</td>
<td>200</td>
<td>192</td>
<td>8</td>
<td>96</td>
</tr>
<tr>
<td>Long E َ</td>
<td>400</td>
<td>360</td>
<td>40</td>
<td>90</td>
</tr>
<tr>
<td>Long O ُ</td>
<td>90</td>
<td>85</td>
<td>5</td>
<td>94</td>
</tr>
<tr>
<td><strong>Avr. Recognition Rate</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>93.33</strong></td>
</tr>
</tbody>
</table>

Table 5. The recognition rate results for long vowels
Experimental-2

In this experiment we study the recognition rates for long vowels connected with other consonant such ل (pronounced as/l/) and ر (pronounced as/r/). Tab.6, reported the recognition rates. The results indicated 88.5% average recognition rate.

<table>
<thead>
<tr>
<th>Long Vowels</th>
<th>Number of Signals</th>
<th>Recognized Signals</th>
<th>Not Recognized Signals</th>
<th>Recognition Rate [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>La ل</td>
<td>60</td>
<td>57</td>
<td>3</td>
<td>95</td>
</tr>
<tr>
<td>Le ل</td>
<td>60</td>
<td>60</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Lo ۲ل</td>
<td>60</td>
<td>42</td>
<td>18</td>
<td>70</td>
</tr>
<tr>
<td>Ra ر</td>
<td>60</td>
<td>54</td>
<td>6</td>
<td>90</td>
</tr>
<tr>
<td>Re ر</td>
<td>60</td>
<td>57</td>
<td>3</td>
<td>95</td>
</tr>
<tr>
<td>Ro ر</td>
<td>60</td>
<td>49</td>
<td>11</td>
<td>81</td>
</tr>
</tbody>
</table>

Table 6. The recognition rate results for long vowels connected with other letters

Probabilistic neural network based speech recognition system is presented in this work. This system was performed using a wavelet feature extraction method. In this work, effective feature extraction method for Arabic vowels system is developed, taking in consideration that the computational complexity is very crucial issue. The experimental results on a subset of recorded database showed that feature extraction method proposed in this work is suitable for Arabic recognition system. Our study of speaker-independent Arabic vowels classifier system performance is performed via two experiments depending on vowel type. The declared results show that the proposed method can make an effective analysis with identification rates may reach 93%.

The proposed future work of this study is to improve the capability of proposed system to work in real time. This may be performed by modifying the recording apparatus and a data acquisition system (such as NI-6024E), and interfacing online with written Matlab code that simulates the expert system.

6. Conclusion

In this work, an expert system for speaker identification was investigated for the analyzing of the speech signals using pattern identification. The speaker identification performance of this method demonstrated on the total 47 individual speakers (31 male speakers and 16 female speakers). LPC in conjunction with framed DWT upon level three features extraction method was developed. For performing the classification process PNN was proposed. The stated results show that the proposed method can make an powerful analysis. The performance of the intelligent system was given in Table 1 and Table 2. The average identification rates were 93.26%, better than other methods. Our investigation of speaker-
independent Arabic vowels classifier system performance is performed via several experiments depending on vowel type. The declared results show that the proposed method can make an effectual analysis with identification rates may reach 93%.

7. References

Discrete Wavelet Transform & Linear Prediction Coding
Based Method for Speech Recognition via Neural Network

Fonseca, E. S., Guido, R. C., Scalassara, P. R., Maciel, C. D., & Pereira, J. C. (2007). Wavelet
timefrequency analysis and least squares support vector machines for the

Ganchev T., Fakotakis N., Kokkinakis G. (2005), Comparative evaluation of various MFCC
implementations on the speaker verification task, in: *Proceedings of the SPECOM-2005*,

Gelfer M. P. and Mikos V.A. (2007), The Relative Contributions of Speaking Fundamental
Frequency and Formant Frequencies to Gender Identification Based on Isolated


Haydar, A., Demirekler, M. & Yurtseven, M.K. (1998), Speaker identification through use of

Hyeon, J. &. Bang, S.Y., Feature selection for multi-class classification using pairwise class

pitch detection of speech signals. *IEEE Transactions on Information Theory*,
32(March), 712–718.


various components of the modulation spectrum for automatic speech recognition.


Lung, S.-Y. (2004), Applied multi-wavelet feature to text independent speaker

selection based multilayered neural network using supervised learning algorithm,
*Pattern Recognition* 40, 3616 – 3620.


the fifth IEEE international conference on fuzzy systems* (pp. 2049–2052).


The discrete wavelet transform (DWT) algorithms have a firm position in processing of signals in several areas of research and industry. As DWT provides both octave-scale frequency and spatial timing of the analyzed signal, it is constantly used to solve and treat more and more advanced problems. The present book: Discrete Wavelet Transforms - Biomedical Applications reviews the recent progress in discrete wavelet transform algorithms and applications. The book reviews the recent progress in DWT algorithms for biomedical applications. The book covers a wide range of architectures (e.g. lifting, shift invariance, multi-scale analysis) for constructing DWTs. The book chapters are organized into four major parts. Part I describes the progress in implementations of the DWT algorithms in biomedical signal analysis. Applications include compression and filtering of biomedical signals, DWT based selection of salient EEG frequency band, shift invariant DWTs for multiscale analysis and DWT assisted heart sound analysis. Part II addresses speech analysis, modeling and understanding of speech and speaker recognition. Part III focuses biosensor applications such as calibration of enzymatic sensors, multiscale analysis of wireless capsule endoscopy recordings, DWT assisted electronic nose analysis and optical fibre sensor analyses. Finally, Part IV describes DWT algorithms for tools in identification and diagnostics: identification based on hand geometry, identification of species groupings, object detection and tracking, DWT signatures and diagnostics for assessment of ICU agitation-sedation controllers and DWT based diagnostics of power transformers. The chapters of the present book consist of both tutorial and highly advanced material. Therefore, the book is intended to be a reference text for graduate students and researchers to obtain state-of-the-art knowledge on specific applications.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:


InTech Europe
University Campus STeP Ri
InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai

www.intechopen.com
© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.