



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Wavelet Packet and Percent of Energy Distribution with Neural Networks Based Gender Identification System

¹Emad F. Khalaf, ²Khaled Daqrouq and ²Mohamed Sherif

¹Department of Computer Engineering,

²Department of Communication and Electronics Engineering, Philadelphia University, Jordan

Abstract: This research presents the study of gender identification for security systems based on the energy of speaker utterances. The proposed system consisted of a combination of signal pre-process, feature extraction using Wavelet Packet Transform (WPT) and gender identification using artificial neural network. In the signal pre-process, the amplitude of utterances was normalized for preventing an error estimation caused by speakers' change in volume. 128 features fed to Feed Forward Backpropagation Neural Networks (FFPBNN) for classification. The functions of features extraction and classification are performed using the Wavelet Packet and Percent of Energy Distribution and Neural Networks (WPENN) expert system. The declared results showed that the proposed method can make an effectual analysis with average identification rates reached 91.09. Two published methods were investigated for comparison. The best recognition rate selection obtained was for WPENN. The proposed method can offer a significant computational advantage by reducing the dimensionality of the WP coefficients by means of percent of energy distribution. Discrete Wavelet Transform (DWT) was studied to improve the system robustness against the noise of -2 dB. DWT approximation Sub-signal through several levels instead of original imposter had good performance on Additive White Gaussian Noise (AWGN) facing, particularly upon level 4.

Key words: Speech, gender, wavelet, energy, formants, neural network

INTRODUCTION

The most evident difference between male and female voices is fundamental frequency or pitch (Gelfer and Mikos, 2005). The average speaking fundamental frequency for men generally falls between 100 and 146 Hz whereas the average speaking fundamental frequency for women is generally between 188 and 221 Hz. These pitch levels aid a listener correctly identify the speaker's gender (Baken and Orlikoff, 2000). Nonetheless, in addition to fundamental frequency, resonance might also play a role in gender identification. Resonance is function of the supralaryngeal vocal tract. The air in the oral cavity, oropharynx, laryngopharynx and for some phonemes, the nasal cavities and nasopharynx vibrates at various frequencies in reply to the vibratory movement of the vocal folds and air passing through the glottis. These resonant frequencies related on the size and shape of the vocal tract and its constrictions (as well as tongue and lip positions which can change the functional length of the vocal tract). Vocal tract resonances are often studied as vowel formant frequencies. Because the male vocal tract is about 15% longer than the female vocal tract, the speech of men can be probable to have lower

formant frequencies than those considered characteristic of women (Bachorowski and Owren, 1999).

The source-filter model of vocal production (Chiba and Kajiyama, 1941) assumes that there is a source (the vocal cords) and a filter (the supralaryngeal vocal tract, hereafter referred to as vocal tract). In this model, the fundamental frequency of the voice is fixed to the rate of vocal fold vibration, whereas formant frequencies are the resonant frequencies of air in the vocal tract (Titze, 1994). In humans, the fundamental frequency of the voice is considered by the amount of testosterone present at the later stages of puberty which detects laryngeal size and vocal fold length (Butler *et al.*, 1989; Titze, 1994; Harries *et al.*, 1998). Fundamental frequencies are sexually dimorphic and steadily become lower during childhood development until puberty in both sexes. After this, male fundamental frequencies become lower quite rapidly until adulthood (Huber *et al.*, 1999). By contrast, the fundamental frequencies of females decrease at a pretty slower rate than those of males through puberty, resulting in adult fundamental frequencies that are about twice those in males (Bachorowski and Owren, 1999).

