Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information). 9/2 (2016), 80-87
DOI: http://dx.doi.org/10.21609/jiki.v9i2.382

A NOVEL APPROACH TO STUTTERED SPEECH CORRECTION

Alim Sabur Ajibola, Nahrul Khair bin Alang Md. Rashid, Wahju Sediono, and Nik Nur Wahidah Nik Hashim

Mechatronics Engineering Department, International Islamic University Malaysia, Jalan Gombak, Kuala Lumpur, 53100, Malaysia

E-mail: moaj1st@yahoo.com, wsediono@iium.edu.my

Abstract

Stuttered speech is a dysfluency rich speech, more prevalent in males than females. It has been associated with insufficient air pressure or poor articulation, even though the root causes are more complex. The primary features include prolonged speech and repetitive speech, while some of its secondary features include, anxiety, fear, and shame. This study used LPC analysis and synthesis algorithms to reconstruct the stuttered speech. The results were evaluated using cepstral distance, Itakura-Saito distance, mean square error, and likelihood ratio. These measures implied perfect speech reconstruction quality. ASR was used for further testing, and the results showed that all the reconstructed speech samples were perfectly recognized while only three samples of the original speech were perfectly recognized.

Keywords: s*tuttered speech, speech reconstruction, LPC analysis, LPC synthesis, objective quality measure*

Abstrak

Shuttered speech adalah *speech* yang kaya dysfluency, lebih banyak terjadi pada laki-laki daripada perempuan. Ini terkait dengan tekanan udara yang tidak cukup atau artikulasi yang buruk, meskipun akar penyebabnya lebih kompleks. Fitur utama termasuk *speech* yang berkepanjangan dan berulang-ulang, sementara beberapa fitur sekunder meliputi, kecemasan, ketakutan, dan rasa malu. Penelitian ini menggunakan *LPC analysis* dan *synthesis* algoritma untuk merekonstruksi *stuttered speech*. Hasil dievaluasi menggunakan jarak cepstral, jarak Itakura-Saito, mean square error, dan rasio *likelihood*. Langkah-langkah ini terkandung kualitas *speech reconstruction* yang sempurna. ASR digunakan untuk pengujian lebih lanjut, dan hasilnya menunjukkan bahwa semua sampel *speech* yang terekonstruksi dikenali dengan sempurna sementara hanya tiga sampel dari *speech* asli dikenali dengan sempurna.

Kata Kunci: s*tuttered speech, speech reconstruction, LPC analysis, LPC synthesis, objective quality measure*

1. Introduction

The aim of this study is to develop a novel approach for stuttered speech correction using speech reconstruction. Human beings express their feelings, opinions, views and notions orally through speech. Speech includes articulation, voice, and fluency [1,2]. It is a complex naturally acquired human motor skills, an action characterized in normal grownups by the production of about 14 different sounds per second via the coordinated actions of about 100 muscles connected by spinal and cranial nerves. The ease with which human beings speak is in contrast to the complexity of the act, and that complexity may help explain why speech can be exquisitely sensitive to the nervous system associated diseases [3]. Nearly 2% and 5% of adults and children stutter respectively [4,5].

Stuttering can also be defined as a disruption in the normal flow of speech unintentionally by dysfluencies, which include repetitive pronunciation, prolonged pronunciation, blocked or stalled pronunciation at the phoneme or the syllable level [6-8]. Stuttering cannot be permanently cured, however, it may go into remission after some time, or stutterers can learn to shape their speech into fluent speech with the appropriate speech pathology treatment. This shaping has its effects on the tempo, loudness, effort, or duration of their utterances [7,9].

Stuttering has been found to be more prevalent in males than females (ratio 4:1) [1,2,6,9,10]. Stutterers and non-stutterers alike have speech dysfluencies, which are gaffes or disturbances in the flow of words a speaker plans to say, but dysfluencies are more observable in stutterers' speech [11]. Stuttered speech is rich in dysfluencies, usually repetitions. Classical approaches to the

analysis of dysfluencies are in very short intervals, which is sufficient for recognition of simple repetitions of phonemes [12].

