ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 1, February 2017

DYSARTHIC SPEECH RECOGNITION USING RANDOM PROJECTION BASED AUTOMATIC SPEECH RECOGNITION TECHNIQUE

[1]W.Brajula and [2]R.Mohan kumar

[1] P.G Scholar, Dept. of Electronics and Communication Engineering, Udaya School of Engineering, Vellamodi Tamilnadu, India
[2]Assistant Professor, Electronics and Communication Engineering, Udaya School of Engineering, Vellamodi, Tamilnadu, India.

ABSTRACT **:** We investigate the speech recognition of persons with articulation disorders resulting from cerebral palsy in this paper. The articulation of their first speech tends to become more unstable due to strain on speech-related muscles and that causes degradation of speech recognition. We here propose a feature extraction method based on RP (Random Projection) and the final output will be in the form of continuous speech for dysarthric speech recognition. Random projection has been suggested as a means for space mapping, where the original data are projected onto a space using a random matrix based on orthogonal projection and the distance between the words are preserved.. It represents a computationally simple method that approximately preserves the Euclidean distance of any two points through the projection of the words. The Automatic Speech Recognition (ASR) plays a main role in the conversion from speech to text. Moreover, we are able to produce various random matrices, there may be some possibility of finding a random matrix that gives better speech recognition accuracy among these random matrices in order to get a clear speech. To obtain an optimal result from many of the random matrices, a vote-based combination is introduced in this paper. ROVER combination is applied to the recognition results, obtained from the ASR systems created from each of the RP-based feature. The feature gives only the words in the form of text and it is finally converted into a fine and a continious speech with the help of Text –to – Speech API. Its effectiveness is confirmed by word recognition experiments.

Index Terms **—** Articulation Disorder, Cerebral Plasy, Speech Recognition, Random Projection, Automatic Speech Recognition , ROVER

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 1, February 2017

1. INTRODUCTION

Recently, the importance of information technology in the welfare - related fields has increased more and more. For example, sign language recognition using image recognition technology [1][2][3].Such as, text-reading systems from natural scene images [5][6], and the design of wearable speech synthesizers for voice disorders[7][8] and the text to speech conversion from the scanned papers have been studied. There are 34,000 people with speech impediments associated with articulation disorders in Japan alone, and it is hoped that speech recognition systems will one day be able to recognize their voices as few of their words could be grumbling.

Articulation disorder is generally classified into two types, namely: Functional Disorder and Organic Disorder. When, structure, hearing and observable physical systems appear to be normal, the articulation function is termed to be functional in nature and its origin attributed to be faulty learning. Most articulation disorder belongs to this type. If the defective utterance is related to structural deformity or physical defect, the articulation problem is termed to be organic in nature and this may be caused by some insult to the brain such as stroke, infections, tumours or trauma. Christo Ananth et al. [4] presented an automatic segmentation method which effectively combines Active Contour Model, Live Wire method and Graph Cut approach (CLG). The aim of Live wire method is to provide control to the user on segmentation process during execution.

One of the causes of speech impediments is due to cerebral palsy. Cerebral palsy results from the damage to the central nervous system, and the damage causes movement disorders. Three general times are given for the onset of the disorder may be before birth, at the time of delivery, and after birth. It is a non-curable, Life-long condition and the damage do not get worsen day by day. Cerebral palsy can be classified as: 1) spastic type 2) athetoid type 3) ataxic type 4) atonic type 5) rigid type, and a mixture of types[9].

In this paper, we focus on persons with articulation disorders resulting from the athetoid type of cerebral palsy. Athetoid symptoms develop in about 10-15% of cerebral palsy sufferers from unstable movement mostly at their first speech. In the case, Person with this type of articulation disorder, the first movements are sometimes more unstable than usual. That means, in the case of speaking-related movements, the first utterance is often unstable or unclear due to the athetoid symptoms, that causes degradation of speech recognition. Therefore, we record speech data for the persons with articulation disorders, who uttered each of the words several times, and then investigated the influence of the unstable speaking style caused by the

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 1, February 2017

athetoid symptoms.

The goal of front-end speech processing in ASR is to obtain a projection of the speech signal to a compact parameter space where the information related to speech content can be extracted clearly. In current speech recognition technology, MFCC (Mel-Frequency Cepstrum Coefficient) is being widely used as a filter. The feature is uniquely derived from the mel-scale filter-bank output by the DCT (Discrete Cosine Transform).

The low-order MFCCs account for the slowly changing in the spectral envelope, while the high -order ones describe the fast variations of the spectrum. Therefore, a large number of MFCCs are not used for speech. Because, we are only interested in the spectral envelope, not in the fine structure. As suggested by H. Matsumasa, we have proposed a robustic feature extraction based on PCA (Principal Component Analysis)[10] with more stable utterance data instead of DCT in a dysarthric speech recognition task. Also C. Miyamoto, used MAF (multiple acoustic frames) as an acoustic dynamic feature inorder to improve the recognition rate of a person with articulation disorders[11], especially in speech recognition using dynamic features only.

These methods improved the recognition rate of accuracy, but the performance for articulation disorders was not sufficient when compared to that of persons with no disability. Random projection has been suggested as a means of spacial mapping, where a projection matrix is composed of the columns and rows as defined by the random values chosen from a probability distribution.