Some researchers have attempted to empirically separate fundamental frequency from formant frequencies

through the use of alternative voicing sources. For example, Schwartz and Rine (1968) studied gender identification for the whispered vowels /i/ and /α/. Outcomes showed that listeners were extremely accurate in their gender identifications. No misidentifications of the gender of the speaker were made for /α/ whereas four errors of gender identification were made for /i/. Spectral analysis showed that formants created by women were higher in frequency than that by men. The findings of Schwartz and Rine (1968) supported the hypothesis that vowel formant frequencies can be cues for gender perception. However, it could be argued that a male whisper is acoustically different from a female whisper and that, in fact, voice source information in addition to resonance information was available to listeners in identifying gender. Coleman (1971) degreased the variable of voice source by having all of his speakers use an electrolarynx with a fundamental frequency of 85 Hz. In study of Gelfer and Mikos (2005) 10 male and 10 female speakers generated the vowels /i/ and /u/ in isolation and read portions of the Rainbow Passage using a Western Electric Company Model 5 electrolarynx (Lookout Mountain, TN). The readings of the Rainbow Passage were introduced to 15 undergraduate speech students who were asked to recognize the speaker's gender and to rate their confidence in their selection. Results of the gender recognition (identification) procedure showed that the gender of male speakers was correctly recognized 98% of the time whereas the gender of female speakers was correctly recognized only 79% of the time. This difference in correct identifications might have been caused by the very low fundamental frequency of the electrolarynx which was more suitable to male speakers than to female speakers. Results of spectrographic analysis for formant frequencies in isolated vowels showed that, on the average, women had higher formant frequencies than men but statistical analyses were not completed. It was concluded that, even with a single-frequency sound source (the electrolarynx) and a complete absence of vocal fold vibratory characteristics, correct gender identification was possible; therefore, some information regarding gender must be conveyed by formant frequencies or vocal tract resonances.

Because the male vocal tract is about 15% longer than the female vocal tract, the speech signal taken from men are predicted to have lower formant frequencies than those considered characteristic of women (Bachorowski and Owren, 1999). Average Vowel Formants (AVF) (in Hertz) for Male Subjects, Transgender Subjects and Female Subjects are extracted by Gelfer and Mikos (2005), for F1 of /i/ vowel (in instance) as 283.16, 272.40 and 323.01, respectively for F2 as 2200.71, 2365.44 and

2614.15, respectively and for F3 as 2770.79, 3138.77 and 3230.26, respectively. A Fast Fourier Transform (FFT) was conducted with a single 1024-point Hamming window and Linear Predictive Coding (LPC). A frequently cited acoustic parameter as a cue to body size is mean fundamental frequency (F0) of voice (Morton, 1977; Laver and Trudgill, 1979; Xia and Espy-Wilson, 2000).

Experimental literature proposes solution for formant tracking, many are discussed by Malkin *et al.* (2005), such as Acero (1999) uses LPC spectral analysis to estimate potential formant frequencies. There have also been other types of formant trackers such as HMM-based methods (Deng *et al.*, 2003), approaches using nonlinear predictors (Deng *et al.*, 2004) and a current one using a Kalman filtering framework, to name a few.

In latest years, multi-resolution analysis based on wavelet theory was useful in many recognition tasks. Wavelet theory was proposed in 1984 where Goupillaud *et al.* (1984) introduced a new transformation for the frequency analysis of the discretized signals. The transform is known as wavelet transform (Goupillaud *et al.*, 1984; Daqrouq, 2011; Afshari, 2011). In the aspect of speaker identification, many works had developed the Mel filter-like structure to integrate the concept of Mel scale and multiresolution capabilities (Farooq and Datta, 2001; Karam *et al.*, 2000; Torres and Rufiner, 2000).

The ideal features for representing a speaker's identity should be substituted by some representative parameters to avoid complex computing (Wu and Lin, 2009). The advantage of WP parameters presented in energy form is that the model of extracted features will approach humans' auditory system; moreover, the number of parameters will be decreased. In this study, a WPT based percentage of energy corresponding to WP nodes approach is presented to improve the performance of gender identification system. The motivation of driving this work is to develop a different thought for linguistic recognition.

