World Applied Sciences Journal 8 (10): 1260-1266, 2010
ISSN 1818-4952

Reducing the Codebook Search Time in G.728 Speech Coder Using Fuzzy ARTMAP Neural Networks

Mansour Sheikhan and Sahar Garoucy

Department of Electrical Engineering, Islamic Azad University, South Tehran Branch, Tehran, Iran

Abstract: Codebook search has high computational load in code excited linear prediction (CELP) speech coders. In this paper, a fuzzy ARTMAP neural network (FAMNN) is used to determine the best index of shape codebook in ITU-T G.728 speech coding algorithm. In this way, the gain value is calculated according to this index and the best index of gain codebook is determined based on the minimum distance to each of eight gain codebook values. Empirical results show that the proposed model leads to 50.7% reduction in codebook search time as compared to the traditional implementation of ITU-T G.728 encoder. However, the degradations in mean opinion score (MOS), perceived evaluation of speech quality (PESQ) and segmental signal to noise ratio (SNR_{seg}) are not significant, as well.

Key words: Fuzzy ARTMAP % Codebook search % Speech encoder

INTRODUCTION

The high delay of conventional code excited linear prediction (CELP) algorithm degrades the communication quality [1]. So, a low delay-CELP (LD-CELP) algorithm was adopted by the International Telephone and Telegraph Consultative Committee (CCITT) for the coding of speech at 16 kbps with toll quality and became standardized in 1992 as G.728 [2, 3].

The application of LD-CELP is broad, including video-telephony, digital circuit multiplication equipment (DCME), packet circuit multiplication equipment (PCME), voice over Internet protocol (VoIP) and personal communication systems (PCS).

On the other hand, artificial neural networks (ANNs) have been used extensively for a variety of applications in speech and language technology (e.g. in speech synthesis [4, 5], automatic speech recognition (ASR) [6, 7] and natural language processing (NLP) [8] that experienced by the author for Farsi language). The researches on using ANNs in speech coding can also be classified into two main domains: neural predictors which improve the quality of coder [9-13] and reduction the computational complexity [14-19].

In the family of CELP coders, codebook search has high complexity. ANNs can be used to reduce this complexity. In this way, fuzzy adaptive resonance theory mapping (ARTMAP) is a neural network architecture that can establish the correct mapping between real-valued input patterns and correct labels in a variety of classification problems. In general, this family of neural networks include ART_1, ART_2 [20], ART_3 [21], ARTMAP [22], Fuzzy ART [23], ART-EMAP [24], dARTMAP [25], Boosted ARTMAP [26], Fuzzy ARTVar [27], Gaussian ARTMAP [28], µ_ARTMAP [29] and Fuzzy ARTMAP [30].

LD-CELP belongs to a family of speech coding algorithms, where a speech signal is modeled as the output of a linear prediction filter excited by an appropriate sequence. The 16 kbps LD-CELP coder consists of the following major components: an excitation sequence shape codebook, an excitation gain codebook, a 50^{th} order linear prediction coding (LPC) predictor, a 10^{th} order perceptual weighting filter and a mean square estimation computational block. The predictor and excitation gain are updated in a backward adaptive fashion by analyzing the previously quantized speech samples and excitation signal. Speech samples are buffered in blocks of 5 samples [2].

At 16 kbps, the LD-CELP encoder transmits a 10-bit index for every block of 5 speech samples which corresponds to an encoding rate of 2 bits per sample. The LD-CELP receiver decodes the best matched excitation vector out of 128 possible vectors, which is then multiplied by a decoded gain to form a final excitation sequence. This sequence becomes the input to the 50^{th} order LPC filter to synthesize the output speech.