In addition to it, the Euclidean distance of any two points is approximately preserved through the projection. Therefore, random projection has also been suggested as a means of dimensionality reduction as said by N. Goel, G. Bebis, and A. Nefian [12]. In contrast to conventional techniques such as PCA, which find a subspace by optimizing certain criteria, random projection does not use such criteria; therefore, it is data independent. Moreover, it is represents as computationally simple and efficient method that preserves the structure of the data without introducing significant distortion [13]. Goel et al[13] have reported that random projection has been applied to various types of problems, including information retrieval, sign based language[14], image processing [15][16], machine learning [17]-[19]), and so on. Although it is based on a simple idea, random projection has demonstrated good performance in a number of applications, as it yields results comparable to conventional dimensionality reduction techniques, such as PCA.

In this paper, we investigated the feasibility of random projection for speech

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

feature transformation in order to improve the recognition rate of persons with articulation disorders. In our proposed method, original speech features (MFCCs) are transformed using various random matrices composed of rows and columns. we use the same number of dimensions for the projected space as that of the original space. There may be some possibility of finding a random matrix that gives better speech recognition accuracy among random various matrices.As, we are able to produce various RP-based features (using various random matrices).

Therefore, a vote-based combination method is introduced here,in order to obtain an optimal result from many (infinite) random matrices, where ROVER combination[20] is applied to the results from the ASR systems created from each RP-based features .The rest of the paper is organized as follows. Section 3.1 describes a feature projection method using random orthogonal matrices. In Sect ion 3.3, a vote-based combination method is explained. Results and discussion for the experiments on a dysarthric speech recognition task are presented in Section 4.

2. RELATED WORKS

2.1.Mel-Frequency Cepstum Coefficient

The data is recorded using a data recording digital audio tape recorder. Then it is digitized using a 24 bit digital I/O card. An Artificial Neural Network (ANN) is used as the classifying tool. It consists of a potentially large number of simple processing elements called nodes, which influence each other's behavior via a network of excitatory or inhibitory weights. Each node simply computes a nonlinear weighted sum of its inputs and transmits this result over its outgoing connections to other units. A training set consists of patterns of values that are assigned to designate input and/or output units.

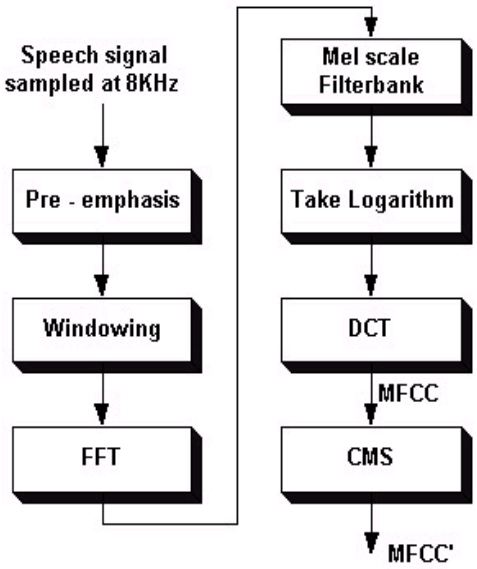


Fig.1.Mel-Frequency Cepstum Coefficient

Here the Discrete Cosine Transform (DCT) is used, which depends on the cosine function. It completely dependsonly on the

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

development of the spectrum and hence the Space between the words are more and also leads to Loss of useful data. Therefore, in case of DCT the Discrete Fourier Transform (DFT) is used. The DCT fist finds the spectrum. The bank of filters based on Mel-Scale is applied and each of the filter output is the sum of its filtered spectral components this leads to Less speech clarity, Low accuracy and Lack of stability. It focuses only on fine structure of the waveform and has no concern about the formation of envelope.

2.2 HMM – Hidden Markov Model

In case of, Existing system Hidden Markov Model (HMM) is used. The Hidden Markov Model is a finite set of states, each of which is associated with (generally multi dimensional) probability distribution. Transitions among each states are governed by a set of probabilities called transition probabilities. In a particular state of an outcome or observation can be generated, according to the associated probability distribution. It is only the outcome, not the state which is visible to an external observer and therefore states are "hidden"' to the outside; hence the name Hidden Markov Model. Finally, high frequency sounds signals are less informative, so can be sampled using a log scale.

3.PROPOSED METHOD

3.1. Random Orthogonal Projection

This section presents a feature projection method using random orthogonal matrices. The main idea of random projection arises from the Johnson-Lindenstrauss lemma [21]; namely, if original data are projected onto a randomly selected subspace using a random matrix of the orthogonal projection, then the distances between the data are approximately preserved by Euclidean space. Random project ion is a simple yet powerful technique, and it has another benefit. Dasgupta[17] has suggested that even if distributions of original data are highly skewed (have ellipsoidal contours of high eccentricity) and their transformed counterparts will be more spherical than usual.