FEATURES EXTRACTION

For a certain orthogonal wavelet function, a library of wavelet packet bases is generated. Each of these bases offers an exacting way of coding signals, preserving global energy and reconstructing exact features. The wavelet packet is used to extract additional features to guarantee superior recognition rate. In this study, WPT is applied at the stage of feature extraction but these data are not proper for classifier due to an immense amount of data length. Thus, we have to seek for a better representation for the speech features. Previous studies

showed that the use of entropy of WP as features in recognition tasks is competent. Wu and Lin (2009) proposed a method to calculate the energy value of the wavelet coefficients in digital modulation recognition. Avci (2007) proposed features extraction method for speaker recognition based on a combination of three entropies types (sure, logarithmic energy and norm). Lastly, Avci (2009) investigated a speaker identification system using adaptive wavelet sure entropy. For a better demonstration of the sub-band signals, the energy of speech is commonly computed. Previous investigations showed that the utilization of an energy index as features in recognition roles is efficient. Kotnik *et al.* (2003) proposed a robust speech recognition scheme in a noisy environment by means of wavelet-based energy as a threshold for de-noise estimation. In the biomedical field, Behroozmand and Almasganj (2007) introduced a combination of genetic algorithm and wavelet packet transform used in the pathological evaluation and the energy features are computed from a group of wavelet packet coefficients. In study of Wu and Lin (2009), the energy indexes of WP were proposed for speaker identification.

As seen in above studies, the energy of the specific sub-band signal may be employed as features for recognition tasks. In this study, the energy corresponding to WP nodes approach will be employed for gender identification. The wavelet packet features extraction method can be summarized as follows (Derbel *et al.*, 2008):

- 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 the proposed work, the signals are normalized by using the following formula (Wu and Lin, 2009):

$$S_{Ni} = \frac{S_i - \bar{S}}{\sigma} \tag{1}$$

where, S_i is the i th element of the signal S , \bar{S} and σ are the mean and standard deviation of the vector S , respectively S_{Ni} is the i th element of the signal series S_N after normalization:

- Decomposing the speech signal by wavelet packet transform at level 7, with Daubechies type (db1)
- For a better demonstration of the sub-band signals, the energy corresponding to WP nodes approach will be calculated for gender identification (Fig. 1)

CLASSIFICATION

Speaker recognition with NN has recently undergone a significant development. Early experiments have exposed the potential of these methods for tasks with limited complexity. Many experiments have then been performed to test the ability of several NN models or approaches to the problem. Although most of these preliminary studies deal with a small number of speakers, they have shown that NN models were serious candidates for speaker identification tasks. NN classifiers like FFBPNN may lead to very good performances because they allow to take into account inter-speaker information and to build complex decision regions. However, the complexity of classification training procedures forbids

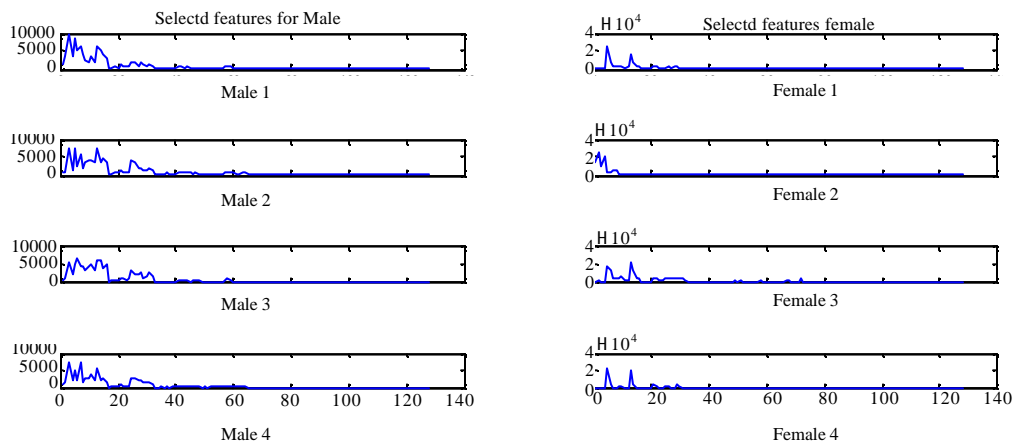


Fig. 1: The comparison between male and female using the proposed feature extraction method