In order to achieve the reconstruction, the linear prediction coefficient (LPC) was used. It was used because its algorithm models the human speech production. The reconstructed speech was then evaluated using objective speech quality measures such as cepstral distance (CD), mean square error (MSE), Itakura-Saito distance (IS) and likelihood ratio (LR). Automatic speaker recognition (ASR) system was developed to further evaluate and compare between the original speech and the reconstructed speech.

2. Methods

The methodologies used for the actualization of this research are described in this section. The LPC analysis and synthesis, the line spectral frequency (LSF) for feature extraction and the multilayer perceptron (MLP) as classifier are explained.

LPC Speech Reconstruction

Linear predictive coding (LPC) is most widely used for medium or low bit-rate speech coders [13]. From each frame of the speech samples, the reflection coefficients are computed. Because important information about the vocal tract model is extracted in the form of reflection coefficients, the output of the LPC analysis filter using reflection coefficients will have less redundancy than the original speech. Thus, less number of bits is required to quantize the residual error. This quantized residual error along with the quantized reflection coefficients are transmitted or stored. The output of the filter, termed the residual error signal, has less redundancy than original speech signal and can be quantized by a smaller number of bits than the original speech. The speech is reconstructed by passing the residual error signal through the synthesis filter. If both the linear prediction coefficients and the residual error sequence are available, the speech signal can be reconstructed using the synthesis filter.

Speech Analysis Filter

Linear Predictive Coding is the most efficient form of coding technique [14, 15] and it is used in different speech processing applications for representing the envelope of the short-term power spectrum of speech. In LPC analysis of order '*p*' the current speech sample *s*(*n*) is predicted by a linear combination of p past samples *k*, and given by equation(1) [16].

$$\hat{s}(n) = \sum_{k=1}^{p} a_p(k).s(n-k) \tag{1}$$

where $\hat{s}(n)$ is the predictor signal and $\{a_p(1), ..., a_p(p)\}$ are the LPC coefficients. The residual signal $e(n)$ is derived by subtracting $\hat{s}(n)$ from $s(n)$ and the reduced variance is given by the equation-(2).

$$\begin{aligned} e(n) &= s(n) - \hat{s}(n) \\ &= \hat{s}(n) - \sum_{k=1}^{p} a_p(k).s(n-k) \end{aligned} \tag{2}$$

By applying the Z-transform to the equation which gives rise to the equation(3).

$$E(z) = A_p(z).S(z) \tag{3}$$

where $S(z))$ and $E(z)$ are the transforms of the speech signal and the residual signal respectively, and $A_p(z)$ is the LPC analysis filter of order $'p'$ as given by equation(4).

$$A_p(z) = 1 - \sum_{k=1}^{p} a_p(k)\, z^{-k} \tag{4}$$

The short-term correlation of the input speech signal is removed by giving an output $E(z)$ with more or less flat spectrum. After implementation of analysis filter, the quantization techniques are implemented and the speech signal is to be brought from the quantized signal at the receiver and so the quantized signal is to be synthesized to get the speech signal.

Speech Synthesis Filter

The short-term power spectral envelope of the speech signal can be modelled by the all-pole synthesis filter as given by equation(5) [16]:

$$H_p(z) = \frac{1}{A_p(z)} = \frac{1}{1 - \sum_{k=1}^{p} a_p(k)\, z^{-k}} \tag{5}$$

The equation(5) is the basis for the LPC analysis model. On the other hand, the LPC synthesis model consists of an excitation source $E(z)$, which provides input to the spectral shaping filter H_p (z), which will give the synthesized output spe-ech $S(z)$ as given by equation(6) [14]:

$$S(z) = H_p(z).E(z) \tag{6}$$

In order to identify the sound whether it is voiced or unvoiced, the LPC analysis of each frame can act as a decision-making process. The impulse train is used to represent voiced signal, with non-zero taps occurring for every pitch period. To determine the correct pitch period/frequency, a pitch-detecting algorithm is used. The pitch period can be estimated using autocorrelation function. However, if the frame is unvoiced, then the white noise is used to represent it and a pitch period of T=0 is transmitted [14-15].

Therefore, either white noise or impulse train becomes the excitation of the LPC synthesis filter. Hence, it is important to emphasize on the pitch, gain and coefficient parameters that will be varying with time and from one frame to another. The above model is often called the LPC Model. This model speaks about the digital filter (called the LPC filter) whose input is either a train of impulses or a white noise sequence and the output is a digital speech signal [14-15].