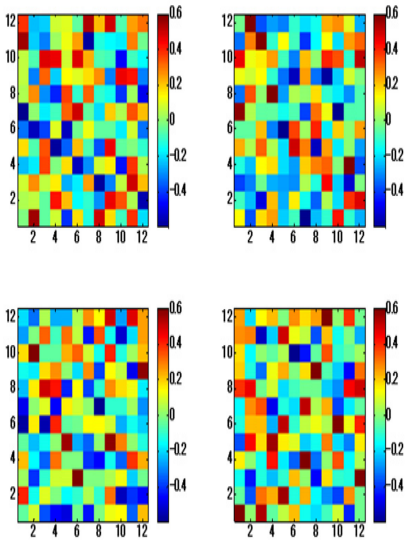


Fig.2. Examples of random matrices 12 dim. (12 × 12)

The algorithm can be given as:

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 1, February 2017

- The original n-dimensional vector X, is projected onto the d-dimensional subspace using the l-random matrices.

- Choose each entry of the matrix from an independent and identifiably Normal distributed N(0,1)value.

- Make the orthogonal matrix using the Gram Schmidt Orthogonal algorithm, and then normalize it to unit length.

- HMM is used for speech recognition, as it has diagonal convergence matrix.

First, we choose an n-dimensional random vector,**P** and let $\mathbf{P}^{(l)}$ be the *l*-th n x d matrix whose columns are vectors $P1^{(l)}$, $P2^{(l)}$, $P3^{(l)}$…$pd^{(l)}$,Then, an original n-dimensional vector X , is projected onto a d-dimensional subspace using the l^{th} -random matrix, $P^{(l)}$, where we compute a d-dimensional vector, X' , whose coordinates are the inner products of the terms,

$$\mathbf{X_1'=P_1^{(l)}x,\ldots,x_d'=P_d^{(l)}x \quad and}$$
$$\mathbf{x' = P^{(l)T} . x}$$

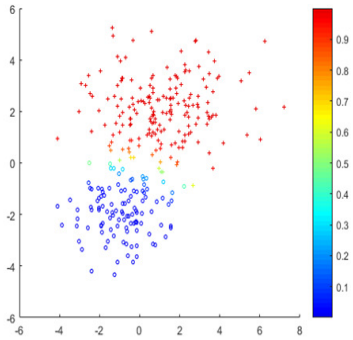


Fig.3. Scattering of signals in plots

We here investigate the feasibility of random projection for speech feature transformation as a main key in this paper. As described above, a random projection from n-dimensions to d = n dimensions is represented by an n x d matrix, P. It has been shown that if the random matrix, P, is chosen from the standard normal distribution, with mean 0 and variance 1, referred to as N(0,1). Then, the projection preserves the structure of the data[21]. In this paper we use N(0,1) for the distribution of the coordinates. The random matrix, P, can be calculated using the algorithm [13][17].

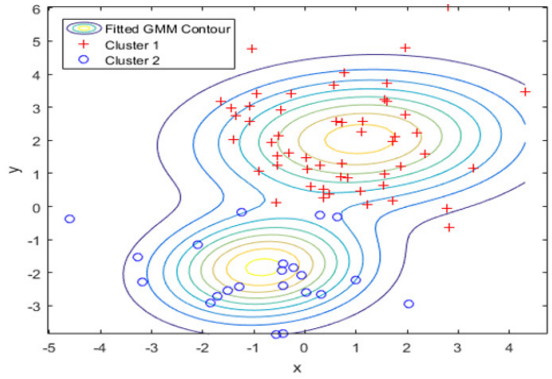


Fig.4. Data cluster arrangement

3.2. Vote-Based Combination

As described in Section 3.1, we can make many (infinite) random matrices from *N*(0,1) (Fig. 2). Since there may be some possibility of finding a random matrix that

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

gives better performance, we will have to select the optimal matrix or the optimal recognition result from those.

To obtain the optimal result, a majority vote-based combination is introduced in this paper, where ROVE(Random Orthogonal vote based Evaluation) combination is applied to the results from the ASR systems created from each RP-based feature. Fig. 5 shows an overview of the vote-based combination method. Figure.6 shows an overview of vote – based combination method.

First, random matrices $\mathbf{P}^{(l)}$ $(l,1,L)$ are chosen from the standard normal distribution, with mean 0 and variance 1. Original speech features (MFCCs) are projected using each random matrix. An acoustic model corresponding to each random matrix is also trained.

For the test utterance, using each acoustic model, an ASR system outputs the best scoring word by itself. To obtain a single hypothesis from among all the results from the random projection, voting is performed by counting the number of occurance of the best word for each RP – based feature. For example, in the case of l = 20, 20 kinds of new feature vectors are calculated using 20 kinds of random matrices. Then, we train the 20 kinds of acoustic models using 20 kinds of new feature vectors.

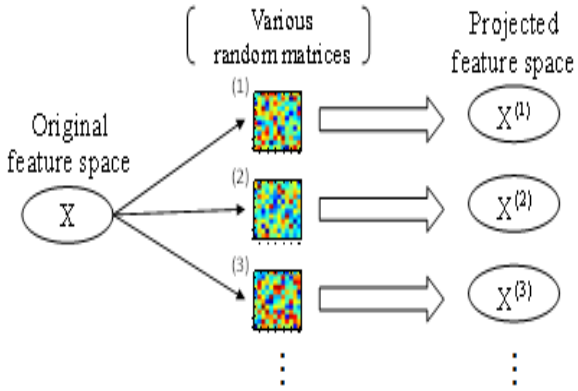


Fig.5. Random projection on the feature domain

It is little bit lengthier as takes more time. In the test process, 20 kinds of recognition results are obtained using 20 kinds of acoustic models based on pitch ,length and sound. To obtain a single hypothesis from among 20 kinds of recognition results, voting is performed as a selection process.

