RECEIVER OPERATING CHARACTERISTICS MEASURE FOR THE RECOGNITION OF STUTTERING DYSFLUENCIES USING LINE SPECTRAL FREQUENCIES

NAHRUL KHAIR BIN ALANG MD RASHID, SABUR AJIBOLA ALIM*, NIK NUR WAHIDAH NIK HASHIM AND WAHJU SEDIONO

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

**Corresponding author: moaj1st@yahoo.com*

(Received: 23rd Aug. 2015; Accepted: 27th Sep. 2016; Published on-line: 30th May 2017)

ABSTRACT: Stuttering is a motor-speech disorder that has features in common with other motor control disorders such as dystonia, Parkinson's disease, and Tourette's syndrome. Stuttering results from complex interactions between factors such as motor, language, emotions, and genetic systems. This study used Line Spectral Frequency (LSF) for feature extraction, while using three classifiers for the identification purpose, Multilayer Perceptron (MLP), Recurrent Neural Network (RNN) and Radial Basis Function (RBF). The UCLASS (University College London Archive of Stuttered Speech) release 1 was used as the database in this research. These recordings were from people of ages ranging from 12y11m to 19y5m, who were referred to clinics in London for assessment of their stuttering. The performance metrics used for interpreting the results are sensitivity, accuracy, precision, and misclassification rate. Only M1 and M2 had below 100% sensitivity for RBF. The sensitivity of M1 was found to be between 40% & 60%, therefore categorized as moderate, while that of M2 falls between 60% & 80%, classed as substantial. Overall, RBF outperforms the two other classifiers, MLP and RNN for all the performance metrics considered.

ABSTRAK: Gagap adalah gangguan motor pertuturan, mempunyai ciri-ciri yang sama dengan lain-lain gangguan kawalan motor seperti dystonia, penyakit Parkinson dan sindrom Tourette. Keputusan kegagapan daripada interaksi kompleks antara faktor-faktor seperti motor, bahasa, emosi dan genetik. Kajian ini menggunakan Frekuensi Line spektral (LSF) untuk pengekstrakan ciri, semasa menggunakan tiga penjodoh untuk tujuan mengenal pasti, Multilayer Perceptron (MLP), Rangkaian Neural Berulang (RNN) dan Radial Asas Fungsi (RBF). The UCLASS (University College London Arkib Stuttered Ucapan) melepaskan 1 digunakan sebagai pangkalan data dalam kajian ini. Ini rakaman adalah dari orang-orang peringkat umur 12y11m untuk 19y5m, yang dirujuk kepada klinik di London untuk penilaian kegagapan mereka. Metrik prestasi yang digunakan untuk mentafsir keputusan yang sensitif, ketepatan, ketepatan dan kadar misclassification. Hanya M1 dan M2 mempunyai di bawah 100% kepekaan untuk RBF. Kepekaan M1 didapati antara 40% & 60%, oleh itu dikategorikan sebagai sederhana, manakala M2 jatuh antara 60% & 80%, dikelaskan sebagai besar. Secara keseluruhan, RBF melebihi performa dua penjodoh lain, MLP dan RNN untuk semua metrik prestasi dipertimbangkan.

KEYWORDS: *stuttering; dysfluencies; line spectral frequency; radial basis function*

1. INTRODUCTION

Stuttering is a neurological trait that may involve specific abnormalities of speech motor control in the human brain. Stutterers experience reduced blood flow and decreased or increased electrical activity in areas of the brain associated with speech production. The cause of this malfunction is linked to the brain [1, 2]. Stuttering is a severe complication focused on by speech pathologists [3, 4]. Stuttering is primarily viewed as a motor-speech disorder, sharing features with other kinds of motor control disorders such as dystonia, Parkinson's disease, and Tourette's syndrome. However, there is evidence that linguistic factors also play a role in stuttering [5].

It was found that many complex interactions between the language and motor systems, have led researchers to believe that there is no single cause of stuttering. Stuttering is the result of a complex interaction among numerous factors such as motor, language, emotional, and genetic systems. About 80% of stuttering occurrences automatically disappear. However, the remaining 20%, which is about 1% of the entire world's population, have difficulties returning to normal speech [6]. Stuttering has been found to more prevalent in males than females (by a ratio of 4:1) [1, 3, 4, 7].