Feature Extraction

In general, most speech feature extraction methods fall into the following two categories: modelling the human voice production system or modelling of the peripheral auditory system [17]. Feature extraction consists of computing representations of the speech signal that are robust to acoustic variation but sensitive to linguistic con-tent [18]. It is executed by converting the speech waveform to some type of parametric representation for further analysis and processing. This representation is effective, suitable and discriminative than the original signal [19]. The feature extraction plays a very important role in speech identification. As a result of irregularities in human speech features, human speech can be sensibly interpreted using frequency-time interpretations such as a spectrogram [20].

Line Spectral Frequency (LSF)

Line Spectral Frequency (LSF) exhibits ordering and distortion independence properties. The-se properties enable the representation of the high frequencies associated with less energy using fewer bits [21]. LSF's are an alternative to the direct form predictor coefficients or the lattice form reflection coefficients for representing the filter response. The direct form coefficient representation of the LPC filters is not conducive to an efficient quantization. Instead, nonlinear functions of the reflection coefficients are often used as transmission parameters. These parameters are preferable because they have a relatively low spectral sensitivity [22]. It has been found that the line spectral frequency (LSF) representation of the predictor is particularly well suited for quantization and interpolation. Theoretically, this can be motivated by the fact that the sensitivity matrix relating the LSF-domain squared quantization error to the perceptually relevant log spectrum is diagonal [23].

Classification

In order to classify and recognize the eight speakers, an MLP (multilayer perceptron) type of neural network was used. Since neural networks are very good at mapping inputs to target outputs, this feature was used to the advantage of this study. The MLP was used to map the input to the output and it is described below.

Multilayer Perceptron (MLP)

Multilayer perceptron (MLP) is one of many different types of existing neural networks. It comprises a number of neurons connected together to form a network. This network has three layers which are input layer, one or more hidden layer(s) and an output layer with each layer containing multiple neurons [24]. A neural network is able to classify the different aspects of the behaveours, knows what is going on at the instant, diagnoses whether it is correct or faulty, forecasts what it will do next, and if required responds to what it will do next. For an MLP network with b input nodes, one-hidden-layer of c neurons, and d output neurons, the output of the network is given by equation(7) [25-26]:

$$Y_k = \phi_k\left(\sum_{j=1}^{c} w_{jk}\phi_j\left(\sum_{i=1}^{b} w_{ij}x_i\right)\right) \qquad (7)$$

where ϕ_j and ϕ_k are the activation functions of the hidden-layer neurons and the output neurons, respectively; $w_{ij}$and w_{jk} are the weights connected to the output neurons and to the hidden-layer neurons, respectively; x_i is the input.

All nodes in one layer are connected with a specific weight to every node in the following layer, without interconnections within a layer. Learning takes place in the perceptron by varying connection weights after each piece of data is processed, based on the quantity of error in the output judged against the anticipated result. This is an example of supervised learning and is achieved through back propagation, a generalization of the least mean squares algorithm [27]. However, a common problem when using MLP is the way to choose the number of neurons in the hidden layer [28].

TABLE 1
SUMMARY OF SAMPLES USED FOR THE ASR

Sample	Stutterer type			
	B	R	BL	I
F1			x	x
F2	x	x	x	x
F3	x	x	x	
F4	x	x	x	
M1	x		x	x
M2	x		x	x
M3	x		x	x
M4	x	x		x

Performance Analysis

Performance analysis is the process of evaluating how the designed system is or would be functioning. By evaluating the system, it is possible to determine if something could be done to speed up a task, or change the amount of memory required to run the task without negatively impacting the overall function of the system. Performance analysis also helps to adjust components in a manner that helps the design make the best use of available resources. The confusion matrix labelling for the computation of the ROC.

The major metrics that are extracted from the confusion matrix are sensitivity, accuracy, specificity, precision, and misclassification rate [29]. Sensitivity (Sen) or recall is a measure of the proportion of actual positives which were correctly identified (true positive rate), accuracy (Acc) is a measure of the degree of closeness of the predicted values to the actual values, precision (Pres) is a measure of repeatability or reproducibility and misclassification rate (MR) is the number of incorrectly identified instances divided by the total number of instances.