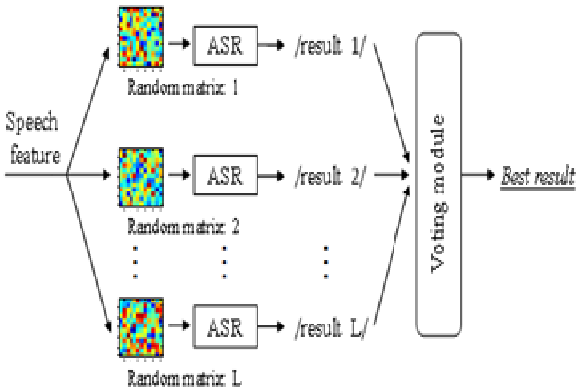


Fig.6.Overview of the vote-based combination

3.3. Automatic speech recognition

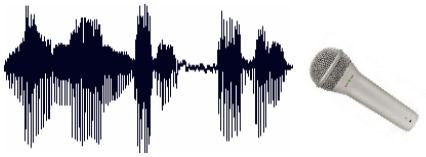
ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 1, February 2017

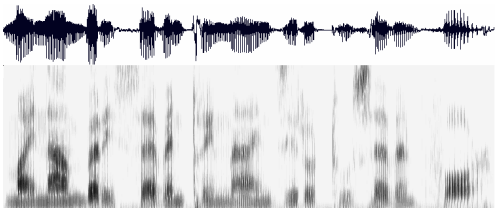
Articulation produces sound waves which the ear conveys to the brain for the purpose of processing. Digitization helps in Converting analogue signal into digital representation. Uses filter to measure energy levels for various points on the frequency spectrum. By,Knowing about the relative importance of different frequency bands (for speech) makes this process more efficient.E.g. high frequency sounds are less informative in nature, so should be sampled using a broader bandwidth (log scale).

Signal processing Separates speech from background noise. Noise cancelling microphones are used, two mics, one facing speaker, the other facing away.Ambient noise is roughly same for both the mics.Knowing which bits of the signal relate to speech Spectrograph analysis is carried out. Variation among speakers due to Vocal range (f0, and pitch range), Voice quality (growl, whisper, physiological elements such as nasality, adenoidality, etc)ACCENT !!! (especially vowel systems, but also consonants, allophones, etc.).

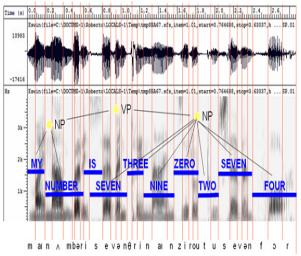
Variations within speakers are due to health emotional state. Speaker-dependent systems require "training" to "teach" the system your individual idiosyncracies and ambient conditions. User is asked to pronounce some key words which allow computer to infer details of the user's accent and voice. In Speaker-independent systems, language coverage is reduced to compensate need to be flexible in phoneme identification. Some ASR systems include a grammar which can help disambiguation.



AcousticWaveform



Acoustic Signal



Speech recognition

Fig.7.Automatic Speech recognition

Phonetics finds Variability in human speech. Phonology helps in recognizing individual sound distinctions (similar phonemes). Lexicology and syntax are disambiguating homophones and finds the features of continuous speech. Syntax and pragmatics interprets prosodic features in a best way.Pragmatics helps in filtering of the performance errors (disfluencies) .Discontinuous speech is much easier to

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 1, February 2017

recognize. Since,S ingle words tend to be pronounced more clearly.Continuous speech involves contextual co-articulation effects are Weak forms,Assimilation,Contractions.

3.4. Interpreting Prosodic Features

Pitch, length and loudness are used to indicate "stress". All of these are relative.On a speaker-by-speaker basis and in relation to context.Pitch and length are phonemic in some languages. Pitch contour can be extracted from speech signal.But pitch differences are relative. One man's high is another (wo)man's low.Pitch range is variable.Pitch contributes to intonation.But has other functions in tone languages.Intonation can convey meaning. Length is easy to measure but difficult to interpret Again, length is relative always.

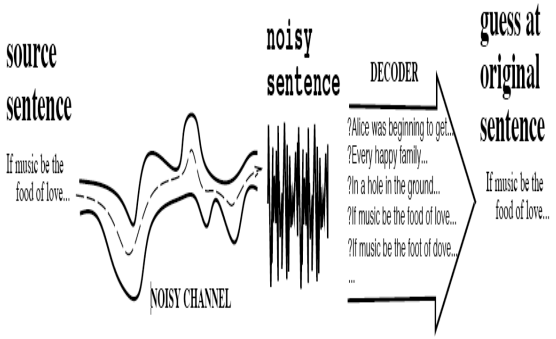


Fig.8.The Noisy Channel Model

It is phonemic in many languages.Speech rate is not constant – slows down at the end of a sentence. Loudness is easy to measure but difficult to interpret.Again, loudness is relative. It Identify individual phonemes, Identify words, Identify sentence structure and/or meaning and Interpret prosodic features (pitch, loudness, length). Performance "errors" include Non-speech sounds,Hesitations, False starts, and repetitions.Filtering implies handling at syntactic level or above.