Corresponding Author: Dr. Mansour Sheikhan, P.O. Box 11365/4435, Post-Graduate Center, South Tehran Branch, Islamic Azad University, Iran, Tel: +98 21 88605565

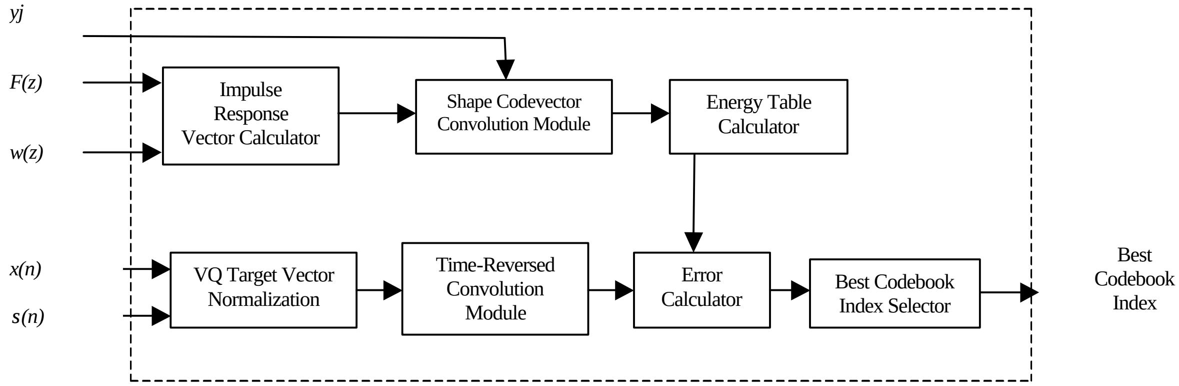

Fig. 1: Block diagram of search codebook module in G.728 [3]

In this paper, a fuzzy ARTMAP neural network (FAMNN) is used as shape codebook search module for G.728 speech coding algorithm. So, this paper is organized as follows. In Section 2, the codebook search module in G.728 encoder is introduced. Section 3 gives a review on FAMNN. The details of shape codebook search using FAMNN are discussed in Section 4. The simulation and empirical results are reported in Section 5 and conclusions are drawn in Section 6.

CODEBOOK SEARCH IN G.728: In the G.728 encoder structure, zero-input response vector *r(n)* subtracts from the vector quantization (VQ) weighted speech vector *v(n)* to obtain the VQ codebook search target vector *x(n)*. The excitation gain *F(n)* is also obtained with closed loop method. Each of the 1024 candidate codevectors is scaled by the current excitation gain and then is passed through a cascaded filter consisting of the synthesis filter *F(z)* and the perceptual weighting filter *W(z)* [3].

The detailed block diagram of G.728 search codebook module is shown in Figure 1. The following formula is used for codebook search [3]:

$$D = s^2(n)\left\|\hat{x}(n) - g_i H(n) y_j\right\|^2 \tag{1}$$

in which, $\hat{x}(n)$ is the target vector adjusted by *F(n)*. *H(n)* is the unit impulse response matrix of the short–term predictor, g_i is the ith level in the 3-bit gain codebook and y_j is the jth codevector in the 7-bit shape codebook [3].

According to G.728 recommendation, minimizing *D* is equivalent to maximizing D_{max} [3]:

$$D_{\max} = 2 g_i p^T(n) y_j - g_i^2 E_j \tag{2}$$

in which, $p(n) = H^T \hat{x}(n)$ and $E_j = \left\|Hy_j\right\|^2$.

The inner product term, $p_j = H^T \hat{x}(n) y_j$, which solely depends on *j*, takes the most of the computation in determining D_{max}. Once the best indices for *i* and *j* are identified, they are concatenated to form the output of the codebook search module (a single 10-bit index). The 10-bit codebook index consists of two portions: 3 bits for gain codebook (b_0-b_2: 8 scalar values) and 7 bits for shape codebook (b_3-b_9: 128 codevectors).

FUZZY ARTMAP NEURAL NETWORK: The fuzzy ARTMAP neural network (FAMNN) has been introduced by Carpenter *et al*. [30]. The FAMNN has been successfully applied in many tasks such as data mining, remote sensing and pattern recognition. FAMNN is considered fast among members of ARTMAP family due to the computationally cheap mapping between inputs and outputs. Furthermore, compared to the standard nearest neighbor techniques which are also commonly used, FAMNN requires less memory since it uses a compressed representation of the data and for the same reason FAMNN requires less classification time. The FAMNN is a supervised network which is composed of two fuzzy ART modules, ART_a and ART_b, as shown in Figure 2. Those modules are linked together via an inter-ART module, F^{ab}, called a map field. The map field is used to realize the match tracking rule, whereby the vigilance parameter of ART_a increases in response to a predictive mismatch at ART_b. Match tracking reorganizes category structure so that predictive error is not repeated on subsequent presentations of the input.