Stutterers and non-stutterers alike have speech disfluencies that can be gaffes or disturbances in the flow of words a speaker plans to say, but disfluencies are more observable in stutterers' speech [8]. Stuttered speech is rich in dysfluencies, usually repetitions. Classical approaches to the analysis of dysfluencies are done over very short intervals, which is sufficient for recognition of simple repetitions of phonemes [9].

The usual practice in the stuttered speech recognition or dysfluency recognition reported in previous research findings was in form of identification accuracy. It is, however, essential to use metrics such as sensitivity, accuracy, precision and misclassification rate, which better analyze how a stuttered speech recognition system would behave in real life. These performance metrics form a part of the receiver operating characteristics (ROC). This experiment serves as a furtherance of the previous experiments, gives further confidence in stuttering recognition research, and serves as a precursor to research on stuttered speech correction via reconstruction.

2. FEATURE EXTRACTION

Line Spectral Frequency (LSF) is an alternative to the direct form predictor coefficients or the lattice form reflection coefficients for representing filter response. The direct form coefficient representation of the linear prediction coefficient (LPC) filters is not perfect for efficient quantization. Nonlinear functions of the reflection coefficients are often used as transmission parameters in place of direct form coefficients. These parameters are preferable because they have a relatively low spectral sensitivity [10]. 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 [11].

The linear prediction polynomial $A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k}$ can be expressed as $A(z) = 0.5[P(z) + Q(z)]$ [12].

where $P(z) = A(z) + z^{-(p+1)} A(z^{-1})$, $Q(z) = A(z) - z^{-(p+1)} A(z^{-1})$, $P(z)$ is the vocal tract with the glottis closed, $Q(z)$ is the vocal tract with the glottis opened, $A(z)$ is the LPC analysis filter of order p, and a_k is the LPC coefficient.

The roots of P and Q lie on the unit circle in the complex plane. The roots alternate as they travel through the circle. The coefficients of P and Q are real. As such, the roots occur in conjugate pairs and are the same as the LPC order.

3. CLASSIFICATION

3.1 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 generally has three layers: the input layer, one or more hidden layer(s), and the output layer, with each layer containing multiple neurons [13]. A neural network is able to classify the different aspects of the behaviors of a data structure, understand what is going on at every instant of time, analyze whether it is correct or faulty, forecast what to do next, and give the desired response [14, 15].

3.2 Recurrent Neural Network (RNN)

Feed-forward multi-layer neural networks are unable to deal with time-varying information like time-varying spectra of speech sound. One way to cope with this problem is to incorporate feedback into the networks to provide them with the capacity to learn incoming time-varying information [16]. Recurrent Neural Networks (RNN) have a feedback connection that is used to pass the output of a neuron in a certain layer to the previous layer(s). RNN have the ability to process short-term spectral features but yet respond to long-term temporal events [17]. However, RNN have limited memory, and underperform once the memory is surpassed [15].

3.3. Radial Basis Functions (RBF)

Radial Basis Function (RBF) Networks are derived from the theory of function approximation. They are two-layered feed-forward networks, containing a hidden layer and an output layer. The hidden layer nodes implement a set of radial basis activation functions, while the output layer nodes compute the linear summation functions just as in MLP. The network training is divided into two phases: first the weights from the input to hidden layer are determined, then the weights from the hidden to output layer. The learning or training is very fast and the networks are very good at interpolation [18]. The response of the functions rises or falls monotonically with distance from their center and are controlled by local measurements [19].

4. METHODOLOGY

The UCLASS (University College London Archive of Stuttered Speech) release 1 was used as the database for this research. These recordings were from people of ages 12y11m to 19y5m, who were referred to clinics in London for assessment of stuttering. The recordings were all monologs [20]. All the samples were quantized at a bit rate of 16 bits. Table 1 shows the age and sex of the 8 samples used for this experiment. Each sample was divided into smaller bits of 10 seconds and 11 sub-samples for each sample.