3. Results and Analysis

The stuttered speech samples that were obtained for use in this research is the University College London Archive of Stuttered Speech (UCLASS) release 1 database. The recordings of the stuttered speech were collected at University College London (UCL) over a number of years. The recordings are mostly from children who were referred to clinics in London for assessment of stuttering. The Release One recordings have only monolog speech with an age range from 5 years 4 months to 47 years. For the convenience of users, they were prepared in CHILDES, PRAAT, and SFS formats, all of which are freeware available on the Internet. The speech recordings included both male and female speakers. Table 1 shows the eight samples used and the types of stuttering present in them. The categories of the stuttering present are burst stuttering (B), reciprocating stuttering (R), blocking stuttering (BL), and interjection (I).

TABLE 2
MODIFIED ANALYSIS TOOL

Range		Assigned class
0 - <= 0.10	(0 - <= 10%)	negligible (N)
> 0.10 - <= 0.20	(> 10 - <= 20%)	poor (P)
> 0.20 - <= 0.40	(> 20 - <= 40%)	low (L)
> 0.40 - <= 0.60	(> 40 - <= 60%)	moderate (M)
> 0.60 - <= 0.80	(> 60 - <= 80%)	substantial (S)
> 0.80 - <= 0.90	(> 80 - <= 90%)	Considerable (C)
> 0.90 - <= 1.00	(> 90 - <= 100%)	high (H)

The dysfluencies associated with stuttering can be classed into the following categories [2, 8, 9, 11, 30]:

Bursts stuttering (B)
A syllable is repeated when speaking ("He wa-wa-was a good king") or (caaaaaaaaaaaaaaaaake).

Reciprocating stuttering (R)
Some syllables are repeated when speaking ("He wwwas a good king") or ("u-um-um") or prolonged ("uuuum") or repeated syllable before pronunciation ("wa wa wa water").

Blocking stuttering (BL)
A word is difficult to pronounce in a sentence, for a few seconds unsuccessful ("He w——as a good king").

Interjection (I)
Some interjections are added to the sentence ("I have um, um, a test to-day") or ("School is, well, fine") or ("The test was, you know, hard").

The analysis tool for evaluating the automatic speaker recognition (ASR) systems was a modification of the analysis tool developed by Best in 1981. In order to cater for more distinction at the boundaries of the analysis tool, it was modified to enhance its ability to effectively handle probability values that are exactly on the edges such as 20, 40, 60 and 80. Furthermore, the categories *negligible* and *high* were divided into 2 each. This was done in order to enhance the grouping of probabilities into the two classes and to reduce the band of the two classes. The modifications introduced are described in Table 2.

Objective Measure

The cepstral distance (CD), mean square error (MSE), Itakura-Saito distance (IS) and likelihood ratio (LR) measure between the original speech and the reconstructed speech can be seen in Table

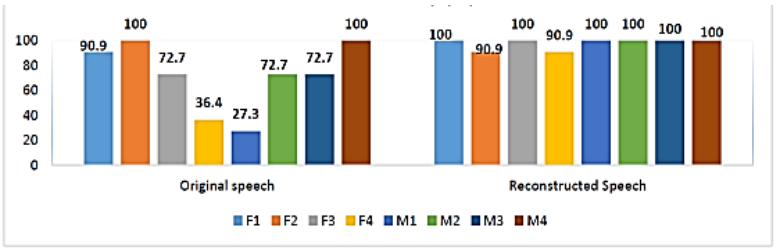

Figure 1. Sensitivity of the ASR.

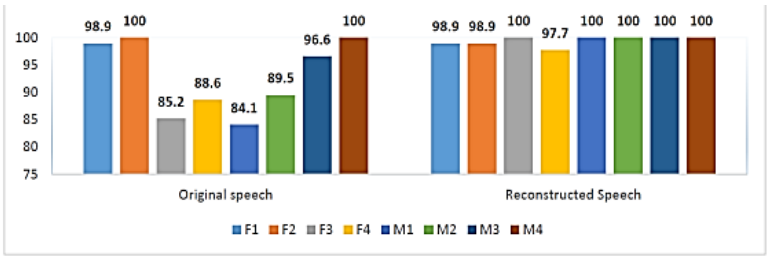

Figure 2. Accuracy of the ASR.