As shown in Figure 2, variables in ART_a and ART_b modules are shown by *a* or *b* subscripts or superscripts. Inputs to each module are in the complement code form, $A = (a, a^c)$ for ART_a module and $B = (b, b^c)$ for ART_b module, respectively. For ART_a module, $x^a = (x_1^a, \ldots, x_{2M_a}^a)$ represents the F_1^a output vector, $y^a = (y_1^a, \ldots, y_{2N_a}^a)$ represents the F_2^a output vector and $w_j^a = (w_{j1}^a, \ldots, w_{j,2M_a}^a)$

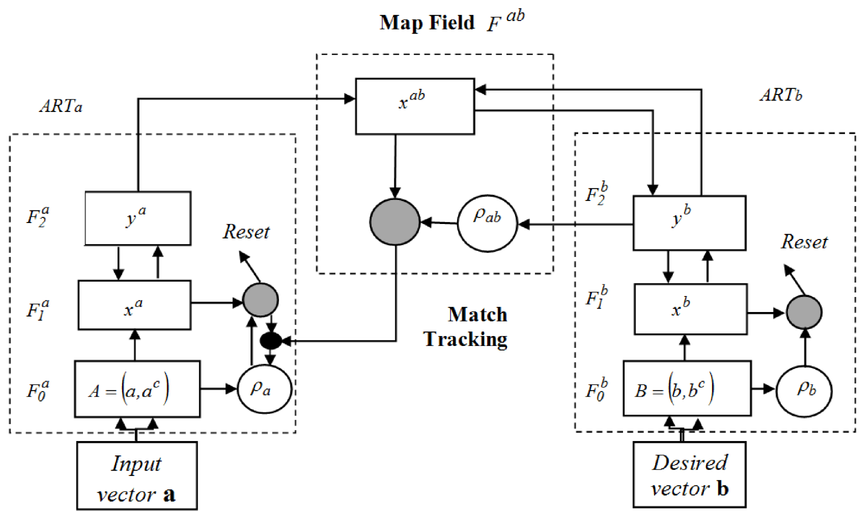

Fig. 2: Fuzzy ARTMAP structure

represents the jth weight vector. For ART$_b$ module, $x^b=(x_1^b,...,x_{2M_b}^b)$ represents the F_1^b output vector and $y^b=(y_1^b,...,y_{N_b}^b)$ represents the F_2^b output vector and $w_k^b=(w_{k1}^b,...,w_{k,2M_b}^b)$ represents the kth weight vector. For the map field, let $x^{ab}=(x_1^{ab},...,x_{N_b}^{ab})$ represents the F^{ab} output vector and $w_j^{ab}=(w_{j1}^{ab},...,w_{jN_b}^{ab})$ represents the weight vector from jth F_2^a node to F^{ab}. The map field F^{ab} is activated whenever one of the ART_a or ART_b categories is active. If node J of F^{ab} is chosen, then its weights w_J^{ab} activate F^{ab}. If node K of F_2^b is active, then the node K in F_2^b is activated by 1-1 pathways between F_2^b and F^{ab}. The F^{ab} output vector x^{ab} obeys:

$$x^{ab}=\begin{cases} y^b \wedge w_J^{ab}; & J^{th} \textit{ node of } F_2^a \textit{ is active and } F_2^b \textit{ is active} \\ w_J^{ab}; & J^{th} \textit{ node of } F_2^a \textit{ is active and } F_2^b \textit{ is inactive} \\ y^b; & F_2^a \textit{ is inactive and } F_2^b \textit{ is active} \\ 0; & F_2^a \textit{ and } F_2^b \textit{ are inactive} \end{cases} \quad (3)$$

