SPEAKER IDENTIFICATION USING HYBRID OF SUBTRACTIVE CLUSTERING AND RADIAL BASIS FUNCTION

YAP TECK ANN

UNIVERSITI TEKNOLOGI MALAYSIA

SPEAKER IDENTIFICATION USING SUBTRACTIVE CLUSTERING AND RADIAL BASIS FUNCTION

YAP TECK ANN

A thesis submitted in fulfillment of the
requirements for the award of the degree of
Master of Science (Computer Science)

Faculty of Computing
UniversitiTeknologi Malaysia

JUNE 2013

ACKNOWLEDGEMENT

This thesis has accomplished with the contribution directly and indirectly of numerous people to whom I wish to express my sincere gratitude.

First of all, I would like to thank to my family who have been supporting and encouraging me to complete this thesis. Their encouragement is my greatest motivation to continue this research.

A special thank to my supervisor, Assoc. Prof. Dr. MohdShafry Bin Mohd Rahim, for all his guidance, encouragement and patience in supervised my studies. Thanks so much for his advices, comments and suggestions to me during the research. This thesis is a result of his vision and ideas.

I also thank Assoc. Prof. Abdul Manan bin Ahmad previously as my supervisor as well for his advices on how to improve this research and provided a working environment with great facility.

Furthermore, I would like to take this opportunity to thank all my friends and lab-mates who give me encouragement and lend me a helpful hand when I am in troubles.

ABSTRACT

Speaker identification is the computing task of identifying unknown identities based on voice. A good speaker identification system must have a high accuracy rate to prevent incorrect detection of the user's identity. This research proposed a hybrid of Subtractive Clustering and Radial Basis Function (Sub-RBF) which is the combination of supervised and unsupervised learning. Unsupervised learning is more suitable for learning large and complex models than supervised learning. This is because supervised learning increasing the number of connections between sets in the network. If the model contains a large and complex dataset, supervised learning is difficult. In addition, K-means is faced with improper initial guessing of first cluster centre and difficulty in determining the number of cluster centres. The proposed technique is introduced because subtractive clustering is able to solve these problems. RBF is a simple network structures with fast learning algorithm. RBF neural network model with subtractive clustering proposed to select hidden node centers, can achieve faster training speed. In the meantime, the RBF network was trained with a regularization parameter so as to minimize the variances of the nodes in the hidden layer and perform more accurate prediction. The accuracy rate for subtractive clustering is 8.125% and 11.25% for training dataset 1 and training dataset 2 respectively. However, Sub-RBF provides 76.875% and 71.25% accuracy rate for training dataset 1 and training dataset 2 respectively. In conclusion, Sub-RBF has improved the speaker identification system accuracy rate.

ABSTRAK

Sistem pengecaman suara adalah tugas mengecam identiti manusia berasaskan suara. Sistem pengecaman suara yang baik mesti mempunyai kadar pengecaman yang tinggi untuk mengelakkan daripada salah pengesanan identiti pengguna. Kajian ini mencadangkan hibrid Kelompok Subtraktif dan Fungsi Asas Jejarian (Sub-RBF) yang merupakan gabungan pembelajaran tak diselia dan pembelajaran diselia. Pembelajaran tak diselia lebih sesuai untuk mempelajari model yang besar dan kompleks berbanding dengan pembelajaran diselia. Ini adalah kerana pembelajaran diselia meningkatkan bilangan sambungan set dalam rangkaian. Mempelajari model pembelajaran diselia adalah sukar jika model mengandungi set data yang besar dan kompleks. Selain itu, K-means menghadapi masalah tekaan awalan tentang pusat kluster pertama dan kesukaran untuk menentukan bilangan kluster. Teknik yang dicadangkan ini diperkenalkan kerana Kelompok Subtraktif berupaya menyelesaikan masalah tersebut. RBF merupakan struktur rangkaian yang ringkas dan algoritma pembelajaran yang lebih pantas. Model rangkaian neural RBF menggunakan Kelompok Subtraktif untuk memilih pusat nod tersembunyi dapat mencapai kelajuan latihan dengan lebih cepat. Pada masa yang sama, rangkaian RBF yang dilatih dengan parameter diregularisasi dapat mengurangkan varians nod pada lapisan tersembunyi dan melaksanakan ramalan yang lebih tepat. Kadar pengecaman Kelompok Subtraktif ialah 8.125% dan 11.25% bagi dataset latihan 1 dan dataset latihan 2. Namun begitu, Sub-RBF menyediakan kadar pengecaman 76.875% dan 71.25% bagi dataset latihan 1 dan dataset latihan 2. Kesimpulannya, Sub-RBF telah meningkatkan kadar pengecaman untuk sistem pengecaman suara.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	-	Artificial Neural Network
DTW	-	Dynamic Time Warping
VQ	-	Vector Quantization
HMM	-	Hidden Markov Model
GMM	-	Gaussian Mixture Mode
GD	-	Gradient Descent
RPROP	-	Resilient Back-Propagation
SVM	-	Support Vector Machine
LBG	-	Linde-Buzo-Grey
ASI	-	Automatic Speaker Identification
LPC	-	Linear Predictive Coding
MFCC	-	Mel-frequency Cepstral Coefficients
PLP	-	Perceptual Linear Predictive
DCT	-	Discrete Cosine Transform
DFT	-	Digital Fourier Transform
ML	-	Maximum Likelihood
EM	-	Expectation Maximazation
MLP	-	Multilayer Perceptron
FE	-	Feature Extractor
FCM	-	Fuzzy C-means
RBF	-	Radial Basis Function
LMS	-	Least Mean Squares
Sub-RBF	-	Hybrid Subtractive Clustering and Radial Basis Function

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introductions

Speaker recognition technologies have two major applications that are speaker identification and speaker verification. The goal of speaker identification is to recognize the unknown speaker from a set of N known speakers. On the other hand, the goal of speaker verification is to evaluate whether the claimed identity is correct or not when the unknown speaker presents a speech sample.To build a robust speaker identification system, it is often difficult because the performance of the speaker identification is dependent upon few factors such as amount of data, environment for speech producing, age of the speakers, accuracy rate of the system and the time processing. The performance of the speaker identification system must be near perfection becausethose technologies are currently applying in access control system, security control for confidential information, transaction authentication and telephone banking.

In this research, accuracy rate of the speaker identification will be focused to improve the performance of the speaker identification. The accuracy rate for speaker identification, in other word is percentage of correct identification is the main performance measurement. Once the accuracy rate for speaker identification is unsatisfactory, the other performance measurement such as time processing and amount of data will become unimportant.

Speaker identification can be divided into text dependent and text independent. For the text dependent, speaker must use the same utterance for the training and testing phase in the system. But in text independent, user can simply use whatever utterance in training and testing phase. A matter of course, this project will concentrate on text independent.

1.2 Problem Background

A robust speaker recognition system is influenced by few factors. Those factors or named as speech variation can be classified into six categories. There are intra-speaker variations, inter-speaker variations, model size, robustness, modelling and accuracy (El Hannani and Petrovska-Delacrétaz, 2006).

Intra-speaker variation is generally interpreted as variation in correctness. Every human being can use his language in more than one way. The voice could change in time due to aging, illness and emotions. These reasons may influence the result of the speaker recognition system. To solve this problem, better enrollment techniques are needed to increase the accuracy for the speaker identification.

Inter-speaker variations can be explained as each of the speakers will produce the different speech signal even they are uttered the same utterance. The most vital source of this variation is the physiological difference between speakers, such as the vocal tract length, physiology of the vocal folds, shape of the nasal tract, etc. The inter-speaker variations are also influence by the age, gender, speaking style and others related with the physiological difference.

The model size is the amount of the training data used to build the speaker model for the recognition system. Large amount of the training data is a large impact for the accuracy for the recognition system. The complexity of the training data increases proportionally to the error rate of the inter-speaker variation, memory and time.

In therobust speech variants, the production, perception and acoustic representation of a speech signal are affected by the environment in which the speech is produced. There are two categories of environment aspects that induce the variations, static elements and dynamic elements. The static elements are caused by the room acoustics, reverberation and etc. The dynamic elements are caused by the background noise, microphone placement and etc. The differences in recording devices and environments can introduce discrepancies and influence the accuracy of the system.

In addition to the speech waveform, a recorded signal may contain acoustical background noise and the effects of microphone characteristics and electrical transmission. The noise of the acoustical background and the transmission will be used to train the speaker model. Some of the speaker models capture the speaker characteristics and the noise together. This will influence the accuracy when the speaker model is used to recognition system.

The first step of the recognition system is the enrollment processes which record the speaker's voice and extract the features from the speaker’s voice. There are several ways to extract the features from the voice to build the speaker model by using statistical method. Statistical method can be divided into generative and discriminative models. Generative model are probability density estimators which model the acoustic feature vectors, discriminative model are optimized to minimize the error on a set of training samples of the target and non-target (imposters) classes. So a suitable model will increase the accuracy of the recognition system.

Pattern classification plays an important role in speaker modelling component chain. The result of the pattern classification will affect the performance of the speaker recognition system in testing phase. Dynamic Time Warping (DTW) and Hidden Markov Model (HMM) are famous pattern classification technique but due to the characteristic which not suitable for text independent recognition, those technique are starting be eliminated in speaker identification system(Loh, 2010).

In order to solve the problem of text independent recognition, some approaches have been introduced by researchers such as Vector Quantization (VQ), Gaussian mixture model (GMM), Support Vector Machine (SVM)and etc. According to Kekre and Kulkarni(2010), vector quantization is a very simple technique but the accuracy rate decrease when the number of speaker increases. VQ is a process of mapping a large set of vectors to produce a smaller set of vectors which represents the centroids or called as codewordof the distribution. Collection of all the codewordis called codebook. To form a codebook, the training data has to cluster and the original algorithm involves in producing the codebook is Linde-Buzo-Grey (LBG). LBG algorithm is one of the most popular algorithms and has an advantage of simplicity in learning. But LBG is a slow learning algorithm and this characteristic causes LBG not suitable to learn a large set of data.

According toKinnunen(2000), the clustering algorithm involved in speaker identification are Linde-Buzo-Grey (LBG), Self-organizing maps (SOM), Pairwise nearest neighbour (PNN), Principal component analysis (PCA) and Randomized local search algorithm (RLS). Each of the clustering algorithms is success in the speaker identification system. The research can prove that clustering algorithm is one of the methods in speaker identification and has a high potential to enhance theperformancein speaker identification.

According to Suvarnaet al.(2010), GMM has the advantages of minimum model order needed to adequately model speakers and achieve good identification performance and maintain high identification performance with increasing population(Bagul and Shastri, 2012). But the GMM will have the difficulty to estimate the covariance matrices when one of the objects has insufficiently points per mixture. The characteristics of GMM are insensitive to the model initialization method and variance limiting which are very important in training in order to avoid model singularities. Xuetal. (2007) stated that GMM reduces the likelihood of the data and many approaches are presented by researcher to compensate the losing likelihood.

SVM is a binary classification method that finds the optimal linear decision surface based on the concept of structural risk minimization (Raghavan et al., 2006). The decision surface is a weighted combination of elements of a training set. These elements are called support vectors, which characterize the boundary between the two classes (labeled +1 and -1).Schmidt and Gish, (1996) declared that,SVM is inefficient when the number of training frames is large and Vincent (2003) state that SVM in speaker recognition need a normalization process to transform the signal into fixed length due to SVM only can process fixed-length input but speech signals are non-stationary. In order to allow SVM to process speech signals, some pre-processing need to apply to the speech signal. According to Liet al.(2012), in order to enable the SVM to classify the speech signal, a novel kernel function based on GMM supervector or called NAP mappling KL divergence linear kernel function is proposed. This technique has the advantages of channel subspace which cause variability, can be removed in kernel space and improved the classification performance of SVM.

As an alternative, hybrid approachnormally used in current research for pattern recognition. For example, hybrid GMM/ANN(Xiang and Berger, 2003), hybrid HMM/ANN (Heckmannet al., 2000), hybrid GMM/VQ (Pelecanos et al., 2000),hybrid clustering and RBF network(Mashor,2000). Those researches have shown that, hybrid method improve the current traditional method by taking the advantages of two typical pattern classification approaches.

In this research, a hybrid approach will proposed - hybrid of Subtractive Clustering and Radial Basis Function. From the analysis of Subtractive Clustering and RBF network, Subtractive clustering solves the major problem of K-means and Fuzzy C-means (FCM) which face the improper initial guesses of cluster center. Subtractive clustering obtains the cluster centers by compute the density of the data point and subtractive clustering grow exponentially with the size of the data, not the dimension data. RBF have the simpler network structures and faster learning algorithm (Lim and Zainuddin, 2008). RBF finds the input to output map using the local approximators which will combine the linear of the approximators and cause the linear combiner have few weights. Besides that, RBF network is trained with a regularization term to minimize the variances of the nodes and perform more accurate prediction (Yang et al., 2009).

1.3 Problem Statement

There are many recent advances and successes in speaker recognition have been achieved, but a better technique in speaker recognition is still in need. Based on the analysis on the previous techniques in speaker identification, those techniques still suffer from several problems:

i. The most common clustering algorithm involved in those techniques is K-means or fuzzy C-means which has the problem of improper initial guesses of cluster center (Leeet al., 2005).
ii. Mountain clustering depends heavily on grid resolution and the dimension of data which will face the problem of efficiency if the dataset is in high dimension(Hammouda, 2006).
iii. Hybrid of ANN with other technique faces the problem of Multilayer Perceptron (MLP) network which will fall into poor local minima when increasing the number of connection(Cheang, 2009).
iv. Hybrid clustering and RBF network need a suitable clustering algorithm to prevent from lack of ability to choose the most accurate cluster center(Yang et al., 2009).

1.4 Project Aim

The project aims is to propose a new technique in speaker identification by hybrid the subtractive clustering and RBF network which will improve the accuracy rate for speaker identification. It constructs a front-end processing, subtractive clustering for finding cluster center and RBF model for identification task.

1.5 Objectives of the Study

The objectives of the research are:

i. To develop a speaker models by hybrid the subtractive clustering and RBF.
ii. To compare the accuracy rate among the proposed technique with the Subtractive clusteringtechnique.
iii. To evaluate the risk of wrong detection in speaker identification.
iv. To construct a Sub-RBF model based on text-independent environment.

1.6 Scopes of the Study

This research is bound to the following scopes:

i. This research will focus on the model-based approach by subtractive-RBF as a framework for improving speaker recognition.In RBF network model,the hidden node centers of the network is obtained by applying clustering algorithm which can achieve faster training in the network.
ii. Subtractive clustering is chosen to solve the improper initial cluster center and able to train a large set of data.

iii. The result of the proposed method will be compared with the subtractive clustering method from the aspect of accuracy rate and the ability to prevent wrongly identified.

iv. Data set involves is TIMITAccoustic-Phonetic Continuous Speech Courpus which taken from eight different dialect regions and include male and female speakers.

1.7 Thesis Structure

Chapter 1 introduces speaker recognition pattern classification approach and background of the research proposal. The aim, objectives and scope of the research are stated clearly.

Chapter 2 review the general components of speaker recognition application and framework; analyze feature extractor for speaker identification, some pattern classification approach. Besides that, this chapter also analyze the hybrid method in order to increase the accuracy in speaker identification.

Chapter 3 provides a discussion about methodology and theoretical framework of this research. The methodology consists of several procedures, there are planning and literature review, data collection, design of hybrid Subtractive Clustering and Radial Basis Function, evaluate and analysis results and thesis writing.

Chapter 4considers the implementation of the proposed method. In this chapter, the model of the proposed method is designed and implement with the TIMIT dataset. The model consists of three phase - pre-classifier using Subtractive Clustering, Classifier using Radial Basis Function network and Decision phase.

Chapter 5 presents the preparation for the experimental setup and the detail of the conducted experiments with the proposed method in this research. The results of the experiment are discussed in this chapter. Chapter 6 concludes the thesis. This chapter describes the suggestions for future work to improve the proposed method.

REFERENCES

Adrian G. Bors (2001). Introduction of the Radial Basis Function (RBF) Networks. Online Symposium for Electronics Engineers, issue 1, vol. 1, DSP Algorithms: Multimedia, Feb. 13 2001, pp. 1-7.

Adrian G. Bors and Pitas I. (1996) "Median radial basis functions neural network," *IEEE Trans. On Neural Networks*, vil. 7, no. 6, pp 1351-1364.

Bagul, S.G. and Shastri, R.K. (2012). Text Independent Speaker Recognition System using GMM. International Journal of Scientific and Research Publications. Volume 2, Issue 10.

Bataineh, K.M., Naji, M., Saqer, M. (2011). A Comparison Study between Various Fuzzy Clustering Algorithms. Journal of Mechnanical& Industrial Engineering, volume 5(4):335.

Bishop, C. (1995). Neural Networks for Pattern Recognition. Oxford: University Press.

Bourland, H. and Wellekens, J. (1987).Multilayer Perceptron and Automatic Speech Recognition.IEEE Neural Networks.

Cha, I., Kassam, S.A. (1996) "RBFN restoration of nonlinearly degraded images," IEEE Trans. On Image Processing, vol 5, no 6, pp. 964-975

Chakroborty, S., Roy, A. and Saha, G. (2008). Improved Closed set Text-Independent Speaker Identification by Combining MFCC with Evidence from Flipped Filter Banks. International Journal of Signal Processing. Volume 4(2):114-122.

Cheang, S.Y. (2009). Speaker Identification Using Artificial Neural Network By Combination OfResilent Back-propagation And Gradient Descent. UniversitiTeknologiMalayasia: Msc Thesis.

Chiu, S.L. (1994): Fuzzy model identification based on cluster estimation, Journal of Intelligent and Fuzzy systems.

El Hannani, A., and Petrovska-Delacretaz, D. (2006).Fusing acoustic, phonetic and data-driven systems for text-independent speaker verification.

Hammouda, K. (2006). A Comparative Study of Data Clustering Techniques. Future Generation Computer Systems. Volume 13(2-3):149-159.

HarriValpola (2000). Bayesian Ensemble Learning for Nonlinear Factor Analysis Neural Networks Research Centre, Helsinki University of Technology: PhD Thesis.

Haykin, S. (1994). Neural Networks: A Comprehensive Foundation. New York, Macmillan Publishing.

Heckmann, M., Berthommier, F., Savario, C. and Kroschel, K. (2000). Labeling audio-visual speech corpora and training an ANN/HMM audio-visual speech recognition system.

Hermansky, H. (1990). Perceptual Linear Predictive (PLP) Analysis of Speech. Journal of the Acousical Society of America. Volume 87(4):1738-1752

Heyer, L.J., Kruglyak, S. and Yooseph, S. (1999). Exploring expression data: identification and analysis of coexpressed genes. Genome Research - Genome Res, Volume 9(11):1106-1115.

Higgins, A. and Vermilyea, D., (1995). KING Speaker Verification Linguistic Data Consortium, Philadelphia.

Hossain, M., Ahmed, B. and Asrafi, M. (2007)."A real time speaker identification using artificial neural network".Prof. of the 2007 Computer and Information Technology 10th International Conference. December 27-29.

John, H. and Wendy, H. (2002).Speech Synthesis and Recognition. (2nd edition) Bristol, PA, USA : Taylor & Francis, Inc.

KamrulMohmmad, Hasan, Simon Sisavanh, ShrutiTripathi, Emir Hodzic (2006). Team D Speaker Recognition. <http://www.softwarepractice.org/wiki/Team_D_Speaker_Recognition>

Kaski, S. (1997).Clustering methods.<http://users.ics.tkk.fi/sami/thesis/node9.html>

Kekre H. B. and Kulkarni, V. (2010) "Speaker Identification by using Vector Quantization." International Journal of Engineering Science and Technology, Vol. 2(5):1325-1331.

Kinnunen, T., Kilpelinen, T., FrÄnti, P. (2000).Comparison Of Clustering Algorithm In Speaker Identification.In Proc. IASTED Int. Conf. SignalProcessing and Communications (SPC 2000), pages 222–227, Marbella, Spain.

Kothari, R. and Pittas, D. (1999).On finding the number of clusters, Pattern Recognition Letters.

Kotsiantis, S.B., Zaharakis, I.D., Pintelas, P.E. (2007). Supervised Machine Learning: A Review of Classification Techniques.

Lee, Y.J, Lee, J., Lee, K.Y. (2005). Speaker Identification Based on Subtractive Clustering Algorithm with Estimating Number of Clusters. Proceedings of the 8th international conference on Text, Speech and Dialogue.Page 249-256.

Lee, Y.J, Lee, K.Y., Rheem, J.Y. (2003). Efficient Speaker Identification Based on Robust VQ-PCA.ICCSA'03 Proc. of the 2003 international conference on Computational science and its applications: PartII.

Li, H.J., Xing, Y.J. and Tan, P. (2012). SVM speaker verification based on NAP sequence kernels. Audio, Language and Image Processing (ICALIP), 2012 International Conference.

Li, S.Z., Zhang, D., Ma, C.Y., Shum, H.Y. and Chang, E. (2003). Learning to Boost GMM Based Speaker Verification. Eurospeech, Geneva, Switzerland. Sep 2003.

Lian, H., Wang, Z., Wang, J.J., and Zhang, L.M. (2004).Speaker Identification Using Reduced RBF Networks Array.*Advances in Neural Networks*. Volume 3(173):924-929.

Lim, E.A. and Zainuddin, Z. (2008). An Improved Fast Training Algorithm for RBF Networks Using Symmetry-Based Fuzzy C-Means Clustering. MATEMATIKA, Volume 24, Number 2, 141–148, Department of Mathematics, UTM.

Lippmann, R. (1989). Review of Neural Networks for Speech Recognition.Neural Macmillan College Publishing Company, Inc.

Loh, M.Y. (2010). Speaker Identification Using Distributed Vector Quantization And Gaussian Mixture Models. UniversitiTeknologiMalayasia: Msc Thesis.

Loh M.Y. and Abdul Manan, A. (2008). Advances in Speaker Recognition: Techniques and Technology. First edition 2008. Publishedby Universiti Tecknologi Malaysia.

Mashor, M. Y. (2000). Hybrid Training Algorithm for RBF Network.International Journal of The Computer, The Internet and Management. Volume 8(2):50-65.

Milner, B. (2002). A Comparison of Front-End Configurations for Robust Speech. Proceedings of the Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on 13-17 May. Pages 797-800.

Nijhawan, G., Soni, M.K. (2012). A Comparison Study of Two Different Neural Models For Speaker Recognition Systems. International Journal of Innovative Technology and Exploring Engineering.Volume 1(1).

Patnaik, L.M. and Rajan, K. (2000).Target detection through image processing and resilient propagation algorithms.*Neurocomputing*, 35(1-4):123–135.

Pelecanos, J., Myers, S., Sridharan, S. and Chandran, V. (2000). Vector Quantization Based Gaussian Modeling for Speaker Verification. 15th International Conference on Pattern Recognition. 3-7 September. Barcelona, Spain, 3298-3302.

Rabiner, L and Juang, B. H. (2003). Fundamentals of Speech Recognition. (2nd ed.) U.S.A. :Pearson Education

Rabiner, L. R. (1989). A Tutorial On Hidden Markov Models And Selected Applications In Speech Recognition. Proceedings of the IEEE, 77(2), pp 257—286

Rabiner, L. R. and Juang, B.H. (1993).Fundamentals of Speech Recognition, Prentice-Hall, Englewood Cliffs, N.J.

Raghavan, S., Lazarou, G. and Picone, J. (2006). Speaker Verification using Support Vector Machines. Proceedings of the IEEE Southeast Conference. 31 March - April 2. Nashville, Tennessee, 188 - 191.

Recognition. Proc. ICASSP ’2002. 1(1), 797–800.

Reynodls, D., Rose, R., (1995). Robust text-independent speaker identification using Gaussian mixture speaker models.Proceedings of the Speech and Audio Processing, IEEE Transactions on Jan 1995: IEEE, 72-83.