Table 1: Parameters used for the network

Functions	Description
Network type	Feed forward back propagation
No. of layers	Four layers: Input, two hidden and output
No. of neurons in layers	128- Input, 30-Hidden and -4 Output
Weight function	DOTPROD
Training function	Levenberg-marquardt backpropagation
Activation functions	Log- sigmoid
Performance function (mse)	10 ⁻⁵
No. of epochs	200

the use of this simple approach when dealing with a large number of speakers (Qureshi and Jalil, 2002). Two solutions do emerge for managing large databases: Modular classification systems which a how to break the complexity of single NN architectures, or NN predictive models which tender a large variety of possible implementations.

This operation performs the intelligent classification by means of features obtained from feature extraction phase. In this study FFBPNN is used. The training condition and the structure of the NNT used in this study are as tabulated in Table 1. These were selected for the best performance. That is accomplished after several experiments, such as the number of hidden layers, the size of the hidden layers, value of the moment constant and type of the activation functions or transfer functions. 128×24 feature matrix which is obtained in features extraction stage is given to the input of the Feed-forward networks consist of several layers using the DOTPROD weight function, NETSUM net input function and the particular transfer functions. The weights of the first layer come from the input. Each network layer has a weight coming from the previous layer. All layers have biases. The last layer is the network output which we call target (T). In this study target is designed as a four binary digits for each features vector:

$$T = \begin{bmatrix} 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 1 & \dots & 1 \\ 1 & 0 & 1 & \dots & 0 \end{bmatrix} \quad (2)$$

The mean square error of the NN is achieved at the final of the training of the ANN classifier by means of Levenberg-Marquardt Backpropagation. Backpropagation is used to compute the Jacobian jX of performance with respect to the weight and bias variables X. Each variable is adapted according to Levenberg-Marquardt:

$$\begin{aligned} j_j &= jX * jX \\ j_e &= jX * E \\ dX &= -(j_j + I * Mu) \setminus j_e \end{aligned} \quad (3)$$

where, E is all errors and I is the identity matrix. The adaptive value Mu is increased by 10 Mu increase factor

until the change above outcomes in a reduced performance value. The change is then made to the network and Mu is decreased by 0.1 Mu decrease factor. After training the 24 (12 male and 12 female) speakers features, imposter simulation is performed. The imposter Simulation Result (SR) is compared with each of the 24 patterns target (P_n, n=1, 2, ..., 24) in order to determine the decision by:

$$C_n = 100 - [100 * \sqrt{(\sum (P_n - SR)^2 / \sum P_n^2)}] \quad (4)$$

where, C_n is the similarity percent between imposter simulation results and pattern target P_n. The gender is identified as patterns of maximum similarity percent. For instant, when most higher magnitudes of C_n belong to male patterns then decision is male.

RESULTS AND DISCUSSION

A testing database was created from Arabic language. The recording environment is a normal office environment via PC-sound card, with frequency 4000 Hz and sampling frequency 16000 Hz.

These utterances are Arabic spoken digits from 0 to 14. In addition, each speaker read ten separated 30 sec different Arabic texts. Total 45 individual speakers (19 to 40 years old) who are 22 individual male and 23 individual female spoken these Arabic words and texts for training and testing phases. The total number of tokens considered for training and testing was 1080.

It were performed experiments using total 1080 the Arabic utterances of total 29 individual speakers (22 male speakers and 23 female speakers). For each of these speakers, 24 speech signals were used. 6 of these signals were used for training. All of these signals were used for testing the WPENN expert system (Fig. 2). In this experiment, 89.35% correct classification was obtained by means of WPENN among the 45 different speaker signal classes. Testing results are tabulated in Table 2. It clearly indicates the usefulness and the trustworthiness of the proposed approach for extracting features from speech signals gender identification system.