Fuzzy ARTMAP module vectors are based on two separately distance criteria, match and choice. The match function is defined by:

$$S_j(I)\equiv\frac{|I\wedge w_J|}{|I|} \quad (4)$$

where w_j is an analog-valued weight vector associated with cluster j, ∨ denotes the fuzzy AND operator, $(p\wedge q)_i\equiv\min(p_i,q_i)$ and $|.|$ denotes the norm operator. The choice function is defined in Equation (5), where ″ is a small constant. Input vector (I) is assigned to the category which maximizes $T_j(I)$ while satisfying $S_j(I)$ $D, where the vigilance, D, is a constant, 0#D#1. The fuzzy ARTMAP learning rule is given in Equation (6):

$$T_j(I)=\frac{|I\wedge w_j|}{a+|w_j|} \quad (5)$$

$$w_{Ji}^{new}=\begin{cases} w_{Ji}^{old} & ; w_{Ji}\le I_i \\ w_{Ji}^{old}-b(w_{Ji}^{old}-I_i) & ; w_{Ji}>I_i \end{cases} \quad (6)$$

where 0<$#1. All w_{ij} are initially set to 1.

The map field is fundamentally a look-up table, retrieving an analog-valued weight w_{JL}^{ab} when node J of module a and node L of module b are active. Note that only one node of each module is active at a given time. If $w_{JL}^{ab}<r^{ab}$, then the vigilance of module a, D^a, is raised until node J becomes inactive (and some other node becomes active). This process is repeated until $w_{JL}^{ab}\ge r^{ab}$. When the next input is presented, D^a is returned to its baseline value. All w_{JL}^{ab} are initially set to 1. During learning, when $w_{JL}^{ab}\ge r^{ab}$ and J and L nodes become active, w_{Jl}^{ab}; l … L is reduced in value (typically set to 0).

FAMNN-BASED SHAPE CODEBOOK SEARCH: In this paper, the FAMNN is used to determine the best index of shape codebook. The occurrence frequency characteristics of codevectors in shape codebook have not the uniform distribution [31]. The experiments show that the occurrence frequency characteristics of shape codevectors have not so significant dependency on the gain scalar values, as well. According to the results that are reported in [32], occurrence probability of shape codebook index values in the range of 65 to 128 is much higher than index values in the range of 1 to 64.

The proposed model for FAMNN-based shape codebook search is shown in Figure 3. In this way, to determine the best index of gain codebook, the gain is calculated by Equation (7):

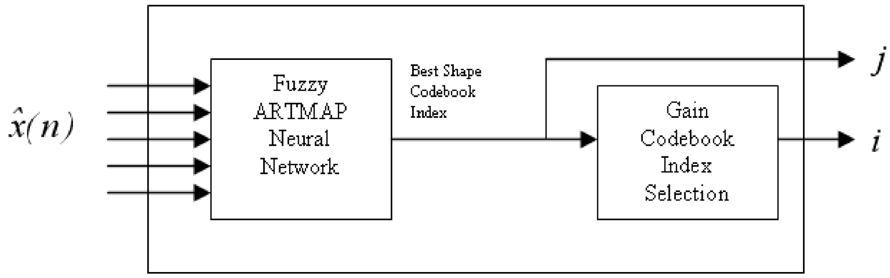

Fig. 3: Proposed FAMNN-based codebook search module

Table 1: Specifications of FAMNN in the proposed model

Specification	Value
Learning rate	0.98
Vigilance parameter	0.99
Number of F_0 nodes	10
Number of F_1 nodes	1235
Number of F_2 nodes	1235
Training time (sec)	721
Number of classes	128
Number of training samples	100,000
Number of test samples	3200
Correct identification rate	95%

Table 2: Performance of the proposed model equipped with neural search shape codebook module

System	Codebook search time (sec)/frame	SNR_{seg} (dB)	MOS	PESQ
Traditional G.728 [2, 34]	5.56/200	18.45	3.91	3.43
Proposed model	2.74/200	18.36	3.83	3.37

$$\hat{g} = \frac{p_j}{E_j} \tag{7}$$

Based on the calculated gain, the best index of gain codebook is determined by finding the one of eight scalar values of gain codebook that is closest to $\hat{g}$.