Statistics-based approach Can be seen as an extension of the template-based approach, using more powerful mathematical and statistical tools.Sometimes seen as "anti-linguistic" approach. Collect a large corpus of transcribed speech recordings. Train the computer to learn the correspondences ("machine learning").At run time, apply statistical processes to search through the space of all possible solutions, and pick the statistically most likely one.

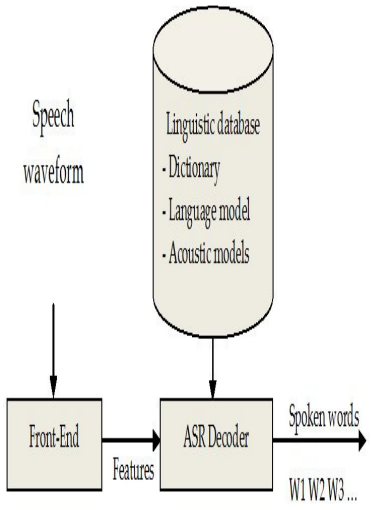


Fig.9.ASR Decoder

Search through space of all possible

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 1, February 2017

sentences.Pick the one that is most probable given the waveform. It Uses the acoustic model to give a set of likely phone sequences.Use of the lexical and language models to judge which of these are likely to result in probable word sequences.

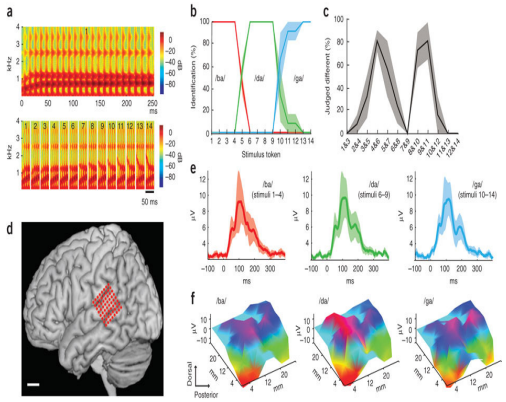


Fig.10 Overall evaluation of the signal

The trick is having sophisticated algorithms to juggle the statistics.A bit like the rule-based approach except that it is all learned automatically from data. The output of ASR is usually in the form of words. The Text – to – speech (TTS) API makes converting text-to-speech easier than ever. The Text – to- speech conversion occurs with the help of Digital- to- Analogue converter and they are finally filtered and amplified and stored in the computer as analogue signal through the sound card of the PC.

4. Evaluation

4.1 Experimental Conditions

The proposed method was evaluated on a word recognition task for three males suffering from articulation disorders namely (Speaker A, B, C). For the conducted experiments, we recorded 210 words including the ATR Japanese speech database from each speaker. Each of the 210 words was repeated five times (Fig. 10). The speech signal was sampled at 16 kHz and windowed with a 25-msec Hamming window for every 10 msec. Fig. 11 shows an example of a spectrogram spoken by a person with an articulation disorder due to cerebral plasy.

Fig.12 shows a spectrogram spoken by a physically unimpaired person but can speech with no distortions, doing the same task.

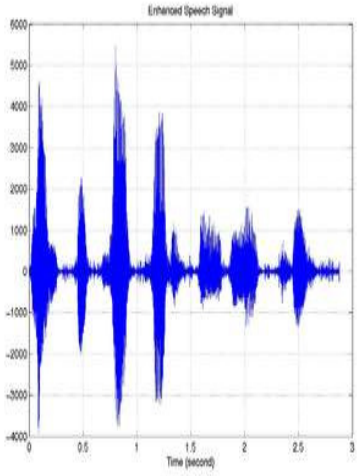


Fig.11. Recorded speech data

The recognition results for a speaker-independent model are shown in Fig.1.2, where the speech data uttered by unimpaired

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 1, February 2017

persons were used. As can be seen in the Fig. 12, the recognition rate of a physically unimpaired person was around 90%. However, the result of a person with an articulation disorder (speaker A) is only 3.5%,while comparing with the unimpaired It is clear that the speaking style of a person with an articulation disorder differs considerably from that of the physically unimpaired persons.

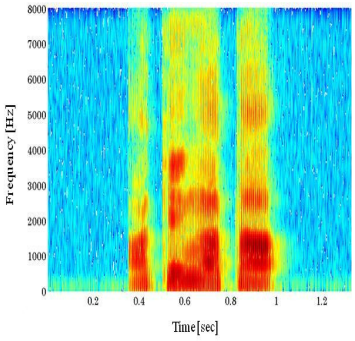


Fig.12.Spectrogram from a person with an articulation disorder**.**

Also, Fig. 12 shows that it was difficult to recognize the utterances of articulation disorders using an acoustic model trained by utterances of physically unimpaired persons. Therefore, in this paper, we trained the acoustic model using the utterances of a person with an articulation disorder. The acoustic model consists of a HMM set with 54 context-independent phonemes and 8 mixture components for each state in this model. Each HMM has three states and three self-loops.