Table 3 shows the experimental results of different approaches used in the experimental investigation for comparison. WPENN, DWT with proposed feature extraction method (at level 11), formants and pitch based gender identification method (FPFI) (Gelfer and Mikos, 2005) were investigated for comparison. The recognition

Table 2: WPENN identification rate results

Speaker	No. of signals	Recognized signals	Not recognized Signals	Recognition rate (%)
Sp.1	24	22	2	91.66
Sp.2	24	24	0	100.00
Sp.3	24	24	0	100.00
Sp.4	24	19	5	79.16
Sp.5	24	24	0	100.00
Sp.6	24	22	2	91.66
Sp.7	24	24	0	100.00
Sp.8	24	19	5	79.16
Sp.9	24	22	2	91.66
Sp.10	24	21	3	87.50
Sp.11	24	19	5	79.16
Sp.12	24	24	0	100.00
Sp.13	24	19	5	79.16
Sp.14	24	20	4	83.33
Sp.15	24	24	0	100.00
Sp.16	24	20	4	83.33
Sp.17	24	24	0	100.00
Sp.18	24	19	5	79.16
Sp.19	24	24	0	100.00
Sp.20	24	19	5	79.16
Sp.21	24	20	4	83.33
Sp.22	24	22	2	91.66
Sp.23	24	24	0	100.00
Sp.24	24	19	5	79.16
Sp.25	24	24	0	100.00
Sp.26	24	21	3	87.50
Sp.27	24	24	0	100.00
Sp.28	24	22	2	91.66
Sp.29	24	24	0	100.00
Sp.30	24	22	2	91.66
Sp.31	24	20	4	83.33
Sp.32	24	24	0	100.00
Sp.33	24	20	4	83.33
Sp.34	24	19	5	79.16
Sp.35	24	19	5	79.16
Sp.36	24	19	5	79.16
Sp.37	24	20	4	83.33
Sp.38	24	19	5	79.16
Sp.39	24	22	2	91.66
Sp.40	24	22	2	91.66
Sp.41	24	22	2	91.66
Sp.42	24	19	5	79.16
Sp.43	24	19	5	79.16
Sp.44	24	24	0	100.00
Sp.45	24	22	2	91.66
Total	1080	965	115	89.35

Table 3: Comparison of different classification approaches

Identification method	Identification system	No. of signals	Identification rate (%)
WPENN	Text-independent	1080	89.35
DWFNN	Text-independent	1080	82.25
GWPNN	Text-independent	1080	88.50

rate of DWFNN reached the lowest value. The best recognition rate selection obtained was 89.35% for WPENN.

To improve the robustness of WPENN to Additive White Gaussian Noise (AWGN), same wavelet packet decomposition process was applied to DWT approximation Sub-signal via several levels instead of original imposter (Daqrouq *et al.*, 2010). Afterwards, the features extraction was applied to each of the obtained

Table 4: WPENN Identification Rate results through DWT with SNR=-2dB

Speaker	No. of signals	Recognized signals (100%)			
		Level 1	Level 2	Level 3	Level 4
Sp.1	24	0	0	20	50
Sp.2	24	0	25	87	12
Sp.3	24	0	25	25	25
Sp.4	24	0	0	50	100
Sp.5	24	10	50	50	30
Sp.6	24	25	10	25	100
Total	144	5.83	18.33	42.83	52.91