SIMULATION AND EMPIRICAL RESULTS: In this work, as the first step, LD-CELP vocoder is implemented based on the ITU-T G.728 recommendation [3]. The simulation of encoder and decoder is performed by MATLAB v.7.2 simulation software. The sampling frequency is 8 kHz and the frame size is 20 samples.

In this work, we use Farsi speech data files of FARSDAT [33]. FARSDAT is a continuous speech Farsi corpus including 6000 utterances from 300 speakers with various accents. Training dataset in our work includes 100,000 frames of speech from 24 male and 28 female speakers. Test dataset includes 3200 frames of speech from one male and two female speakers, as well. In the training dataset, the utterance of two sentences by each speaker is included. The test dataset includes the utterance of one sentence by each speaker. It is noted that the proposed model can show its efficiency in reducing the codebook search time, if it performs well on a rich small-size test dataset, too.

It is noted that the input of FAMNN is gain-normalized VQ target vector, $\hat{x}(n)$ (Fig. 3). The specification of simulated FAMNN is reported in Table 1. The effect of vigilance parameter and learning rate values on prediction accuracy is investigated in this work. The value of D is set to 0.99 in our experiments (Figure 4) to have an acceptable number of F_2 nodes and prediction accuracy. The value of $ is set to 0.98, as well (Figure 5). The performance of proposed model, that is equipped with neural shape codebook index prediction, is compared with the traditional G.728 [2, 34] in terms of segmental SNR (SNR_{seg}), mean opinion score (MOS) and perceived evaluation of speech quality (PESQ) measures (Table 2). SNR_{seg} is an important factor in determining the quality of audio data [35]:

$$SNR = 10\log\left(\frac{\sum_n x^2(n)}{\sum_n (x^2(n) - y^2(n))^2}\right) \tag{8}$$

where *x(n)* is the input signal to encoder and *y(n)* is the output signal from decoder. SNR_{seg} is defined as the average of SNR measurements:

$$SNR_{seg} = \frac{1}{N}\sum_{m=1}^{N} SNR_m \tag{9}$$

in which, *N* is the number of frames.

It is noted that MOS provides a numerical indication of the perceived quality of received media after compression and/or transmission. The MOS is expressed as a single number in the range of 1 to 5, where 1 is the lowest perceived audio quality and 5 is the highest perceived audio quality measurement [36].

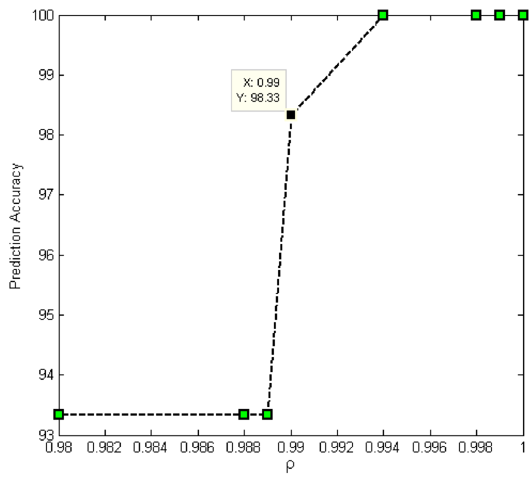

Fig. 4: Effect of vigilance parameter on prediction accuracy ($=1)

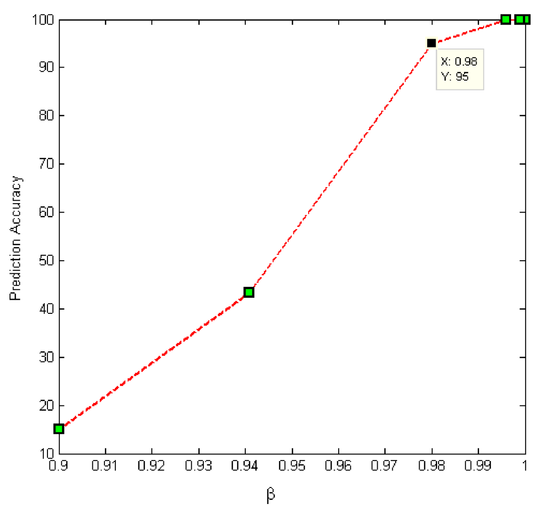

Fig. 5: Effect of learning rate on prediction accuracy (D=0.99)

