ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 2, March 2016*

National Conference on Future Technologies in Power, Control and Communication Systems (NFTPCOS-16)

on 10, 11 and 12th March 2016

Organised by

Dept. of EEE, College of Engineering Perumon, Kollam, Kerala – 691601, India

Text-Indicated Voice Verification with Message Authentication Code

Jincy C[1], Rinku K R[2]

Assistant Professor, Dept. of ECE, Sivaji College of Engineering & Technology, Manivila, Tamilnadu, India[1]

PG Student [Communication Systems], Dept. of ECE, Sivaji College of Engineering & Technology, Manivila, Tamilnadu, India[2]

ABSTRACT: Speech processing is emerged as one of the important application area of digital signal processing. The main objective of Speaker recognition is its computational task of validating a person's identity based on their voice. It is the process of automatically recognizing who is speaking on the basis of individual information included in speech waves. The two phases of a speaker recognition system are the enrollment phase where speech samples from different speakers are turned into models and in the verification phase where a sample of speech is tested to determine if it matches a proposed speaker. In a text-indicated system, there are no constraints in the words or phrases used during verification. A speaker verification function protects private information and separates the user's voice from that of people nearby who are also speaking. A text-indicated speaker recognition system uses CELP Parameters. A CELP-based speaker verification method is used to match the audio stream by comparing the trajectories of continuous phonemes and its analysis techniques will be implemented in Proteus. The goal of this project is to build a simple, yet complete and representative automatic speaker recognition system. It is possible to use the speaker's voice to verify their identity and control access to services.

KEYWORDS: Speaker Recognition, Consumer Electronic Devices, CELP, LSP, Speaker Verification.

I.INTRODUCTION

Humans have the innate ability to recognize familiar voices within seconds by hearing a person speaking. Researches in speaker recognition or verification, the computational task of validating a person's identity based on their voice, began in 1960 with a model based on the analysis of X- rays of individuals making specific phonemic sounds. With the advancements in technology over the past 50 years, robust and highly accurate systems have been developed with applications in automat password and it will helps to prevent the misuse by a malicious user and reset capabilities, forensics and home healthcare verification.

There are two phases in a speaker recognition system: an enrollment phase where speech samples from different speakers are turned into models and the verification phase where a sample of speech is tested to determine if it matches a proposed speaker. It is assumed that each speech sample pertains to one speaker. A robust system would need to account for differences in the speech signals between the enrollment phase and the verification phase that are due to the channels used to record the speech (landline, mobile phone, handset recorder) and in consistencies within a speaker (health, mood, effects of aging) which are referred to as channel variability and speaker dependent variability respectively.

In text-dependent systems, the words or a phrase used for verification is known be forehand and are fixed. In a text - independent system, there are no constraints on the words or phrases used during verification. These make them more convenient to use but increases the risk of deception because an imposter can easily use a voice recording of the authorized user, so which is not appropriate when high security is required. Text-indicated methods were developed to overcome this problem. This project will focus on text-indicated speaker verification systems. This paper details the construction and building of a stand-alone and very less expensive speech recognition technique that may be used to control just about anything such as electrical appliances, robots, test instruments, Fans, Light, TV's, etc.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 2, March 2016*

National Conference on Future Technologies in Power, Control and Communication Systems (NFTPCOS-16)

on 10, 11 and 12th March 2016

Organised by

Dept. of EEE, College of Engineering Perumon, Kollam, Kerala – 691601, India

In this paper, it proposes of two states, a training state, and a test/verification state: The first step of training is to transform the speech signal to a set of vectors, and to obtain a new representation of the speech which is more suitable for statistical modelling. Firstly the speech signal is broken up into short frames, then windowed to minimize distortion. The signal is then analysed and stored as that user's template. The first step in the testing stage is to extract features from the input speech, similar to that during training, compare the input speech to all other stored templates and select the most accurately matching template and ID the speaker. The main process in the speaker identification process is the feature extraction. The representation of speaker verification is as shown below. Speaker/Voice identification is the process of determining which registered speaker provides a given utterance. Speaker verification, on the other hand, is the process of accepting or rejecting the identity claim of a speaker. Speaker recognition is a difficult task and it is still an active research area. Automatic speaker recognition works based on the premise that a person's speech exhibits characteristics that are unique to the speaker. However this task has been challenged by the highly variant of input speech signals.