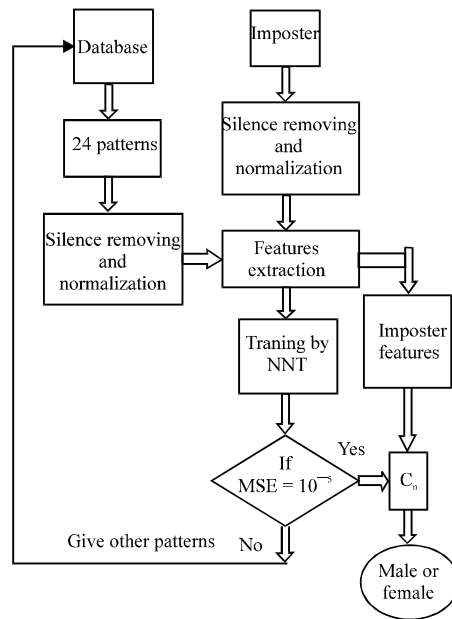


Fig. 2: Proposed expert system flow diagram of the WPENN

wavelet packet decomposition sub-signals. In Table 4, the percent of recognized signal is included upon four DWT levels. Six speakers are involved in the experiment. The 24 signals are utilized for each speaker. 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 (Table 4). In this experiment it was found that the recognition rates were not improved upon increasing the DWT level more than four.

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.

CONCLUSION

In this study, a new method for feature selection of gender based on percent of energy distribution of WP tree nodes and neural network using feed forward backpropagation algorithm is proposed. It leads to a significant improvement of the dimensional reduction from 128 features. The function of feature extraction and classification is performed using the WPENN expert system. The declared results show that the proposed method can make an efficient analysis. The performance of the proposed system was given in Table 2 and 3. The average identification rates were 89.35, better than other methods published before. The proposed method can offer a significant computational advantage by reducing the dimensionality of the WP coefficients by means of percent of energy distribution. DWT approximation Sub-signal through several levels instead of original imposter had good performance on AWGN facing, particularly upon level 4.

REFERENCES

- Acero, A., 1999. Formant analysis and synthesis using hidden markov models. Proceedings of the 6th European Conference on Speech Communication and Technology, Sept. 5-9, Budapest, Hungary, pp: 1047-1050.
- Afshari, M., 2011. An algorithm to analyze of two-dimensional function by using wavelet coefficients and relationship between coefficients. *Asian J. Applied Sci.*, 4: 414-422.
- Avci, E., 2007. A new optimum feature extraction and classification method for speaker recognition: GWPNN. *Expert Syst. Appl.*, 32: 485-498.
- Avci, D., 2009. An expert system for speaker identification using adaptive wavelet sure entropy. *Exp. Syst. Appl.*, 36: 6295-6300.
- Bachorowski, J.A. and M. Owren, 1999. Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *J. Acoust. Soc. Am.*, 106: 1054-1063.
- Baken, R. and R. Orlikoff, 2000. *Clinical Measurement of Speech and Voice*. 2nd Edn., Singular Publishing Group/Thomson Learning, San Diego, CA.
- Behroozmand, R. and F. Almasganj, 2007. Optimal selection of wavelet packet-based features using genetic algorithm in pathological assessment of patients speech signal with unilateral vocal fold paralysis. *Comput. Biol. Med.*, 37: 474-485.
- Butler, G.E., R.F. Walker, R.V. Walker, P. Teague, D. Riad-Fahmy and S.G. Ratcliffe, 1989. Salivary testosterone levels and the progress of puberty in the normal boy. *Clin. Endocrinol.*, 30: 587-596.
- Chiba, T. and J. Kajiyama, 1941. *The Vowel: Its Nature and Structure*. Tokyo- Kaiseikan Publishing Company Ltd., Tokyo.
- Coleman R.O., 1971. Male and female voice quality and its relationship to vowel formant frequencies. *J. Speech Hear. Res.*, 14: 565-577.
- Daqrouq, K., 2011. Wavelet entropy and neural network for text-independent speaker identification. *Eng. Appl. ArtificialIntell.*, 10.1016/j.engappai.2011.01.001
- Daqrouq, K., I.N. Abu-Isbeih, O. Daoud and E. Khalaf, 2010. An investigation of speech enhancement using wavelet filtering method. *Int. J. Speech Technol.*, 13: 101-115.
- Deng, L., I. Bazzi and A. Acero, 2003. Tracking vocal tract resonances using an analytical nonlinear predictor and a target-guided temporal constraint. Proceedings of the 8th European Conference on Speech Communication and Technology, Sept. 1-4, Geneva, Switzerland, pp: 73-76.
- Deng, L., L. Lee, H. Attias and A. Acero, 2004. A structured speech model with continuous hidden dynamics and prediction-residual training for tracking vocal track resonances. *Acoust. Speech Signal Process.*, 1: I-557-I-560.
- Derbel, A., F. Kallel, M. Samet and A.B. Hamida, 2008. Bionic wavelet transform based on speech processing dedicated to a fully programmable stimulation strategy for cochlear prostheses. *Asian J. Scientific Res.*, 1: 293-309.
- Farooq, O. and S. Datta, 2001. Mel filter-like admissible wavelet packet structure for speech recognition. *IEEE Signal Process. Lett.*, 8: 196-198.
- Gelfer, M.P. and V.A. Mikos, 2005. The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *J. Voice*, 19: 544-554.
- Goupillaud P., A. Grossmann and J. Morlet, 1984. Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, 23: 85-102.
- Harries, M., S. Hawkins, J. Hacking and I. Hughes, 1998. Changes in the male voice at puberty: Vocal fold length and its relationship to the fundamental frequency of the voice. *J. Laryngol. Ontol.*, 112: 451-454.
- Huber, J.E., E.T. Stathopoulos, G.M. Curione, T.A. Ash and K. Johnson, 1999. Formants of children, women and men: The effects of vocal intensity variation. *J. Acoust. Soc. Am.*, 106: 1532-1532.