The inherent problem in subjective MOS measurement is that it is slow, time consuming, expensive and cannot be used for long-term or large scale. This has made objective methods very attractive for meeting the demand for voice quality measurement in communication networks. Objective measurement of voice quality in modern communication networks can be intrusive or non-intrusive [37]. Intrusive methods are more accurate, but normally are unsuitable for monitoring live traffic because of the need for a reference data and to utilize the network. A typical intrusive method is based on the ITU-T P.862 standard, perceived evaluation of speech quality (PESQ) measurement algorithm [38]. This involves a comparison of the reference and the degraded speech signals to predict the listening-only one-way MOS score.

The output of PESQ, termed PESQ score, has high correlation with MOS for a wide range of subjective tests spanning many different l anguages and network types. However, PESQ score was calibrated against an essentially arbitrary objective distortion scale. It was not designed to be on exactly the same scale as MOS, either in general or for any specific subjective test. PESQ score may be between -0.5 and 4.5, while absolute category rating (ACR) listening quality MOS is on a 1-5 scale [36]. In this paper, MOS and PESQ scores are used to assess the voice quality of proposed model.

As shown in Table 2, using FAMNN as a neural search shape codebook in G.728 speech coder reduces the computational complexity noticeably, without significant degradation in MOS, PESQ and SNR_{seg}. As a visual indication of coder performance, in addition to subjective measures, the input and output of the proposed enhanced G.728 speech vocoder are depicted in Fig. 6, as well.

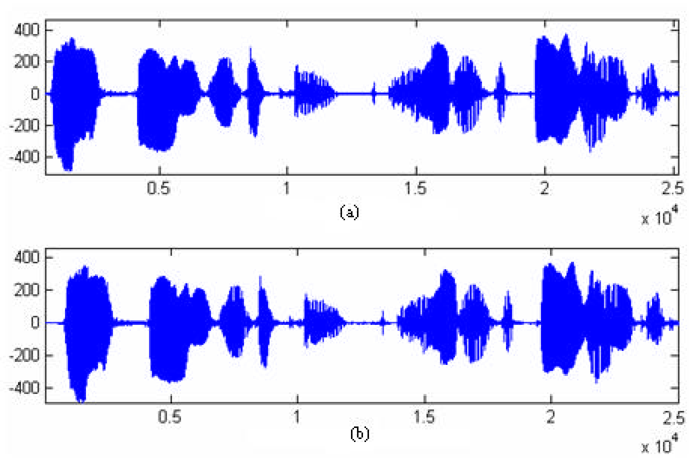

Fig. 6: Input and output of the proposed enhanced speech vocoder, a) input, b) output

CONCLUSIONS

In this paper, shape codebook search part in the structure of ITU-T G.728 speech coder was replaced by a fuzzy ARTMAP neural network (FAMNN). In this way, gain value was calculated analytically according to the best shape codebook index. This index was determined by FAMNN (Equation 7). Then, the best gain codebook index was selected based on the minimum distance to each of 8 gain codebook values.

The proposed model led to 50.7% reduction in codebook search time as compared to the traditional implementation of ITU-T G.728 encoder (Table 2). However, SNR_{seg}, MOS and PESQ were reduced 0.09 dB, 0.08 and 0.06, respectively. Therefore, the proposed structure led to noticeable reduction in computational complexity, without significant degradation in SNR_{seg}, MOS and PESQ.