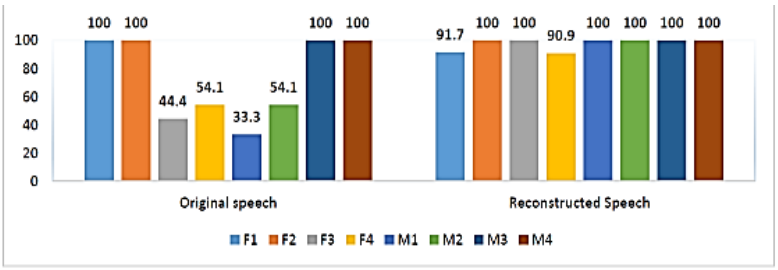

Figure 3. Precision of the ASR

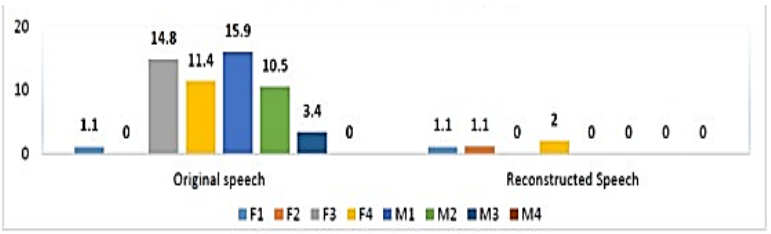

Figure 4. Misclassification Rate of the ASR.

3. The MSE between the original speech and the reconstructed speech for all the speech samples is zero, implying that the reconstruction was perfect with excellent quality of speech and a mirror reflection of the original speech. Similarly, the IS between the original and the reconstructed speech signals was zero. Since the MSE for all samples implied excellent reconstruction, it could be inferred that IS value of zero means perfect reconstruction quality.

The LR for all the 8 samples is also zero. The CD for samples F2, F4, M1 and M4 are not zero, implying that these four samples do not have perfect reconstruction. This, however, is in contrast to the result interpretation of the other 3 metrics. And because it is known that whenever signal processing techniques are applied to any signal, reversing the process cannot give exactly the same signal as the original signal. Either there is an improvement of the signal or it is degraded.

ASR Evaluation

The LSF-MLP, feature extractor classifier was selected for developing the ASR. The developed ASR was applied on the original stuttered speech signal; this is to serve as a benchmark for the purpose of comparison with the reconstructed spe-

TABLE 3
OBJECTIVE MEASURES OF THE RECONSTRUCTED SPEECH

Sample	CD	MSE	IS	LR
F1	0	0	0	0
F2	0.0291	0	0	0
F3	0	0	0	0
F4	0.0305	0	0	0
M1	0.0499	0	0	0
M2	0	0	0	0
M3	0	0	0	0
M4	0.0882	0	0	0

TABLE 4
SUMMARY OF THE PERFORMANCE ANALYSIS OF LSF-MLP FOR THE ORIGINAL SPEECH

Sample	Sen	Acc	Prec	MR
F1	H	H	H	N
F2	H	H	H	N
F3	S	C	M	P
F4	L	C	M	P
M1	L	C	L	P
M2	S	C	M	P
M3	S	H	H	N
M4	H	H	H	N

ech. The performance metrics that have been discussed above were applied to evaluate the responsiveness of the ASR to the original speech. The performance metrics of systems are plotted in Figures 1-4.

The reconstructed speech was also used with the developed ASR system. Though the reconstructted speech has been evaluated using some objective measure, the methodologies used are just distance measures and the MSE. These distance measures only evaluate the closeness between the original speech and the reconstructed speech. And these measures were not able to effectively differentiate between the original speech and the reconstructed speech. As a result, using ASR for proper evaluation of how the speech would be recognized is compulsory. The results of the performance of the ASR are as discussed below.

For the original speech, the sensitivity of the ASR to samples F1, F2 and M4 was high, while the system sensitivity to F3, M2 and M3 was substantial and F4 and M1 had low sensitivity. The accuracy was high for F1, F2, M3 and M4 and considerable for the remaining four samples. The precision of the system to M1 was low, F3, F4, and M2 were moderate and high for the other samples. The misclassification rate was poor for F3, F4, M1 and M2 and negligible for the other samples.