II.SYSTEM MODEL AND SPEAKER VERIFICATION METHOD

The proposed system supports voice operation by using the CELP parameters that are encoded in mobile phones. Speaker verification and a speech recognition function is needed to support voice operation are the important function in this system. It is done by using line spectrum pairs encoded by CELP parameters. LSP corresponds to a parameter for articulation of speech. The verification algorithm used for this "CELP-based speaker verification" is the focus of this paper and also provides efficiency. CELP parameters are used in the speaker verification and speech recognition functions, which are assumed to be performed on consumer electronic devices to prevent misuse by a malicious user. A flow of the proposed CELP-based speaker verification is shown in Fig. 1. Speaker verification has two phases, enrollment and verification.

In this paper, we will discuss only the text independent but speaker dependent Speaker Recognition system. All technologies of speaker recognition, identification and verification, text-independent and text dependent, each has its own advantages and disadvantages and may require different treatments and techniques. The choice of which technology to use is application-specific. At the highest level, all speaker recognition systems contain two main modules: feature extraction and feature matching. Feature extraction is the process that extracts a small amount of data from the voice signal that can later be used to represent each speaker. Feature matching involves the actual procedure to identify the unknown speaker by comparing extracted features from his/her voice input with the ones from a set of known speakers.

A wide range of possibilities exist for parametrically representing the speech signal for the speaker recognition task, such as Linear Prediction Coding (LPC), Mel-Frequency Cepstrum Coefficients (MFCC).LPC analyses the speech signal by estimating the formants, removing their effects from the speech signal, and estimating the intensity and frequency of the remaining buzz. The process of removing the formants is called inverse filtering, and the remaining signal is called the residue. Another popular speech feature representation is known as RASTA-PLP, an acronym for Relative Spectral Transform - Perceptual Linear Prediction. PLP was originally proposed by Hynek Hermansky as a way of warping spectra to minimize the differences between speakers while preserving the important speech information. RASTA is a separate technique that applies a band-pass filter to the energy in each frequency subband in order to smooth over short-term noise variations and to remove any constant offset resulting from static spectral coloration in the speech channel e.g. from a telephone line.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 2, March 2016*

National Conference on Future Technologies in Power, Control and Communication Systems (NFTPCOS-16)