REFERENCES

1. Roychoudhuri, L., E. Al-Saher and G.B. Brewster, 2006. On the Impact of Loss and Delay Variation on Internet Packet Audio Transmission. Computer Communications, 29: 1578-1589.
2. Chen, J.H., R.V. Cox, Y.C. Lin and N. Jayant, 1992. A Low Delay CELP Coder for CCITT 16 kb/s Speech Coding Standard. IEEE Journal on Selected Areas on Commmunication, 10: 830-847.
3. ITU-T G.728 Recommendation, 1992. Coding of Speech at 16 kb/s Using Low-Delay Code Excited Linear Prediction.
4. Sheikhan, M., 2003. Prosody Generation in Farsi Language. In the Proceedings of the International Symposium on Telecommunications, pp: 250-253.
5. Sheikhan, M., M. Nasirzadeh and A. Daftarian, 2006. Text to Speech for Iraninan Dialect of Farsi Language. In the Proceedings of the Second Workshop on Farsi Computer Speech, University of Tehran, pp: 39-53.
6. Sheikhan, M., M. Tebyani and M. Lotfizad, 1997. Continuous Speech Recognition and Syntactic Processing in Iranian Farsi Language. International Journal of Speech Technology, 1: 135-141.
7. Sheikhan, M., 2003. Suboptimum Extracted Features and Classifier for Speaker-Independent Farsi Digit Recognizer. In the Proceedings of the International Symposium on Telecommunications, pp: 246-249.
8. Sheikhan, M., M. Tebyani and M. Lotfizad, 1996. Using Symbolic and Connectionist Approaches to Automate Editing Persian Sentences Syntacticly. In the Proceedings of the International Conference on Intelligent & Cognitive Systems, pp: 250-253.
9. Birgmeier, M., 1996. Nonlinear Prediction of Speech Signals Using Radial Basis Function Networks. In the Proceedings of the European Signal Processing Conference, vol. 1, pp: 459-462.
10. Faundez, M., 1999. Adaptive Hybrid Speech Coding with a MLP/LPC Structure. In the Proceedings of the International Work-Conference on Artificial and Natural Neural Networks, 11: 814-823.
11. Sassi, S.B., R. Braham and A. Belghith, 2001. Neural Speech Synthesis System for Arabic Language Using CELP Algorithm. In the Proceedings of the ACS/IEEE International Conference on Computer Systems and Applications, pp: 119-121.
12. Faúndez-Zanuy, M., 2003. Nonlinear Speech Coding with MLP, RBF and Elman Based Prediction. Lecture Notes in Computer Science, 2687: 671-678.
13. Sheikhan, M., V. Tabataba Vakili and S. Garoucy, 2009. Complexity Reduction of LD-CELP Speech Coding in Prediction of Gain Using Neural Networks. World Applied Sciences Journal, 7 (Special Issue of Computer & IT): 38-44.
14. Easton, M.G. and C.C. Goodyear, 1991. A CELP Codebook and Search Technique Using a Hopfield Net. In the Proceedings of the International Conference on Acoustics, Speech and Signal Processing, pp: 685-688.
15. Hernandez-Gomez, L.A. and E. Lopez-Gonzalo, 1993. Phonetically-Driven CELP Coding Using Self-Organizing Maps. In the Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 2: 628-631.
16. Zhang, G., K. Xie, Z. Zhao and C. Xue, 2006. The LD-CELP Gain Filter Based on BP NN. Lecture Notes in Computer Science, 3973: 150-155.
17. Wu, S., G. Zhang, X. Zhang and Q. Zhao, 2008. A LD-CELP Speech Coding Algorithm Based on Modified SOFM Vector Quantizer. In the Proceedings of the International Symposium on Intelligent Information Technology Applications, pp: 408-411.
18. Huong, V., B.J. Min, D.C. Park and D.M. Woo, 2008. A New Vocoder Based on AMR 7.4 kbit/s Mode in Speaker Dependent Coding System. In the Proceedings of the ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, pp: 163-167.