For the reconstructed speech, the ASR had a sensitivity of group high for all the samples, with only F2 and F4 that below 100%. Similarly, all the accuracies can as well be put in the group high,

TABLE 5
SUMMARY OF THE PERFORMANCE ANALYSIS OF LSF-MLP FOR THE RECONSTRUCTED SPEECH

Sample	Sen	Acc	Prec	MR
F1	H	H	H	N
F2	H	H	H	N
F3	H	H	H	N
F4	H	H	H	N
M1	H	H	H	N
M2	H	H	H	N
M3	H	H	H	N
M4	H	H	H	N

with F1, F2 and F4 slightly below 100%. The values of the precision are also in the group high, with F1 and F4 below 100%. In addition, all the misclassification rates are in the category negligible, while only F1, F2, and F4 have values slightly more than 0%.

Tables 4 and 5 show the summary that reduces the calculations and explanations of Figures 1-4. The hyperbolic tangent sigmoid activetion function (tansig) was used for both the hidden layer and the output layer. From Table 5, it can be seen that the ASR excellently senses each input, puts them in their correct classes, with no misfiring. Similarly, all the inputs had very small values for the misclassification rates, implying that almost all the samples were correctly classified. Comparing it with the ASR results of the original speech signals, it would be observed that only the *accuracy* was very good. The results for the *sensitivity* and *precision* had a mixture of the different categories. Also, the *misclassification rates* are all negligible, with values not as low as they should be. Only 4 of the 8 samples had below 5% while the other 4 had values more than 10%.

4. Conclusion

The use of speech reconstruction for stuttered speech for correcting stuttered speech has been enumerated in this study. Since the LPC algorithm used models the human speech production, the reconstructed speech was very similar to the original speech as interpreted by the objective measures. The ASR gave a better picture of the reconstructed speech as all the speech samples were perfectly recognized while only 3 samples of the original speech were perfectly recognized. Therefore, it could be concluded that the reconstructed speech would be better perceived by the stutterers.

References

[1] M. Hariharan, V. Vijean, Y. Chong, and Y. Sazali, “Speech stuttering assessment using sample entropy and Least Square Support Vector Machine,” in 8th International Collo-

quium on Signal Processing and its Applications (CSPA), 2012, pp. 240–245.

[2] G. Manjula and M. Kumar, "Stuttered Speech Recognition for Robotic Control," Int. J. Eng. Innov. Technol., vol. 3, no. 12, pp. 174–177, 2014.

[3] J. Duffy, "Motor speech disorders: clues to neurologic diagnosis," in Parkinson's Disease and Movement Disorders, Humana Press, 2000, pp. 35–53.

[4] E. G. Conture and J. S. Yaruss, "Treatment Efficacy Summary," Am. speech- language Hear. Assoc., no. 1993, p. 20850, 2002.

[5] C. Oliveira, D. Cunha, and A. Santos, "Risk factors for stuttering in disfluent children with familial recurrence," Audiol. Res., vol. 18, no. 1, pp. 43–49, 2013.

[6] L. S. Chee, O. C. Ai, M. Hariaran, and S. Yaacob, "MFCC based recognition of repetitions and prolongations in stuttered speech using k-NN and LDA," in 2009 IEEE Student Conference on Research and Development and Development (SCOReD), 2009, pp. 146–149.

[7] M. Hariharan, L. S. Chee, and S. Yaacob, "Analysis of infant cry through weighted linear prediction cepstral coefficients and Probabilistic Neural Network," J. Med. Syst., vol. 36, no. 3, pp. 1309–15, Jun. 2012.

[8] J. Zhang, B. Dong, and Y. Yan, "A Computer-Assist Algorithm to Detect Repetitive Stuttering Automatically," in 2013 International Conference on Asian Language Processing (IALP), 2013, pp. 249–252.

[9] S. Awad, "The application of digital speech processing to stuttering therapy," in IEEE Sensing, Processing, Networking, Instrumentation and Measurement Technology Conference, IMTC 97, 1997, pp. 1361–1367.