on 10, 11 and 12th March 2016

Organised by

Dept. of EEE, College of Engineering Perumon, Kollam, Kerala – 691601, India

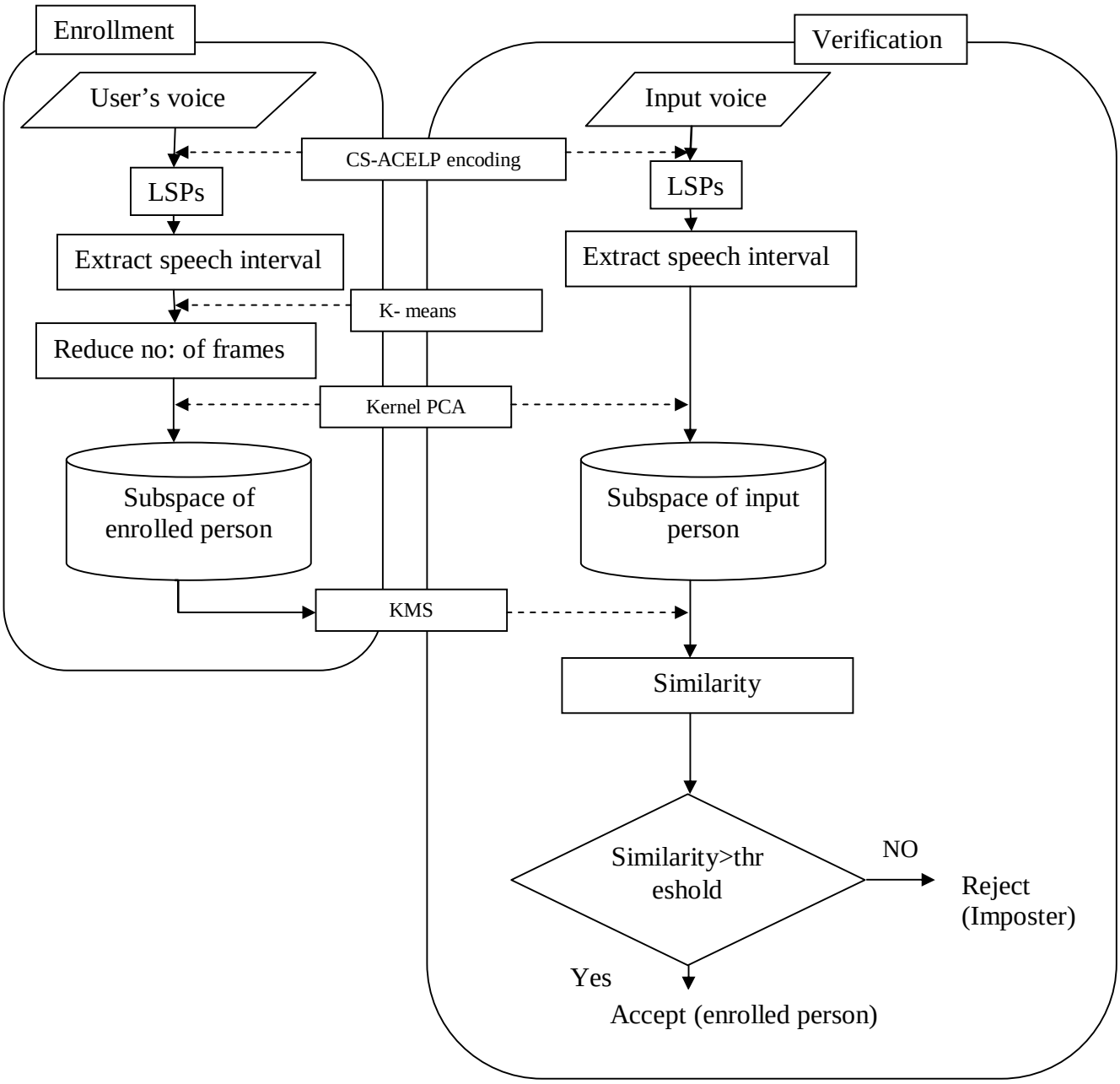

Fig.1 Flow Diagram of Speaker Verification Method

In the enrollment stage, first user's voice is transformed into an electronic signal by using a microphone. Then Encode the electronic signal by using CS-ACELP (conjugate structure algebraic CELP) encoding and extract the LSPs; represent the frame data as a vector concatenating the LSPs extracted from each frame. Then extract a speech interval from the speech signal by using the power of the speech signal (remove silence intervals from speech signal). Reduce the number of frames by replacing the frame data with cluster centres calculated using the K means method. Calculate

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 2, March 2016*

National Conference on Future Technologies in Power, Control and Communication Systems (NFTPCOS-16)

on 10, 11 and 12th March 2016

Organised by

Dept. of EEE, College of Engineering Perumon, Kollam, Kerala – 691601, India

the basis by using the kernel principal component analysis (KPCA) from the reduced frame data and represent the enrollment subspace in terms of the basis. Enrol the enrollment subspace as genuine data in the system database.

In the Verification stage, first transform input voice into an electronic signal by using a microphone. Encode the electronic signal by using CS-ACELP encoding and extract the LSPs; represent the frame data as a vector concatenating the LSPs extracted from each frame. Extract a speech interval from the speech signal by using the power of the speech signal (remove silence intervals from speech signal). Calculate the basis by using the KPCA from the reduced frame data and represent the input subspace in terms of the basis. Calculate the similarity (the square of the cosine of the canonical angle between the enrollment subspace and the input subspace). Compare the similarity with the threshold. If the similarity is greater than the threshold, judges "accept." Otherwise, judge "reject." The threshold value is set by the authorized person/user. According to that parameter accept /reject decision takes place.

III. CELP CODING

In the CELP encoding process, the input speech data is encoded using CELP and the encoding parameters are extracted. With CELP, a linear predictive coding (LPC) is used to calculate a synthesis filter for quantization as line spectral pairs (LSPs). Analysis by synthesis is then used to calculate an excitation signal for vector quantization. The CELP framework is widely used for highly efficient voice encoding for mobile phones and nearly all other mobile equipment.

CS-ACELP uses 8-kbit/s encoding; it produces the same voice quality as 32-kbit/s ADPCM (Adaptive Differential PCM). For efficient encoding process uses CS-ACELP. In addition to being adopted as the full-rate standard codec for digital mobile phones in Japan, it is being used in voice over Internet protocol services and many other applications worldwide. The proposed verification method uses CS-ACELP, which has been standardized as ITU-T G.729. Coder is based on a CELP coding model and uses an analysis-by-synthesis technique to determine the vectors to construct the best excitation which minimizes the perceptually weighted distortion between the original and synthesized speech.

LP analysis is done first, and the speech signal is expressed as the computed linear prediction coefficients (LPC). The quantized LPC coefficients are used in the synthesis filter of this coder. a linear predictive coding (LPC) is used to calculate a synthesis filter for quantization as line spectral pairs (LSPs). The excitation of this filter consists of two parts: one is the adaptive codebook vector which simulates the pitch structure of the voiced sound, and the other is the fixed codebook vector which simulates the unvoiced sound. The two codebook vectors scaled by their respective gains are summed to construct the excitation of the synthesis filter. The synthesized speech is constructed and has the least distortion relative to the original speech, and the searched codebook vectors are the best ones.