- Karam, J.R., W.J. Phillips and W. Robertson, 2000. New low rate wavelet models for the recognition of single spoken digits. Proceedings of the Canadian Conference on Electrical and Computer Engineering, March 7-10, Halifax, NS, Canada, pp: 331-334.
- Kotnik, B., Z. Kacic and B. Horvat, 2003. Noise robust speech parameterization based on joint wavelet packet decomposition and autoregressive modeling. Proceedings of the 8th European Conference on Speech Communication and Technology, Sep. 1-4, Geneva, pp: 1793-1796.
- Laver, J. and P. Trudgill, 1979. Phonetic and Linguistic Markers in Speech. In: Social Markers in Speech, Giles, S. (Ed.). Cambridge University Press, Cambridge, pp: 1-32.
- Malkin, J., X. Li and J. Bilmes, 2005. A graphical model for formant tracking. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, March 18-23, University of Washington, Washington, DC, pp: 913-916.
- Morton, E.S., 1977. On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am. Nat.*, 111: 855-869.
- Qureshi, I.M. and A. Jalil, 2002. Object recognition using ANN with backpropogation algorithm. *J. Applied Sci.*, 2: 281-287.
- Schwartz, M.F. and H.E. Rine, 1968. Identification of speaker sex from isolated, whispered vowels. *J. Acoust. Soc. Am.*, 44: 1763-1767.
- Titze, I.R., 1994. Principles of Voice Production. Prentice Hall, Englewood Cliffs, New Jersey.
- Torres, H.M. and H.L. Rufiner, 2000. Automatic speaker identification by means of Mel cepstrum, wavelets and wavelets packets. *Eng. Med. Biol. Soc.*, 2: 978-981.
- Wu, J.D. and B.F. Lin, 2009. Speaker identification using discrete wavelet packet transform technique with irregular decomposition. *Expert Syst. Appl.*, 36: 3136-3143.
- Xia, K. and C. Espy-Wilson, 2000. A new strategy of formant tracking based on dynamic programming. *Proc. Int. Conf. Spoken Language Process.*, 3: 55-58.