The relevant features were extracted from each sample using Line Spectral Frequency (LSF). A three layer multilayer perceptron, a three layer recurrent neural network and a two layer radial basis function were used for the classification and identification. The neural networks were trained with 88 inputs, 8 outputs, and 215 neurons in the hidden layer. All the layers were trained with a scaled conjugate gradient type of back propagation algorithm, while the tangent sigmoid activation function was used for the

computation of the weight in each layer. The RBF, however, uses the radial basis activation function for the hidden layer and linear activation function for the output layer.

Table 1: Features of samples used

	Age	Sex	
1	15y2m	F	F1
2	17y2m	F	F2
3	15y11m	F	F3
4	12y11m	F	F4
5	16y4m	M	M1
6	17y9m	M	M2
7	19y5m	M	M3
8	16y9m	M	M4
mean	16y5m		

The confusion matrices of the designed systems were plotted. From the confusion matrix plot, the sensitivity, accuracy, precision, and the misclassification rate were computed as performance measures that are the receiver operating characteristics (ROC). Sensitivity measures the number of correctly identified samples, accuracy measures the degree of closeness between the predicted and actual values, precision measures the ability of the algorithm to reproduce the same output for the same set of inputs, and the misclassification rate measures the percentage of incorrectly identified samples with respect to the total number of samples.

5. RESULTS AND DISCUSSION

The results of the performance measures were used to evaluate the three systems, LSF-MLP, LSF-RNN and LSF-RBF. The results are given as percentages for sensitivity, accuracy, precision, and misclassification rate. They help to give a better understanding of the systems for further evaluation, recommendations, and implementation. The three classifiers used have different strengths and weaknesses which would in turn affect their eventual performance. For the interpretation of the results, the tool used is the statistical classification by Best in 1981. It was used to describe the significance of the probability of any experiment. It is listed as follows: 0 - 0.20 (0 - 20%) – negligible, 0.20 - 0.40 (20 - 40%) – low, 0.40 - 0.60 (40 - 60%) – moderate, 0.60 - 0.80 (60 - 80%) – substantial & 0.80 - 1.00 (80 - 100%) – high.

5.1 Sensitivity

Table 2 shows the sensitivity of the systems to each of the samples used. Only M1 and M2 had below 100% sensitivity for RBF. The sensitivity of M1 was found to be moderate, while that of M2 was substantial. All the other samples were high. Similarly, in the case of RNN, the sensitivities of F1, F2, M2, M3 and M4 were high, F3 and F4 were moderate, while M1's sensitivity was negligible. Furthermore, for MLP, the sensitivities of F1, F2 and M4 were high. While F3, M2 and M3, all had substantial sensitivity. M1's sensitivity was moderate. It was observed that for the three classifiers considered, the sensitivity of sample M1 was the lowest.

Table 2: Sensitivity

	MLP (%)	RNN (%)	RBF (%)
F1	90.91	90.91	100
F2	100	90.91	100
F3	72.73	54.55	100
F4	36.36	45.45	100
M1	27.27	9.09	45.45
M2	72.73	90.91	72.73
M3	72.73	100	100
M4	100	100	100

5.2 Accuracy

Table 3 shows the accuracies of the three systems under consideration. The accuracies for all the samples and classifiers are high. In the case of RNN, all the accuracies were 100% except for M1 and M2 whose accuracies were below 90% and 89.77% in both cases. Similarly, for MLP, only F2 and M4 had accuracies that are exactly 100%. The accuracies of F1 and M3 fall between 90 and 100%, while the accuracies of the other samples are between 80 and 90%. Furthermore, for RNN, the accuracies of only M3 and M4 are exactly 100%, the accuracies of F1 and F2 are between 90 and 100%, while the accuracies of the other fall between 80 and 90%. It could also be observed that M1 had the lowest accuracies for both MLP and RBF, while F3 had the lowest accuracy for RNN.