IV. SPEAKER VERIFICATION PROCESS

In the *K means Clustering method,* the continuous speech signal is blocked into frames of N samples, with adjacent frames being separated by M (M < N). The first frame consists of the first N samples. The second frame begins M samples after the first frame, and overlaps it by N - M samples. Similarly, the third frame begins 2M samples after the first frame (or M samples after the second frame) and overlaps it by N - 2M samples. This process continues until all the speech is accounted for within one or more frames. Typical values for N and M are N = 256 (which is equivalent to ~ 30 ms windowing and facilitate the fast radix-2 FFT) and M = 100. The KMS must be able to solve an Eigen value problem in order to represent data in a subspace, which means that the amount of processing is proportional to the cube of the number of frame. When similarity is calculated at the time of recognition, the kernel function must be calculated between all enrollment data and input data targeted for recognition, resulting in a large amount of processing. The K-means clustering method is as follows. Select the initial cluster centre from m frames; set X_i(i=1…..m)randomly. Then divide each X_i(i=1,..m) into cluster C_j of μ_j that minimizes $\|X_{\_i}-\mu_{\_j}\|^2$. Calculate $\sum\sum$ (j=1)^K$\|X_{_i}-\mu_{_j}\|^2$.If the output is less than the threshold value then substitute μ_j(j=1,….K, K ≤ m). Otherwise calculate μ_j=1/N $\sum_{(i=0)}$^n $X_{\_i}$ for each cluster and upgrade the cluster centre.

In the *Kernel Mutual Subspace method,* the accuracy of a recognition system that takes a data stream as input can be improved by increasing the amount of information. The heart of MSM is feature extraction by PCA for both enrollment and input data. Features representing subspaces are calculated from the enrollment and input data. The similarities between the enrollment and input subspaces are represented by the square of the cosine of the canonical angles between

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 2, March 2016*

National Conference on Future Technologies in Power, Control and Communication Systems (NFTPCOS-16)

on 10, 11 and 12th March 2016

Organised by

Dept. of EEE, College of Engineering Perumon, Kollam, Kerala – 691601, India

the subspaces. In the learning phase of MSM, the basis obtained from the enrollment data is registered as a template (which is the same approach as in the conventional subspace method). The recognition phase involves three steps. First, calculate the PCA basis from the input data. Then the matrix is calculated by using the following equation,

$$Z= (Z_{ij}) = (\sum_{m=1..n} (V_i.W_m)(W_m.V_j)) \qquad (1)$$

Where V is the basis of the enrollment data subspace and W is the basis of the subspace obtained from the input data. Obtain the maximum eigenvalue of Z, which is the angle between the two subspaces.MSM is a powerful object recognition method because it can be used to extract stable features from a data stream. It does not work well if the distribution of voice data has a nonlinear structure. To solve this problem, the Kernel Mutual Subspace (KMS) method is introduced. This method combines MSM with KPCA, which provides a powerful nonlinear analytic capability. As a result, KMS is a powerful object recognition method when the voice data distribution is nonlinear.

V. RESULT AND DISCUSSION

Proteus is best simulation software for various designs with microcontroller. It is mainly popular because of availability of almost all microcontrollers in it. So it is a handy tool to test programs and embedded designs for electronics. The voice signals from Android Device are transformed into commands by speech recognition. The voice signals will contains the message signals and some sort noise signals which is caused by the nearer.

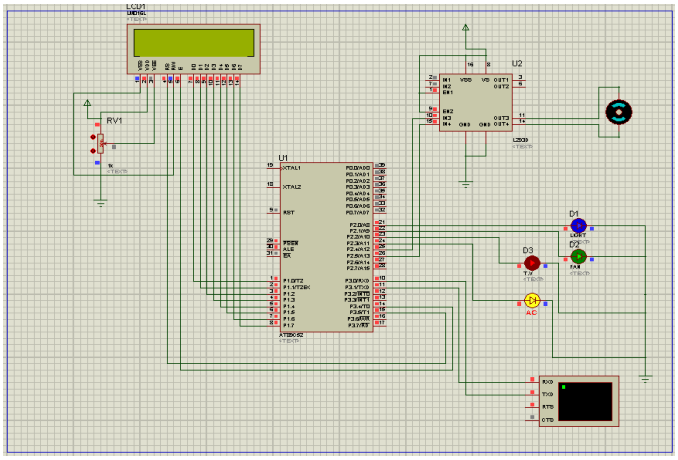

Fig.3 Module after Execution

A CELP-based speaker verification method is used to match the audio stream by comparing the trajectories of continuous phonemes. CELP (Code Excited Linear Prediction) parameters, which are commonly used in speech coding for mobile phones. A speech recognition function is needed to support voice operation. User speech is compared with original speech sampled which is stored in mobile database when the user sample is matched it send command to microcontroller unit by the help of Bluetooth. Microcontroller device which take care LCD Display unit, Bluetooth module and Relay unit. When microcontroller get any signal from mobile Bluetooth, based on the command it controls all the Electronic Device which is connected to the Relay unit. LCD Display shows the operation status of each Electronic Devices.

After giving executes command in the simulation software, the LCD Display will turn ON. According to the commands the corresponding devices will be turn ON/OFF. Since the microcontroller uses Atmel high-density, nonvolatile memory technology and is compatible with the industry-standard 80C51 instruction set and pinout. On-chip flash allows the program memory to be reprogrammed in-system or by a conventional nonvolatile memory programmer. The microcontroller having 8 Input/output lines providing a total of 32 I/O lines. Those ports can be used to output DATA and orders do other devices, or to read the state of a sensor, or a switch. Performing serial data transfer or connecting

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 2, March 2016*

National Conference on Future Technologies in Power, Control and Communication Systems (NFTPCOS-16)

on 10, 11 and 12th March 2016

Organised by

Dept. of EEE, College of Engineering Perumon, Kollam, Kerala – 691601, India

the chip to a computer to update the software. Each port has 8 pins, and will be treated from the software point of view as an 8-bit variable called 'register', each bit being connected to a different Input/output pin.

VI.CONCLUSION

The proposed system defines how a consumer can operate electronic devices by voice using the CELP parameters that are used for speech coding in mobile phones. Private information of the user is being protected by a speaker verification function. It separates the user's voice from that of people nearby who are also speaking. A CELP-based speaker verification method is used to match the characteristics of a set of frames by comparing the trajectories of continuous phonemes. From the simulation results, it is clear that the proposed method is effective for speaker verification using encoded speech information such as speaker verification in mobile communication systems. Experimental evaluation of the verification method demonstrated its effectiveness.

REFERENCES

[1] Y. Yamazaki, Y. Fujita, and N. Komatsu, "CELP-based Speaker Verification: An Evaluation under Noisy Conditions," *IEEE Internation conference on Control, Automation, Robotics and Vision*, pp. 408-412,Dec. 2004

[2] H. Sakano, N. Mukawa and T. Nakamura, "Kernel Mutual Subspace Method and Its Application for Object Recognition," *Electronics and Communications in Japan*, Vol.E88, No.6, pp. 45-53, Jun. 2005.

[3] ITU-T, "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)," ITU-T Recommendation G.729, 1996.

[4] M. Ichino, H. Sakano and N. Komatsu, "Reducing the Complexity of Kernel Mutual Subspace Method Using Clustering," *IEICE TRANS. On Information and Systems (Japanese Edition)*, vol.J90-D, pp.2168-2181,Aug. 2007.

[5] B. Schölkopf, A. Smola and K.R. Müller, "Nonlinear component analysis as a Kernel eigenvalue problem," *Neural Computation*, vol.10,no. 5, pp.1299-1319, 1998.

[6] J. Decuir, "Introducing Bluetooth Smart: Part II: Applications and updates," *IEEE Consumer Electronics Magazine*, pp.25-29, Apr. 2014.

[7] K. M. Lee and J. Lai, "Speech Versus Touch: A Comparative Study of the Use of Speech and DTMF Keypad for Navigation," *International Journal of Human-Computer Interaction*, pp.343-360, Vol. 19, issue 3, Jan. 2005.

[8] S. Itahashi, "Creating Speech Copora for Speech Science and Technology," *IEICE TRANS. on Fundamentals of Electronics,Communications and Computer Sciences*, Vol.E74-A, No.7, pp. 1906- 1910, Jul. 1991.

[9] K. Maeda and S. Watanabe, "A Pattern Matching Method with Local Structure," *IEICE TRANS. on Information and Systems (Japanese Edition)*, vol.J68-D, no.3, pp.345-352, Mar. 1985.

[10] O. Yamaguchi, K. Fukui and K. Maeda, "Voice Recognition System using Temporal Image Sequence," *IEEE International Conference on voice and speech Recognition*, pp. 318-323, Apr. 1998.