[10] L. S. Chee, O. C. Ai, M. Hariharan, and S. Yaacob, "Automatic detection of prolongations and repetitions using LPCC," in International Conference for Technical Postgraduates 2009, TECHPOS 2009, 2009, pp. 1–4.

[11] K. Hollingshead and P. Heeman, "Using a uniform-weight grammar to model disfluencies in stuttered read speech: a pilot study," Oregon, 2004.

[12] J. Pálfy and J. Pospichal, "Pattern search in dysfluent speech," in 2012 IEEE Inter-national Workshop on Machine Learning for Signal Processing (MLSP), 2012, pp. 1–6.

[13] I. Mansour and S. Al-Abed, "A New Architecture Model for Multi Pulse Linear Predictive Coder for Low-Bit-Rate Speech Coding," Dirasat Eng. Sci., vol. 33, no. 2, 2010.

[14] M. Suman, "Enhancement of Compressed Noisy Speech Signal," Koneru Lakshmaiah Education Foundation, 2014.

[15] D. Jones, S. Appadwedula, M. Berry, M. Haun, J. Janovetz, M. Kramer, D. Moussa, D. Sachs, and B. Wade, "Speech Processing: Theory of LPC Analysis and Synthesis," Connexions. June, 2009.

[16] L. Rabiner and R. Schafer, Digital processing of speech signals. Prentice-Hall, 1978.

[17] Q. P. Li, Speaker Authentication. Springer-Verlag Berlin Heidelberg, 2012.

[18] R. L. Venkateswarlu and R. Vasanthakumari, "Neuro Based Approach for Speech Recognition by using Mel-Frequency Cepstral Coefficients," Int. J. Comput. Sci. Commun., vol. 2, no. 1, pp. 53–57, 2011.

[19] C. Cornaz, U. Hunkeler, and V. Velisavljevic, "An automatic speaker recognition system," Lausanne, Switzerland, 2003.

[20] G. T. Tsenov and V. M. Mladenov, "Speech recognition using neural networks," in Neural Network Applications in Electrical Engineering (NEUREL), 2010 10th Sym-posium on, 2010, pp. 181–186.

[21] V. Namburu, "Speech Coder Using Line Spectral Frequencies of Cascaded Second Order Predictors," VirginiaTech, 2001.

[22] P. Kabal and R. Ramachandran, "The computation of line spectral frequencies using Chebyshev polynomials," IEEE Trans. Acoust. Speech Signal Process., vol. 34, no. 6, pp. 1419–1426, 1986.

[23] W. B. Kleijn, T. Bäckström, and P. Alku, "On line spectral frequencies," IEEE Signal Process. Lett., vol. 10, no. 3, pp. 75–77, 2003.

[24] R. Kumar, R. Ranjan, S. K. Singh, R. Kala, A. Shukla, and R. Tiwari, "Multilingual Speaker Recognition Using Neural Net-work," in Proceedings of the Frontiers of Research on Speech and Music, FRSM, 2009, pp. 1–8.

[25] M. A. Al-Alaoui, L. Al-Kanj, J. Azar, and E. Yaacoub, "Speech recognition using artificial neural networks and hidden Markov models," IEEE Multidiscip. Eng. Educ. Mag., vol. 3, no. 3, pp. 77–86, 2008.

[26] S. S. Haykin, Neural networks and learning machines, vol. 3. Pearson Education Upper Saddle River, 2009.

[27] S. Agrawal, A. K. Shruti, and C. R. Krishna, "Prosodic feature based text dependent speaker recognition using machine learning algorithms," Int. J. Eng. Sci. Technol., vol. 2, no. 10, pp. 5150–5157, 2010.

[28] C. L. Tan and A. Jantan, "Digit recognition using neural networks," Malaysian J. Comput. Sci., vol. 17, no. 2, pp. 40–54, 2004.

[29] R. Kohavi and F. Provost, "Glossary of terms," Mach. Learn., vol. 30, no. 2–3, pp.

271–274, 1998.

[30] M. Hariharan, L. S. Chee, O. C. Ai, and S. Yaacob, "Classification of speech dysfluencies using LPC based parameterization techniques," J. Med. Syst., vol. 36, no. 3, pp. 1821–1830, 2012.