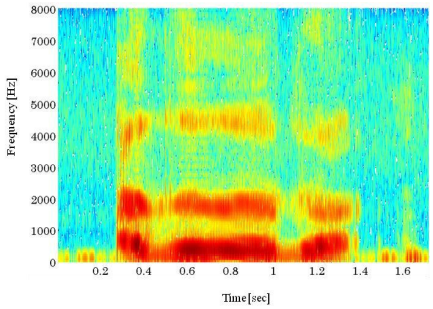


Fig.13. Spectrogram from a person with an articulation disorder**.**

4.2 Experiment on Articulation disorder

In Experiment 1, recognition results were obtained for each utterance using speaker-dependent models of 3 speakers. Each system was trained using 24-dimensional feature vectors consisting of 12-dimensional MFCC parameters, along with their delta parameters also. When we recognized the 1st utterance, the 2nd through 5th utterances were used for training the pulses. We iterated this process for each utterance Table 1 shows the results obtained in Experiment 1 for speaker (A). As can be seen in Table 1, the recognition rate for the 1st utterance was 75.7%. It was significantly lower when compared to other utterances.

TABLE 1.Recognition results[%]for each utterance of speaker(A)

1st	2nd	3rd	4th	5th

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

75.7	86.7	92.9	90.5	88.6

It is considered that the speaker experiences a more strained state during the first utterance compared to subsequent utterances because the first utterance is the first intentional movement in the speech. Therefore, athetoid symptoms occur and articulation becomes difficult here. It is believed that this difficulty causes fluctuations in speaking style and degradation of the recognition rates by omitting the important words.

TABLE 2.Recognition results[%]for each utterance of speaker(B)

1st	2nd	3rd	4th	5th
85.7	91.9	91.4	93.3	95.2

TABLE 3.Recognition results[%]for each utterance of speaker(c)

1st	2nd	3rd	4th	5th
85.7	91.9	91.4	93.3	95.2

Tables 2 and 3 show the recognition results form each utterance for speaker (B) and (C), respectively. As can be seen in Tables 2 and 3, a decrease in recognition rate for the first utterance due to fluctuations in speaking style was confined by evaluation.

4.3 Experiment on Un-impaired Persons

The aim of Experiment 2 is to evaluate the improvement introduced by the use of a RP-based feature projection method for the unstable 1st utterance due to plasy. In the experiments, the following RP-based features were evaluated. Random projection is applied to MFCC at the frame, t-th frame

$$\mathbf{Y(t) = P^{(l)T} x(t)}$$

Then, the new feature also has the delta parameter of original feature, x(t). The final system feature dimensionality is 24(MFCC→12RP→12→ΔMFCC).We have investigated the performance of random projections for various random matrices (l = 20, 40, 60, 80, and 100) sampled from $N(0,1)$.

TABLE 4.Word recognition rate[%]for 1st utterance of speaker(A) using the proposed method in various random matrices.(The recognition rate for the original features is 75.7%)

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

Number of random matrices	RP combination based on ROVER	RP w'o		
		Max	Mean	Min
20	80.5	80	76.8	73.8
40	81.5	80	76.8	72.9
60	81.9	80.5	77	72.9
80	81.4	80.5	76.9	72.9
100	81	80.5	76.8	72.9

Also, even if the number of random matrices is changed, we do not see that subsequent performance variations in our experiments.

Table 4 shows the performance results versus the number of random matrices for speaker (A) who participated in the experiment. As can be seen in Table 4, the results for RP-based feature indicate that the vote-based random-projection combination had improved the recognition rate from 75.7% to 81.9% using the combination of 60 random matrices, and not by using lower order matrices and even the means of random projections based mapping without combination for some random matrices was better than that of the recognition rate of the original features. Also, even if the number of random matrices is changed, we do not see that subsequent performance varies in our experiments.

TABLE 5.Word recognition rate[%]for 1st utterance of speaker(B) using the proposed method in various random matrices.(The recognition rate for the original features is 75.7%)

Number of random matrices	RP combination based on ROVER	RP w'o		
		Max	Mean	Min
20	80.5	80	76.8	73.8
40	81.5	80	76.8	72.9
60	81.9	80.5	77	72.9
80	81.4	80.5	76.9	72.9
100	81	80.5	76.8	72.9

Tables 5 and 6 show the performance of the proposed method for the speaker (B) and (C), respectively. As we can view in Table 5, the recognition rate of 90.5% was obtained using the combination of 60 or 80 random matrices. Also, the results in Table 6 were maintained above 95%, showing that the effectiveness of our proposed method for each person with articulation disorders, who participated in our experiment namely speaker A, Speaker B, Speaker C.

4.4 Experiment on PCA based method

In order to show the superiority of the RP-based feature projection method, in Experiment 3, we compared the proposed method and the PCA-based feature projection method as a result for getting highest percentage of accuracy in the recorded words.