19. Sheikhan, M., V. Tabataba Vakili and S. Garoucy, 2009. Codebook Search in LD-CELP Speech Coding Algorithmin Based on Multi-SOM Structure. World Applied Sciences Journal, 7 (Special Issue of Computer & IT): 59-68.
20. Carpenter, G.A. and S. Grossberg, 1987. ART_2: Self-Organization of Stable Category Recognition Codes for Analog Input Patterns. Applied Optics, 26: 4919-4930.
21. Carpenter, G.A. and S. Grossberg, 1990. ART_3: Hierarchical Search Using Chemical Transmitters in Self-Organizing Pattern Recognition Architectures. Neural Networks, 3: 129-152.
22. Carpenter, G.A., S. Grossberg and J.H. Reynolds, 1991. ARTMAP: Supervised Real-time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network. Neural Networks, 4: 565-588.
23. Carpenter, G.A., S. Grossberg and D.B. Rosen, 1991. Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System. Neural Networks, 4: 759-771.
24. Carpenter, G.A. and W. Ross, 1995. ART-EMAP: A Neural Network Architecture for Learning and Prediction by Evidence Accumulation. IEEE Transactions on Neural Networks, 6: 805-818.
25. Carpenter, G.A., B.L. Milenova and B.W. Noeske, 1998. Distributed ARTMAP: A Neural Network for Fast Distributed Supervised Learning. Neural Networks, 11: 793-813.
26. Verzi, S.J., G.L. Heileman, M. Georgiopoulos and M.J. Healy, 1998. Boosted ARTMAP. In the Proceedings of the IEEE World Congress on Computational Intelligence, pp: 396-400.
27. Dagher, I., M. Georgiopoulos, G.L. Heileman and G. Bebis, 1998. Fuzzy ARTVar: An Improved Fuzzy ARTMAP Algorithm. In the Proceedings of the IEEE World Congress on Computational Intelligence, pp: 1688-1693.
28. Williamson, J., 1996. Gaussian ARTMAP: A Neural Network for Fast Incremental Learning of Noisy Multidimensional Maps. Neural Networks, 9: 881-897.
29. Gomez-Sanchez, E., Y.A. Dimitriadis, J.M. Cano-Izquierdo and J. Lopez-Coronado, 2002. µ_ARTMAP: Use of Mutual Information for Category Reduction in Fuzzy ARTMAP. IEEE Transactions on Neural Networks, 13: 58-69.
30. Carpenter, G.A., S. Grossberg, N. Markuzon, J.H. Reynolds and D.B. Rosen, 1992. Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps. IEEE Transactions on Neural Networks, 3: 698-713.
31. Kawano, N., H. Yajima, A. Hotta and Y. Naito, 1995. A Variable Bit-Rate LD-CELP Speech Coder at 16, 12.8 and 9.6 kbit/s. In the Proceedings of the IEEE Workshop on Speech Coding for Telecommunications, pp: 95-96.
32. CCITT, 1992. A Statistical Characteristic of 16 kbit/s LD-CELP Shape Codebook Used in the Draft Recommenation G.728.COM-XV-175-E.
33. Bijankhan, M., J. Sheikhzadegan, M.R. Roohani, Y. Samareh, C. Lucas and M. Tebiani, 1994. The Speech Database of Farsi Spoken Language. In the Proceedings of the Fifth Australian International Conference on Speech Science and Technology (SST'94), pp: 826-831.
34. Zahang, G., K.M. Xie, X.Y. Zhang and L.Y. Huangfu, 2006. How to Optimize the Gain Filter of LD-CELP. In the Proceedings of the Information and Communication Technologies Conference, 1: 1206-1211.
35. Deller, J.R., J.H.L. Hansen and J.G. Proakis, 2000. Discrete-Time Processing of Speech Signals. IEEE Press.
36. ITU-T P. 800 Recommendation, 1996. Methods for Subjective Determination of Transmission Quality.
37. Al-Akhras, M., H. Zedan, R. John and I. Al-Momani, 2009. Non-Intrusive Speech Quality Prediction in VoIP Networks Using a Neural Network Approach. Neurocomputing, 72: 2595-2608.
38. ITU-T P. 862 Recommendation, 2001. Perceptual Evaluation of Speech Quality (PESQ), An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs.