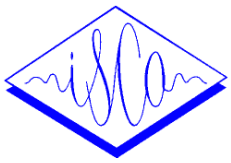

ESCA Workshop on Automatic Speaker Recognition, Identification, and Verification
Martigny, Switzerland
April 5-7, 1994

Speaker Verification over the Telephone Network: Databases, Algorithms and Performance Assessment

Jay Naik

Abstract— **This tutorial is a survey of speaker verification technology and its applications in the telephone network. It describes the algorithms that have been developed to reliably perform personal identity verification over the telephone network. Speech databases for developing and evaluating this technology are described. Assessment of algorithm and system performance and its relevance to practical applications are discussed. Examples of practical implementations of speaker verification technology in the telephone network are also discussed.**

Keywords— **Speaker Verification, Speaker Recognition, Speech databases**

1. Introduction

During the past decade a variety of voice-based services has been successfully deployed in the telephone network in different parts of the world. These services include voice-mail, voice activated dialing and automated telephonic transactions such as banking, reservation systems and call management systems. Automated operator services such as calling card and third-party billing have also become commonplace. As these services proliferate and become increasingly customized to an individual user, the need for securing these transactions has become acute. Access to computer networks, databases and other protected resources is often granted via dial-up telephone lines and require authorization. Speaker verification offers a unique way of performing Personal Identity Verification (PIV) in the telephone network, for a number of attractive reasons.

- PIV based on a person's voice is ideally suited for telephonic transactions
- Easy integration into the existing telephone network
- User interface is simple and cost-effective
- Speaker verification systems can co-exist with and complement speech recognition systems
- User preference is higher than for other biometric PIV's
- Different levels of security can be easily achieved through choice of user dialog and decision criteria

Jay Naik is with the Advanced Voice Services Lab at NYNEX Science and Technology, White Plains, NY. E-mail: naik@nynexst.com

The development of speaker verification technology has largely benefited from advances made in the area of speech recognition[1,2,4]. While these two tasks may differ in a number of ways, such as, speech representation, vocabulary, speaker dependence, training and language/grammar constraints, they share many of the statistical modeling and pattern matching techniques, for example, template matching based on dynamic programming, Hidden Markov Modeling and Multi-layer perceptrons. This parallel development has resulted in integrated system implementations which support both tasks on a single platform, resulting in lower cost and enhanced service. A number of successful systems have already been deployed in the telephone network, with a high degree of user acceptance[2,4,5]. As speaker verification technology leaves the laboratory and enters the marketplace in the form of products and services, we also see a need for standardized procedures and databases for its comparative evaluation. This tutorial will discuss these issues in detail.

2. TASK DESCRIPTION

2.1 Population Size

The task of speaker verification is a subset of the general problem of speaker recognition/identification. The goal of *speaker recognition* is to identify an unknown voice as one or none of a set of known voices. But, *speaker verification* means determining whether an unknown voice matches the voice model of a speaker whose identity is being claimed. Since speaker

verification requires a simple binary comparison, its performance (measured in terms of probability of error) is independent of population size. The speaker identification task requires N+1 decisions for a population size of N speakers (deciding that the unknown voice is one of N known voices or *none* of them) and hence its performance degrades with increasing number of users [2].

2.2 Text Dependence

The performance of the above two tasks is further determined by the type of speech material (often termed "Voice Password") used to validate an identity claim. Fixed-text systems require the recitation of a predetermined text such as a name or a telephone number. Sometimes, the system can prompt the user to say a phrase from a short list of phrases or a random digit string[3,6]. In all these cases, reliable user calibration may be obtained thereby promising high performance. Free-text systems accept speech utterances of unrestricted text. In fixed-text systems, with adequate time alignment one can make precise and reliable comparisons between two utterances of the same text. This is not easily accomplished in free-text systems. Hence fixed-text systems have a higher level of performance than free-text systems. Free-text systems often use long term statistics of the speech signal to extract speaker-specific data and require longer speech material (10-30 sec.) for training and verification. Fixed-text systems typically require 5-10 sec. of speech for training and 2-3 sec. of speech for verification.

For portal access and especially for telephonic access, where verification has to be performed quickly and a high level of user cooperation can be maintained, fixed-text systems have proven to be highly desirable. Free text systems are of value especially in forensic and surveillance applications where the user is not cooperative and often not aware of the task. There are some applications where speaker authentication must be done at the beginning of a service transaction and the channel must be monitored for speaker switching or other misuse of the service. Here, fixed-text systems could perform the initial authentication step and free-text systems, the monitoring step.

3. SPEAKER VERIFICATION ALGORITHMS

The building blocks of a speaker verification system shown in Figure 1. are by and large common to a number of practical implementations of this technology that are currently available in the speech community.

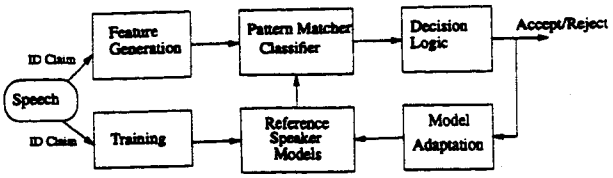

Figure 1. Speaker Verification Task

Enrollment

Creation of a set of speech features for each *valid* user, which represent the dynamics of a speaker's voice characteristics

Verification

Comparison of unknown input speech with reference models for the claimant speaker. Decision based on similarity between input and reference speech models

Reference Update

Adaptation of the speaker reference models to accommodate variations in the valid user's speech. Performed only after successful verification

A variety of speech representations have been developed and tested ([1,2,8,9] are excellent sources for a detailed review). Most fixed-text algorithms use short-term spectral representation, by modeling each speaker's voice as a sequence of vectors, which are derived from short-term spectra of the speech utterance. LPC-cepstral features or LPC-derived filter bank magnitudes are two of the frequently used representations. In recent years, spectral differential information is also used with moderate improvement in speaker discrimination performance[10]. Feature selection methods such as discriminant analysis or Principal Axis methods are also used in a number of algorithms to extract the best set of transformed features, which would reduce storage and computational burden without any noticeable degradation in performance[3,4]. Pitch or a fundamental frequency measure is more resistant to channel degra-

dations than are spectral features, but it is difficult to measure reliably[1]. Pitch is also more susceptible to mimicry than are spectral features. Choosing speech signal features, which are perceptual correlates of *human* speaker recognition or verification have unfortunately not resulted in high performance practical systems. The successful systems have devised *statistical* features, which reliably discriminate between speakers in a population (*low* impostor acceptance), while maintaining stable performance (*high* valid user acceptance) for a speaker, in the presence of speaker variabilities over time and distortions in the signal due to source and channel variabilities. One hopes that a better understanding and modeling of speaker specific characteristics would further enhance the robustness of current speaker recognition/verification systems (for an excellent discussion of this topic, refer to [8] and [9]).

Pattern matching techniques used in fixed-text speaker verification algorithms have closely followed similar work in speech recognition. Dynamic Time Warping (DTW) and Hidden Markov Modeling (HMM) have both been applied to fixed-text speaker verification using telephone speech with operationally acceptable levels of performance [3,4,7,17,21,22,24]. Traditional left-to-right HMM models used in speech recognition have proven to be very robust in fixed-text speaker verification and perform better than DTW based methods for moderate amounts of speech data (two to twenty sec. of speech). The HMM techniques can be designed to perform robust training and to efficiently model non-speech events such as background noise, clicks and non-vocabulary sounds. Their ability to better represent the statistical variability of speech has also contributed to gains in performance over DTW methods [4,23].

Many text-independent speaker verification systems have used Vector Quantization (VQ) methods wherein speaker-specific features are stored as elements of a codebook after a clustering procedure. The speaker specific codebooks obtained with a specific distortion measure are efficient representations of speaker-specific speech signal parameters. During the speaker comparison phase, input vectors for an unknown speaker are compared to the elements of a speaker-trained codebook using the same distortion measure and a nearest-neighbor rule and the cumulative distortion is compared against a decision threshold to make an accept/reject decision [23,25]. A novel method of normalizing decision scores to compensate for trial-to-trial differences is Cohort Normalization [36], which was found to substantially reduce equal-error rates when training and testing were performed on carbon-button and electret microphones respectively.

More recently, Multi-Layer Perceptrons (MLP) have been applied to speech recognition problems [26,27]. Studies show that incorporating the discriminating capabilities and classification power of MLP's with the statistical modeling and temporal segmentation power of HMM's results in a system that is better than either MLP's or HMM's [28]. Our current study of combining HMM's and MLP's to perform text-dependent speaker verification has shown that moderate improvements in speaker discrimination can be obtained from this hybrid approach [29].

The following is a summary of speech feature models used in a number of speaker verification systems reported in the literature.

- LPC and LPC derived Cepstral features and their time derivatives [10]
- Filter-bank and Mel-cepstra [4,30]
- Log-Area Ratios [5]
- Perceptual Linear Predictor(LP) and rasta PLP [35]
- Line Spectrum Pairs [31,32]
- Smoothed Group Delay Spectrum [4]
- Orthogonal transformations of the above feature sets [4]

The performance of a fixed-text speaker verification system developed and evaluated at NYNEX on a telephone speech database is shown in Figure 2. The front-end comprised LPC-Cepstra and their time derivatives computed every 20 msec. and orthogonally transformed. The voice passwords were 10-digit strings spoken by 100 speakers, over a four month period. The database is described in detail in the following section. The pattern matching was based on a continuous density HMM and observation probabilities were modeled as multivariate Gaussian mixtures with state-specific full-covariance matrices. Speaker models comprised segmented whole-word digit models and verification scores were obtained by forced-path Viterbi decoding [33]. A total of 6000 true-speaker trials and 12000 impostor trials were performed on the

telephone database. For an operationally acceptable valid user rejection rate of 1%, the corresponding impostor acceptance rate (performed over speakers who shared same text) was 10.9%. A field trial of this system currently being conducted in the NYNEX telephone network will be briefly described in section 6.

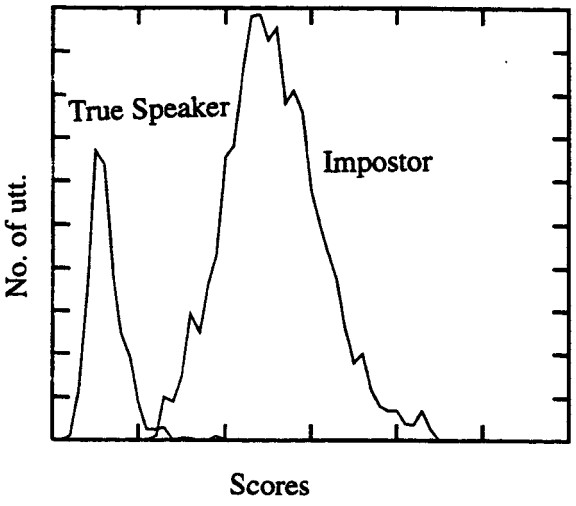

Figure 2. Histogram of Speaker/Impostor Scores

The overall system performance of a speaker verification system is often dependent on choice of voice password, quality of telephone speech database (long distance or local loop, payphone or quiet office), Signal-to-Noise Ratio of the evaluation database, type of training, etc. In general, fixed-text systems operate in the operational region of 0.5-2% valid user rejection and a corresponding 5-20% dedicated impostor acceptance, when evaluated over large telephone speech databases which embody the channel/handset/speaker variabilities commonly encountered in the telephone network. This is quite adequate for operational deployment and easily implementable in hardware. The challenges to building useful services from these systems are cost-effective network integration, efficient user dialog, error recovery and user-friendly call protocol. With the advent of PIV services and a variety of approaches to speech feature modeling and pattern matching techniques, it is imperative that standardized databases be developed and shared in the speech research community to reliably measure progress and to evolve new techniques to improve current methods.

4. SPEECH DATABASES

In the last ten years, a number of standardized speech databases have significantly contributed to the evolution of continuous speech recognition technology. Most notably, the ARPA sponsored TIMIT [11], Resource Management [12] and ATIS [13] databases have helped the speech community to advance the state of the art in speech recognition technology and allowed us to gauge progress under the same set of rules. The importance of such a standardized evaluation methodology cannot be underestimated. It is heartening to see steady progress made in the area of speaker verification and speaker recognition in various parts of the world and a number of practical and productive applications have been deployed in the telephone network. The time is ripe for developing similar standardized procedures for evaluating different speaker verification algorithms as well as their overall system performance in ideal and operational situations. Comparison of fixed-text speaker verification systems is now hindered by the fact that different systems use different protocols such as, type of voice password, decision strategy, enrollment and update methods, etc. Free-text systems are less constrained by some of these issues but type of speech material and amount of testing and training data also vary widely among the systems under development. Both of these modalities must be evaluated with speech databases that reflect operating conditions in the telephone network. In recent years, there is much interest in deploying speaker verification systems in the cellular telephone network, where security has become a prime issue. Several experimental databases have been produced for development and evaluation of different speaker verification systems [4,8,13,15,16], with varying protocols and experimental paradigms. But a set of benchmark databases would serve the purpose of objectively evaluating the merits of each system and providing new directions for further research and development. The following issues should be considered in the design of such a database.

4.1 Choice of text

4.1.1 Fixed-text Systems

A number of user surveys have shown that users prefer to choose their own passwords and a string of seven to ten digits was deemed the best choice, whether they chose a particular digit string or it was assigned to them. A small segment of users prefers names or contrived phrases to digits, in the hope of

greater security but our experience shows that digit strings are highly acceptable. Further, a digit string is well suited for combining recognition and verification of a voice password into a single task [17]. In situations where a high level of security is needed, voice passwords can be constructed on demand as random strings of digits. We can also benefit from the large body of work in connected digit recognition.

4.1.2 Free-text Systems

The choice of text is much less rigid and generally, longer speech material results in higher performance. These systems are less desirable for telephonic applications where authentication has to be performed very reliably and very quickly, but there are a number of applications where 'non-invasive' monitoring of a channel or transaction is desirable and the type of conversation is unconstrained. The recently produced SWITCHBOARD database ([18] and Proceedings of this Workshop) will be an invaluable asset in this domain.

4.2 Population and Session Size

Many of the databases produced for fixed-text algorithm development have a small number of users (twenty to fifty), who provide a small number of calls over a short duration. To reliably predict the performance of the system in the network, a larger and more varied population is desired, especially if some form of speaker-independent recognition is also used as part of the task. A minimum of 100-200 speakers, with an equal number of men and women is very desirable. The performance of a speaker verification system asymptotically approaches a stable level after about ten to fifteen sessions per user, assuming that some form of adaptation of the speaker model is used [4]. Hence the number of sessions from each user must also be large (twenty or more), collected over a duration of several months, at different times of day and calling conditions. Another essential requirement of a fixed-text speaker verification database is to have multiple users produce speech of similar text for impostor testing.

4.3 Telephone Channel Characteristics

The distortions of the speech signal due to variations in the telephone network, both due to the transmission characteristics and handset microphone characteristics have come under serious study in the past few years, for speech and speaker recognition [19]. The database should incorporate a large sampling of easily measurable distortions from a variety of handsets [4] and from a variety of telephone channels across a large telephone network, which are more difficult to control. 'Telephonizing' clean speech will no doubt capture distortions in the transmission channel, but handset microphone distortions and the lack of control over the caller's calling environment are often the more pernicious sources of signal variability.

In recent years, there is a growing interest in the use of microphone arrays as means of suppressing noise in mobile telephony [20]. A speech database collected under these conditions would be of great value, especially with the proliferation of hands-free use of cellular phones in noisy conditions.

5. Example Databases

5.1 Land-line database

A large speech database was collected at NYNEX three years ago to support development of speaker verification and speech recognition services. It is described here as an example of many of the points made earlier. It is hoped that standardized databases for speaker verification evaluation are produced along similar lines and shared among the speech community.

A group of 100 speakers (60 men, 40 women) who were employees of NYNEX volunteered to call an automated data collection system over a period of four months, not more than twice a day. The calls originated mainly in the Northeast U.S., but about 20% of the calls came from other parts of the U.S., placed by traveling NYNEX employees and the calls originated from a total of 28 different area codes in the U.S. A variety of dialects were also represented in the database.

Each call comprised three utterances each of three different ten digit utterances (voice passwords) and a sentence chosen from the TIMIT database sentence list. This set of ten utterances was repeated by the user during every call. The fist voice password was unique to each caller and the other two were shared with at least three other callers of the same gender, for impostor testing. A typical list of utterances is as follows:

"548-076-2931" *(three utterances)*
"968-351-7495" *(three utterances)*
"850-179-4632" *(three utterances)*
Carl lives in a lively home *(one utterance)*

The digit sequences were chosen to yield a uniform distribution of every single digit (zero to nine) and digit pairs in the aggregated database. The participants were instructed to call from as many different phones as they could find. During each session they were prompted to enter a digit to indicate the type of phone from which they called, along with the calling phone number. These phone types included office phones, pay-phones, speaker phones, etc. This was done to gather a general description of the calling environment. About 95% of the calls were made from land-line telephones and the rest from mobile phones. A total of 18,000 ten-digit strings and 2000 sentences were recorded from this group of 100 speakers over a four month period. The speech data was recorded at 16KHz sampling rate from analog phone lines with a Gradient Technology A/D box (It is highly recommended that standardized speech databases in the future should capture data from the T1 channel directly and avoid the variabilities and distortions compounded by the recording equipment). All the data were aurally certified by a trained listener and all anomalies were documented. A variety of speaker verification and connected digit recognition experiments were performed on this database and recent field-trial results show that this database is representative of the channel and user variabilities that are encountered in the real world.

5.2 Cellular Telephone database

A second speech database was collected exclusively on the cellular telephone network, to fuel the development of speech recognition and speaker verification technology in this growing sector of the telecommunication industry. The choice of text for speaker verification was the same as in the land-line case described earlier. In addition a set of six personal names (chosen by the caller), eleven command words ('help', 'yes/no', 'directory', etc.) and eleven isolated digits ('oh', 'zero' to 'nine') were also collected. A total of twenty sessions from each of 80 callers (50 men and 30 women) were recorded using the same recording facility as in the land-line database collection. At the end of each call, the caller described the calling conditions - 'AC on', 'driving on a highway', 'hands-free microphone', etc. These descriptions were also recorded and will be documented as part of each file. In summary, the following class of utterances were recorded:

Familiar Names	*(caller's and five other)*
Isolated Digits	*(0-9 and 'oh')*
Command words	*('directory','redial'..)*
Voice Passwords	*(three 10-digit strings)*

6. AN OPERATIONAL SYSTEM

In recent years, a number of practical speaker verification systems have been tested in the telephone network to combat fraudulent use of network services, banking services, voice-mail, etc. Initial reaction from users of this new service has been positive. While a number of technical and economic issues have to be resolved (pertaining to signal variabilities, user training and adaptation, network management of speaker models and benefit to the user and telephone provider), the future for network-based speaker verification systems is very promising. A field-trial currently underway in the NYNEX telephone network is now briefly described in this section [34].

The traditional method of proffering a digit string via DTMF dialpad as a Personal Identification Number(PIN) to place an automated collect call or bill-to-third party call in the Public Switched Telephone Network is regularly abused by multitudes of fraudulent callers, through theft of the PIN. Speaker Verification is an attractive solution for combating this abuse. NYNEX is presently conducting a field-trial of this service using the speaker verification system described in in section 5. A high-level description of this network integrated service is shown in Figure 3.

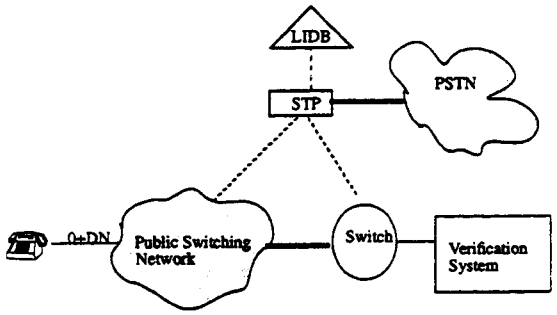

Figure 3. Speaker Verification Service in PSTN

The trial population consisted of over 100 business and residential telephone customers of NYNEX. All calls placed by these callers in the local calling area were routed to the speaker verification system and their identity claim was verified before completing the

call. The voice password was a ten digit phone number chosen by the caller. All verification utterances spoken by the trial participants were saved in their digital form and aurally certified for future enhancements to the SV algorithm. A supervised impostor trial will be conducted this year with a pool of about 100 impostors. Participants were periodically surveyed to obtain their preference ratings and current results show that callers highly approved of this service as a means of combating toll-fraud.

7. CONCLUSION

Speaker Verification technology over the telephone network is now a practical reality. We still have a number of unresolved issues pertaining to signal degradations and variabilities, constraints of fixed-text protocols and implementation of universal services. A formal and rigorous approach to comparative evaluation of speaker recognition/verification technologies is also a necessity. One hopes that the high level of interest in speaker verification generated by this Workshop will serve as a lightning rod to bring about these advances.

REFERENCES

[1] B.S. Atal, *Automatic Recognition of Speakers from their Voices*, Proc. IEEE, vol.64 no.4 pp. 460-475, Apr. 1976.

[2] G.R. Doddington, *Speaker Recognition - Identifying people by their Voices*, Proc. IEEE, vol. 73, pp.1651-1664, No. 1985.

[3] J.M. Naik and G.R. Doddington, *High Performance Speaker Verification Using Principal Spectral Components*, Proc. ICASSP-86, pp.881-884, 1986.

[4] J.M. Naik, L.P. Netsch, and G.R. Doddington, *Speaker Verification over Long Distance Telephone Lines*, Proc. ICASSP-89, pp.524-527, May 1989.

[5] A. Higgins, L. Bahler and J. Porter, *Speaker Verification Using Randomised Phrase Prompting*, Digital Signal Processing, 1, 89-106, 1991.

[6] G.R. Doddington, *Speaker Verification*, Final Rep. RADC-TR-74-179, Griffis Air Force base, Rome N.Y., 1974.

[7] J.M. Naik and G.R. Doddington *Evaluation of a High Performance Speaker Verification System for Access Control*, Proc. ICASSP-87, pp.2392-2395, Apr. 1987.

[8] D. O'Shaughnessy, *Speaker Recognition*, IEEE ASSP Magazine, pp. 4-17, Oct. 1986.

[9] A.E. Rosenberg and F.K. Soong, *Recent Research in Automatic Speaker Recognition*, Advances in Speech Signal Processing, S.Furui and M.M. Sondhi (eds). Marcel Dekker, Inc. 1992.

[10] S. Furui, *Cepstral Analysis Technique for Automatic Speaker Verification*, IEEE Trans. Acoust., Speech and Signal Processing, vol. ASSP-29, no.2, pp.254-272, Apr.1981.

[11] W.M. Fisher, G.R. Doddington and K.M. Goudie-Marshall, *The DARPA Speech Recognition Research Database: Specifications and Status*, Proceedings of the DARPA Speech Recognition Workshop, 1986.

[12] P.J. Price, W.M. Fisher, J. Bernstein and D.S. Pallett, *The DARPA 1000-word Resource Management Database for Continuous Speech Recognition*, Proc. ICASSP-88, New York, April 1988.

[13] C.T. Hemphill, J.G. Godfrey and G.R. Doddington, *The ATIS Spoken Language Systems Pilot Corpus*, Proc. DARPA Speech and Natural Language Workshop, pp. 96-101, June 1990.

[14] M.R. Birnbaum, L.A. Cohen, and F.X. Welsh, *A Voice Password System for Access Security*, AT&T Tech. J., vol.65, no.5, pp.68-74, Sept/Oct 1986.

[15] G. Velius, *Variants of Cepstral Based Speaker Identity Verification System*, Proc. ICASSP-88, pp.583-586, Apr. 1988.

[16] A.E. Rosenberg, C.H. Lee and F.K. Soong, *Sub-Word Talker Verification Using Hidden Markov Models*, Proc. ICASSP-90, vol.1, pp.269-272, April 1990.

[17] A.E. Rosenberg, C.H. Lee and S. Gokcen, *Connected Word Talker Verification Using Whole Word Hidden Markov Models*, Proc. ICASSP-91, pp.381-384, May 1991.

[18] J.G. Godfrey, E.C. Holliman and J. McDaniel, *SWITCHBOARD: Telephone Speech Corpus for Research and Development*, Proc. ICASSP-92, vol. 1, pp.517-520, March 1992.

[19] B.H. Juang *Speech recognition in adverse environments*, Computer Speech and Language, vol.5, no.3, July 1991.

[20] V. Vishwanathan and C. Henry, *Evaluation of multisensor speech input for speech recognition in high ambient noise*, Proc. ICASSP-86, Tokyo, Japan, pp.85-88, April 1986.

[21] S. Furui, *Digital Speech Processing, Synthesis and Recognition*, Marcel Dekker, Inc. NY. 1989.

[22] L. Netsch and G.R. Doddington, *Speaker Verification Using Temporal Decorrelation Post-processing*, Proc. ICASSP-92, San Francisco, pp. II-181-184, March 1992.

[23] S. Furui, *Speaker-dependent-feature extraction, recognition and processing techniques*, Speech Communication, vol.10, pp.505-520, Elsevier Science Publishers, 1991.

[24] T. Matsui and S. Furui, *Speaker Recognition Using Concatenated Phoneme Models*, Proc. Int. Conf. on Spoken Language Processing, pp.603-606, Banff, Canada, 1992.

[25] A.E. Rosenberg and F.K. Soong, *Evaluation of vector quantisation talker recognition system in text independent and text dependent modes*, Computer Speech and Language, vol. 2, no.3/4 Sept/Dec 1987.

[26] Fransini, M. et. al., *A Connectionist Approach to Speech Recognition*, Proc. ICASSP-89, pp.425-428, Glasgow, U.K.,

May 1989.

[27] Bourlard, H. and Wellekens, C.J., *"Speech Pattern Discrimination and Multilayer Perceptrons"*, Computer, Speech and Language, vol.3 pp 1-19, 1989.

[28] Morgan, N. and Bourlard, H., *Continuous Speech Recognition Using Multilayer Perceptrons with Hidden Markov Models*, Proc. ICASSP-90, pp.413-416, Albuquerque, April 1990.

[29] J.M. Naik and D.M. Lubensky, *A Hybrid HMM-MLP Speaker Verification Algorithm for Telephone Speech*, To be presented at ICASSP-94, Adelaide, April 1994.

[30] F. Bimbot and L. Mathan, *Text-Free Speaker Recognition Using an Arithmetic-Harmonic Sphericity Measure*, Proceedings of EUROSPEECH-93, Berlin, Sept. 1993, pp.169-172.

[31] C. Liu, M. Tin, W. Wang and H. Wang, Study of Line Spectrum Pair Frequencies for Speaker Recognition, Proc. ICASSP-90, Albuquerque, pp.277-280, April 1990.

[32] J.P. Campbell, Features and Measures for Speaker Recognition, PhD. Dissertation, Oklahoma State Univ., Dec. 1992.

[33] L.R. Rabiner, *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, Proc. IEEE, vol.77, no.2, pp.257-286, Feb. 1989.

[34] J. Cheng, E. Martinez and J. Naik, *Preventing Calling Card Fraud Using Speaker Verification* BellCore/NYNEX Network Security Symposium, Boston, May 1993.

[35] H. Hermansky, N. Morgan, A. Bayya and P. Kohn, *Compensation for the effect of the communication channel in auditory-like analysis of speech*, Proc. of EUROSPEECH-91, pp. 1367-1370, Genoa, Sept. 1991.

[36] A.E. Rosenberg, J. DeLong, C.H. Lee, B.H. Juang, and F.K. Soong, *The use of cohort normalised scores for speaker verification*, Proc. Int. Conf. on Spoken Language Processing, pp.599-602, Banff, Canada, 1992.