Table 3: Accuracy

	MLP (%)	RNN (%)	RBF (%)
F1	98.86	98.86	100
F2	100	97.73	100
F3	85.23	85.23	100
F4	88.64	87.5	100
M1	84.09	87.5	89.77
M2	89.53	88.64	89.77
M3	96.6	100	100
M4	100	100	100

Table 4: Precision

	MLP (%)	RNN (%)	RBF (%)
F1	100	100	100
F2	100	90.91	100
F3	44.44	42.86	100
F4	57.14	50	100
M1	33.33	50	62.5
M2	57.14	52.63	57.14
M3	100	100	100
M4	100	100	100

5.3 Precision

The precision values of the systems can be seen in Table 4. RNN had high precision for F1, F2, M3 and M4, and is categorized as high. The precisions of F3, F4, M1 and M2 are moderate. Furthermore, for RBF, the values of the precision for all the samples are high, except for M1 and M2. The precision of M1 is substantial, while that of M2 is moderate. In addition, in the case of MLP, the accuracies of F1, F2, M3 and M4 are high. While F3, F4 and M2 have moderate accuracies, and M1's accuracy is low.

5.4 Misclassification Rate

The misclassification rates of LSF-MLP, LSF-RNN and LSF-RBF can be seen in Table 5. It can be seen that all the values obtained are below 20%, putting them in the class of negligible. RNN had misclassification rates of 0% each for M3 and M4, while F1 and F2 had below 10% misclassification rate and the others were between 10 and 20%. Similarly, MLP, both F2 and M4 had misclassification rates of 0%, while F1 and M3 have between 0 and 10% misclassification rates and the remaining fall between 10 and 20% misclassification rate. Finally for RBF, except for M1 and M2, all the other samples had misclassification rates of 0%, while M1 and M2 had misclassification rates of 9.57% and 10.23% respectively.

Table 5: Misclassification Rate

	MLP (%)	**RNN (%)**	**RBF (%)**
F1	1.14	1.14	0
F2	0	2.27	0
F3	14.77	12.79	0
F4	11.36	12.5	0
M1	15.91	12.5	9.57
M2	10.47	11.36	10.23
M3	3.41	0	0
M4	0	0	0

5.5 Discussion

It is expected that for a recognition system to work perfectly, the sensitivity, accuracy and precision should be high, while the misclassification rate should be negligible for each sample. LSF-RBF has high sensitivity to six out of the eight samples having high sensitivity, accuracy and precision, with negligible misclassification rates. RBF has the ability to generate as many as needed hidden neurons in order to effectively map the input to the output. This gives it better performance as compared with the MLP and RNN, where you have to specify a certain number of hidden neurons. As compared with the research conducted by [21], where 39 stuttered speech samples were used, 2 females (5%) and 37 males (95%), from UCLASS which is the same database used for this research.

Table 6: Benchmark

Feature extractor	**Number of coefficients**	**Frame length (ms)**	**Classifier**	**Percentage recognition (%)**
MFCC	15	40	SVM	96.14
WLPCC	10	50	SVM	96.02

They used five feature extractors; LPC, Linear Prediction Cepstral coefficient (LPCC), Weighted Linear Prediction Cepstral Coefficient (WLPCC), Perceptual Linear Prediction (PLP) & Mel Frequency Cepstral Coefficient (MFCC) and k-Nearest Neighbour (kNN), Linear Discriminant Analysis (LDA) & Support Vector Machines (SVM) as classifiers. LPC, LPCC, WLPCC and PLP are LPC based feature extractors as seen in table 6. For MFCC, the best recognition of 96.14% was obtained with SVM classifier, 15 MFCC coefficients and 40 ms frame length. The LPC-based feature extractors, WLPCC-SVM had the best recognition of 96.02% at the 10th LPC order and 50 ms frame length. However, the difference between the current research and the study in [21] is that while the authors evaluated the system using overall system performance. This research was evaluated using the performance and behavior of the system to each test sample.

5. CONCLUSION

The findings of this experiment have been presented above and it was observed that the LSF-RBF performed well in terms of sensitivity, precision, accuracy and misclassification rate for all the samples except M1 and M2. RBF, on the overall, outperforms the two other classifiers, MLP and RNN, for all the performance metrics considered. MLP and RNN performed well for F1, F2, M3 and M4, while giving an average performance for the other samples. It can therefore be said that the LSF-RBF system has a bright future in terms of possible use in real life applications.