Reynolds, D. A., Quatieri, T. F. and Dunn R. B. (2000). Speaker verification using adapted Gaussian mixture models. Digital Signal Processing, 10:19–41.

Rudasi, L., Zahorian, S.A. (1992). Text-independent speaker identification using binary pair partitioned neural networks. International Joint Conf on Neural Networks IJCNN. IEEE press.

Rumelhart, D., Hinton, G. and Williams, G. (1986).Learning Internal Representations by Error Propagation. Parallel Distributed Processing: Explorations in the Micostructure of Cognition. M.I.T. Press.

Schmidt, M. and Gish, H. (1996).Multilayer Perceptron and Automatic Speech Recognition.IEEE Neural Networks.

Seo, C., Lee, K.Y. , Lee, J. (2001). GMM based on local PCA for Speaker Identification. Electronics Letters 37(24), 1486-1488.

Shang, S., Mirabbasi, S., Saleh, R. (2003).A technique for DCoffset removal and carrier phase error compensation in integrated wireless receivers Circuits and Systems", ISCAS apos; 03. Proceedings of the International Symposium onVolume 1, Issue, 25-28 May 2003 Page I-173 - I-176 vol.1.

Stolcke, A., Shriberg, E., Ferrer, L., Kajarekar, S., Sonmez, K. and Tur, G. (2007).Speech Recognition as Feature Extraction for Speaker Recognition. IEEEWorkshop on Signal Processing Applications for Public Security andForensics. 11-13 April. Washington, DC, USA, 1-5.

Sun, J., Shen, R.M. and Yang, f. (2002). An Adaptive Learning Algorithm Aimed at Improving RBF Network Generalization Ability. In Proceedings of Australian Joint Conference on Artificial Intelligence. 363~373.

Suvarna Kumar, G., Prasad Raju, K.A., Mohan Rao, CPVNJ, Satheesh, P. (2010). Speaker Recognition Using GMM.International Journal of Engineering Science and Technology.Volume 2(6): 2428-2436.

Theodoridis, S.,Koutroumbas, K., (2006). Pattern Recognition, Third Edition. Academic Press, Inc. Orlando, FL, USA.

Vapnik, V. (1995).The Nature of Statistical Learning Theory. Springer-Verlag, New York.

Vergin, R., Oapos, Shaughnessy, D. (1995). Pre-emphasis and speech recognition Electrical and Computer Engineering”, Canadian Conference on Volume 2, 5-8 Sep 1995 Page 1062 - 1065.

Vincent, W. (2003). Speaker Verification using Support Vector Machines. Doctor Philosophy, University of Sheffield, U.K.

Xiang B. and Berger, T. (2003). Efficient Text-Independent Speaker Verification With Structural Gaussian Mixture Models and Neural Network. IEEE Trans. Speech Audio Process 2003. 11(5), 447–456.

Xu, L.M., Tang, Z.M., He, K.K., Qian, B. (2007).Transformation-Based GMM with Improved Cluster Algorithm for Speaker Identification.Proceeding PAKDD'07 Proceedings of the 11th Pacific-Asia conference on Advances in knowledge discovery and data mining. Page 1006-1014.

Yager, R. and Filev, D., (1994). Generation of Fuzzy Rules by Mountain Clustering, Journal of Intelligent & Fuzzy Systems, Vol. 2, No. 3, pp. 209-219,

Yang, P., Zhu, Q.S. and Zhong, X. (2009). Subtractive Clustering Based RBF Neural Network Model for Outlier Detection. Journal of Computers, Volume 4(8):755-762.

Yang, Z.R., Zwaliuski, M. (2001).Mutual information theory for adaptive mixture model, IEEE Trans. Pattern and Machine Intelligence.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V. and Woodland, P. (2001) “The HTK Book”, Version 3.0, July 2001. Cambridge University Engineering Department.

Yue X.C., Ye D.T., Zheng C.X., Wu X.Y. (2002). Neural networks for improved text independent speaker identification. IEEE Engineering in Medicine and Biology Magazine, Vol. 21.