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

In the Experiment 3, PCA was applied to 12-dimensional MFCC, and the new feature also had the delta coefficient of the MFCC features. Then, we have computed the eigen-vector matrix using the 2nd through 5th utterances of the words(the more stable utterances) for each speaker A,B and C. A comparison between the PCA-based feature projection method and the results obtained by our proposed method using the combination of 60 random matrices are used for comparison to choose the best method.

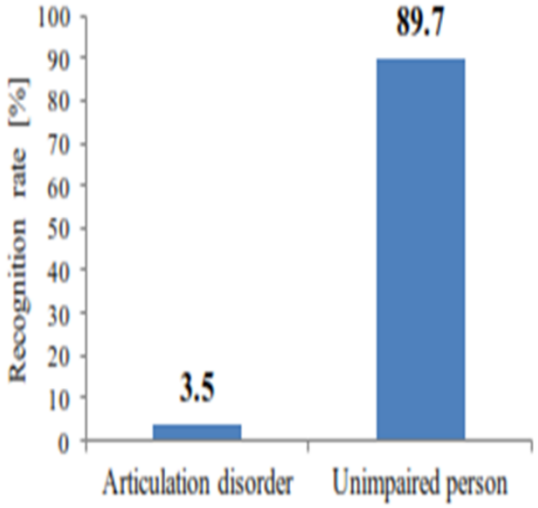


Fig.14 Recognition results[%] for the speaker-independent model using training data uttered by unimpaired persons.

We can see that the combination of random projection and ROVER output performs both the baseline method (MFCCs) and the PCA-based feature extraction method as the combinational process to get best results. This result gives the evidence of the improvements introduced by the speech feature extraction based on random projection and the use of ROVER for getting best results from the existing inputs.

Tables 4 ~ 6 show the recognition rate versus the number of random matrices for each speaker with different kinds of problems. The results of "RP w/o combination" show the maximums, means, and minimums obtained from each random projection without ROVER-based combination.

One of the possible reasons the random projection improves the recognition rates may be that if distributions of original data are skewed (have ellipsoidal contours of high eccentricity), their transformed counterparts will become more spherical as said by S. Dasgupta [17].

Table 6. Word recognition rate[%] for the 1st utterances of speaker (C)using the proposed method in various random matrices. (The recognition rate for the original features is 94.3%)

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

Number of random matrices	RP combination based on ROVER	RP w'o		
		Max	Mean	Min
20	96.2	97.1	95.3	93.8
40	95.7	97.1	95.3	93.8
60	96.7	97.6	95.3	93.8
80	96.7	97.6	95.3	93.8
100	95.7	97.6	95.3	93.8

However, there were 'bad' projections that cause degradation of speech recognition accuracy rates compared with the recognition of original features (Tables 4 ~ 6). Therefore, more research will be needed to investigate the effectiveness of the random projection method for speech features as well.

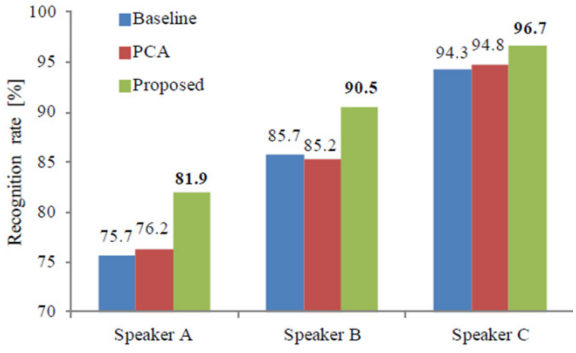


Fig.15. Comparison between the PCA-based feature projection method and the results obtained by our proposed method for the 1st utterance.

5. RESULTS AND DISCUSSON

The original signal is stored in the PC with help of secondary storage devices and these signals are highly distorted. The Fig.15 is the signal obtained from the articulation patient, who spelled a word fifty-fifty.

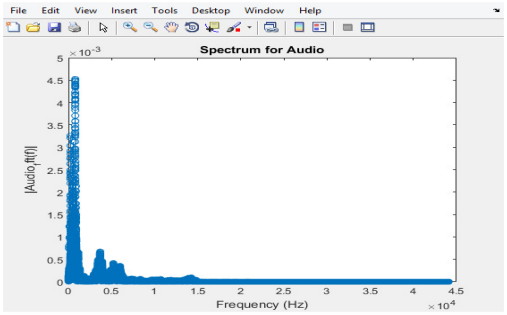


Fig.16 Spectrum of the original signal

From the signal, we are able to analyse that, their first speech tends to be more and more distorted while comparing with the other utterance. It is the MATLAB output obtained from the recorded speech.

The recoded speech is found to be dis- continues in time and it is difficult to be understood.Then, by the ROVER based combination the best result is obtained and the feature of the result is obtained using the Automatic Speech Recognition (ASR).Fig. 16 shows the feature of the signal and the clear signals information is displayed along with the spectrum.

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 1, February 2017

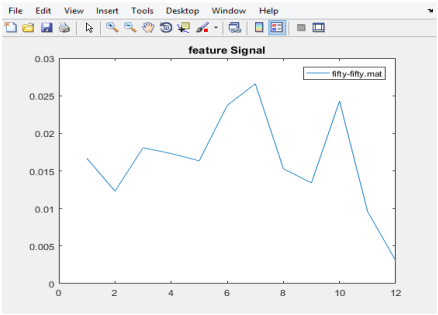


Fig.17 Feature of the signal

The final output can be obtained from the speech- to-text converter in the form of clear speech.