REFERENCES

[1] Awad S. (1997) The application of digital speech processing to stuttering therapy. In IEEE Sensing, Processing, Networking, Instrumentation and Measurement Technology Conference (IMTC), pp1361-1367.

[2] Chee LS, Ai OC, Yaacob S. (2009) Overview of automatic stuttering recognition system. In Proceedings of International Conference on Man-Machine Systems, pp1-6.

[3] Chee LS, Ai OC, Hariharan M, Yaacob S. (2009) Automatic detection of prolongations and repetitions using LPCC. In International Conference for Technical Postgraduates (TECHPOS), pp1-4.

[4] Chee LS, Ai OC, Hariaran M, Yaacob S. (2009) MFCC based recognition of repetitions and prolongations in stuttered speech using k-NN and LDA. In IEEE Student Conference on Research and Development and Development (SCOReD), pp146-149.

[5] Watkins K, Smith S, Davis S, Howell P. (2008) Structural and functional abnormalities of the motor system in developmental stuttering. Brain, 131(1):50-59.

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

[7] Manjula G, Kumar M. (2014) Stuttered Speech Recognition For Robotic Control. Int. J. Eng. Innov. Technol., 3(12):174-177.

[8] Hollingshead K, Heeman P. (2004) Using a uniform-weight grammar to model disfluencies in stuttered read speech: a pilot study. CSLU, OHSU. Available at www.cslu.ogi.edu/publications/ps/hollingshead04.pdf. [Accessed: 20-Jun-2016].

[9] Pálfy J, Pospichal J. (2012) Pattern search in dysfluent speech. In IEEE International Workshop on Machine Learning for Signal Processing (MLSP), pp 1-6.

[10] Kabal P, Ramachandran R. (1986) The computation of line spectral frequencies using Chebyshev polynomials. IEEE Trans. Acoust. Speech Signal Process., 34(6):1419-1426.

[11] Kleijn WB, Bäckström T, Alku P. (2003) On line spectral frequencies. IEEE Signal Process. Lett., 10(3):75-77.

[12] Robinson T. (1998) Speech Analysis. [Online]. Available at http://svr-www.eng.cam.ac.uk/~ajr/SpeechAnalysis/SpeechAnalysis.html. [Accessed: 20-Jun-2016].

[13] Kumar R, Ranjan R, Singh SK, Kala R, Shukla A, and Tiwari R. (2009) Multilingual Speaker Recognition Using Neural Network. In Proceedings of the Frontiers of Research on Speech and Music, FRSM, pp1-8.
[14] Al-Alaoui MA, Al-Kanj L, Azar J, Yaacoub E. (2008) Speech recognition using artificial neural networks and hidden Markov models. IEEE Multidiscip. Eng. Educ. Mag., 3(3):77-86.
[15] Haykin SS. (2009) Neural networks and learning machines. Pearson Education Upper Saddle River.
[16] Ahmad AM, Ismail S, Samaon DF. (2004) Recurrent neural network with backpropagation through time for speech recognition. In IEEE International Symposium on Communications and Information Technology (ISCIT), 1:98-102.
[17] Ismail S, Ahmad A. (2004) Recurrent neural network with backpropagation through time algorithm for arabic recognition. In Proceedings of the 18th European Simulation Multiconference (ESM), pp 13-16.
[18] Bullinaria J. (2004) Introduction to neural networks. University of Birmingham, UK.
[19] Singla P, Subbarao K, Junkins J. (2007) Direction-dependent learning approach for radial basis function networks. IEEE Trans. Neural Networks, 18(1):203-222.
[20] Howell P, Davis S, Bartrip J. (2009) The University College London Archive of Stuttered Speech (UCLASS). J. Speech, Lang. Hear. Res., 52(2):556-569.
[21] Chong YF, Hariharan M, Chee LS, Yaacob S, Adom H. (2013) Comparison of speech parameterization techniques for the classification of speech disfluencies. Turkish J. Electr. Eng. Comput. Sci., 21:1983-1994.