8. Conclusions

As a result of this work, a method for recognizing dysarthric speech using RP-based features has been developed as the best one. The proposed method transforms the conventional speech features such as MFCC using various random matrices by using the random projection. It also introduces the vote-based combination method to obtain an optimal result from the ASR systems created from each RP -based feature. Word recognition experiments were conducted to evaluate the proposed method for three males with articulation disorders. The results of the experiments showed that all the recognition rates of the proposed method outperformed the baseline rate (using MFCCs).

REFERENCES

[1] J. Lin, W. Ying, and T. S. Huang"Capturing human hand motion in image sequences," IEEE Motion and Video Computing Workshop, pp. 99-104, 2002.

[2] T. Starner, J. Weaver, and A. Pentland, "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video," IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(12), pp. 1371-1375, 1998.

[3] G. Fang, W. Gao and D. Zhao, "Large vocabulary sign language recognition based on hierarchical decision trees," Proceedings of the 5th international conference on Multimodal interfaces, pp. 125-131, 2003.

[4] Christo Ananth, S.Santhana Priya, S. Manisha, T.Ezhil Jothi, M.S.Ramasubhaeswari, "CLG for Automatic Image Segmentation", International Journal of Electrical and Electronics Research (IJEER), Vol. 2, Issue 3, Month: July - September 2014, pp: 51-57

[5] V. Wu, R. Manmatha and E. M. Riseman, "Textfinder: an automatic system to detect and recognize text images," IEEE Transactions on Pattern Analysis and Machine Inteligence, 21(11), pp. 1224-1229, 1999.

[6] M. K. Bashar, T. Matsumoto, Y. Takeuchi, H. Kudo, and N. Ohnishi, "Unsupervised Texture Segmentation viaWavelet-based Locally Orderless Images (WLOIs) and SOM," 6th IASTED International Conference COMPUTER GRAPHICS AND IMAGING, 2003.

[7] T. Ohsuga, Y. Horiuchi, and A. Ichikawa, "Estimating Syntactic Structure from Prosody in Japanese Speech," IEICE Transactions on Information and S

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

ystems, 86(3), pp. 558-564, 2003.

[8] K. Nakamura, T. Toda, H. Saruwatari, and K. Shikano, "Speaking Aid System for Total Laryngectomees Using Voice Conversion of Body Transmitted Artificial Speech," INTERSPEECH, pp. 1395-1398, 2006.

[9] S.T. Canale, and W.C. Campbell, "Campbell's Operative Orthopaedics," Mosby-Year Book, 2002.

[10] H. Matsumasa, T. Takiguchi, Y. Ariki, I. Li, and T. Nakabayashi, "PCA-Based Feature Extraction for Fluctuation in Speaking Style of Articulation Disorders," INTERSPEECH 2007, pp. 1565-1568, 2007.

[11] C. Miyamoto, Y. Komai, T. Takiguchi, Y. Ariki, and I. Li, "Multimodal Speech Recognition of a Person with Articulation Disorders Using AAM and MAF," 2010 IEEE International Workshop on Multimedia Signal Processing, pp. 517-520, 2010.

[12] Ella Bingham, and HeikkiMannila, "Random projection in dimensionality reduction: applications to image and text data," KDD'01 Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 245-250, 2001.

[13] N. Goel, G. Bebis, and A. Nefian, "Face Recognition Experiments with Random Projection," SPIE, vol. 5779, pp. 426-437, 2005.

[14] P. Thaper, S. Guha, and N. Koudas, "DynamicMultidimensional Histograms," ACM SIGMOD, pp. 428-439, 2002.

[15] L. Liu, P. Fieguth, G. Kuang, and H. Zha, "Sorted Random Projections for robust texture classification," IEEE International Conference on Computer Vision, pp. 391-398, 2011.

[16] H. T. Ho, and R. Chellappa, "Automatic head pose estimation using randomly projected dense SIFT descriptors," IEEE International Conference on Image Processing, pp. 153-156, 2012.

[17] S. Dasgupta, "Experiments with random projection," UAI, pp. 143-151, 2000.

[18] X. Z. Fern, and C. E. Brodley, "Random Projection for High Dimensional Data Clustering: A Cluster Ensemble Approach," the 20th Int. Conf. on Machine Learning, pp.186-193,2003

[19] S. Lee, and A. Nedic, "Distributed Random Projection Algorithm for Convex Optimization," IEEE Journal of Selected Topics in Signal Processing, Vol. 7, No. 2, pp. 221-229, 2013.

[20] J. G. Fiscus, "A post-processing system to yield reduced word error rates: Recogniser output voting error reduction (ROVER)," IEEE ASRU, pp. 347-352, 1997.

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering
(IJARMATE) Vol. 3, Special Issue 1, February 2017

[21] R. I. Arriaga, and S. Vempala, "An alg orithmic theory of learning: robust con cepts and random projection," IEEE S ymposium on Foundations of Comput er Science, pp. 616-623, 1999.

.