TEXT-INDEPENDENT SPEAKER RECOGNITION USING COMBINED LPC AND MFC COEFFICIENTS

PPS Subhashini[1], Turimerla Pratap[2]

[1]*Associate Prof, ECE Dept., RVR & JC College of Engineering*
[2]*ECE Dept, RVR & JC College of Engineering, Guntur*

Abstract

There are so many techniques for text- independent speaker recognition. However this text-independent speaker recognition is very difficult because the recognition is performed irrespective of what one he is saying. This paper presents a very simple approach to text independent recognition where the recognition is performed by using both LPC and MFC coefficients in parallel and the results of both methods are combined for best matching of the speaker. Here the ANN is used for classification.

Keywords: *LPC-linear predictive coding, MFC- Mel scale frequency cepstrum, ANN- artificial neural network*

--***--

1. INTRODUCTION

Speaker recognition is the process of identifying a person on the basis of speech alone. It is a known fact that speech is a speaker dependent feature that enables us to recognize friends over the phone. During the years ahead, it is hoped that speaker recognition will make it possible to verify the identity of persons accessing systems; allow automated control of services by voice, such as banking transactions; and also control the flow of private and confidential data. While fingerprints and retinal scans are more reliable means of identification, speech can be seen as a non-evasive biometric that can be collected with or without the person’s knowledge or even transmitted over long distances via telephone. Unlike other forms of identification, such as passwords or keys, a person's voice cannot be stolen, forgotten or lost.

Speech is a complicated signal produced as a result of several transformations occurring at several different levels: semantic, linguistic, articulatory, and acoustic. Differences in these transformations appear as differences in the acoustic properties of the speech signal. Speaker-related differences are a result of a combination of anatomical differences inherent in the vocal tract and the learned speaking habits of different individuals. In speaker recognition, all these differences can be used to discriminate between speakers.[1]

In this paper the speaker features are extracted by using LPC and MFC methods and these features are trained and tested by using back propagation neural network. The detailed flowchart of this proposed method is shown in Fig.1

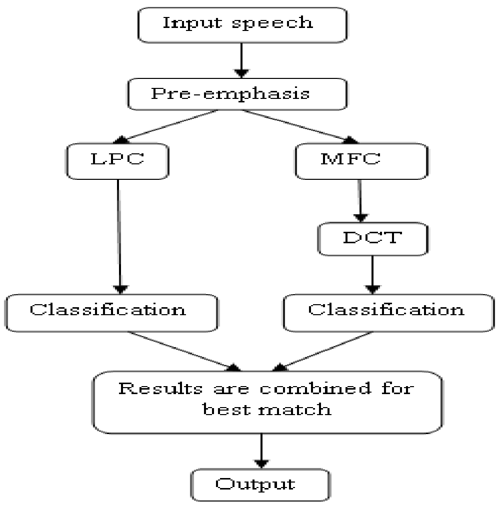

Fig 1 text independent speaker recognition using combined LPC and MFC coefficients

The procedure to find LPC and MFC coefficients are explained in the following chapters.

2. FEATURE EXTRACTION

In this paper two feature extraction techniques are used, one is based on linear predictive coding and another is on Mel-scale frequency coefficients.

2.1 Linear Predictive Coding

Linear predictive coding (LPC) is a tool used mostly in audio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model. It is one of the most powerful speech analysis techniques, and one of the most useful methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters.

2.1.1 Envelope Calculation

The LPC method is quite close to the FFT. The envelope is calculated from a number of formants or poles specified by the user. The formants are estimated removing their effects from the speech signal, and estimating the intensity and frequency of the remaining buzz. The removing process is called inverse filtering, and the remaining signal is called the residue.

2.1.2 LPC Analysis

In this analysis first convert each frame of p+1 autocorrelations into LPC parameter set by using Durbin's method. This can formally be given as the following algorithm

$$E^{(0)} = r(0) \qquad (1)$$

$$k_i = \frac{r(i) - \sum_{j=1}^{l-1} r(|i-j|)}{E^{i-1}} \quad 1 \le i \le p \qquad (2)$$

$$\alpha_i^{(i)} = k_i \qquad (3)$$

$$\alpha_i^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \quad 1 \le j \le i-1 \qquad (4)$$

$$E^{(i)}(1-k_i^2)E^{i-1} \qquad (5)$$

By solving above equations recursively for i=1,2,…..p, the LPC coefficient m is given as

$$\alpha_m = \alpha_m^{(p)} \qquad (6)$$

These LPC coefficients are further statistically analyzed.

2.2 MFCC

MFCCs are based on the known variation of the human ear's critical bandwidths with frequency; filters spaced linearly at low frequencies and logarithmically at high frequencies have been used to capture the phonetically important characteristics of speech. This is expressed in the Mel-frequency scale, which is linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. Here the Mel scale is being used which translates regular frequencies to a scale that is more appropriate for speech, since the human ear perceives sound in a nonlinear manner. Our whole understanding of speech is through our ears only so it is very useful method.

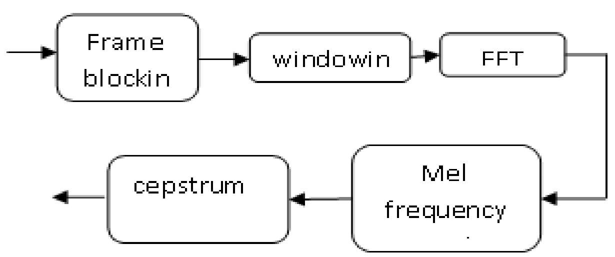

Fig 2 MFCC processor

MFCCs are commonly derived as follows:

1. Take the Fourier transform of (a windowed excerpt of) a signal.
2. Map the powers of the spectrum obtained above onto the Mel scale, using rectangular window.
3. Take the log of the powers at each of the mel frequencies.
4. Take the discrete cosine transform of the list of mel log powers.

The MFCCs are the amplitudes of the resulting spectrum.

3. ARTIFICIAL NEURAL NETWORKS

Neural network or Artificial Neural Network (ANN) is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. A neural network contains a large number of simple neuron like processing elements and a large number of weighted connections encode the knowledge of a network. Though biologically inspired, many of the neural network models developed to duplicate the operation of the human brain.

3.1 Back Propagation Training

Back propagation is a common method of training artificial neural networks used in conjunction with an optimization method such as gradient descent. The method calculates the gradient of a loss function with respects to all the weights in the network. The gradient is fed to the optimization method which in turn uses it to update the weights, in an attempt to minimize the loss function.

Back propagation usually considered to be a supervised learning method. It is a generalization of the delta rule to multi-layered feed forward networks, made possible by using the chain rule to iteratively compute gradients for each layer. Back propagation requires that the activation function used by the artificial neurons (or "nodes") be differentiable.

The back propagation learning algorithm can be divided into two phases: propagation and weight update.

Phase 1: Propagation

Each propagation involves the following steps:

1. Forward propagation of a training pattern's input through the neural network in order to generate the propagation's output activations.
2. Backward propagation of the propagation's output activations through the neural network using the training pattern target in order to generate the deltas of all output and hidden neurons.

Phase 2: Weight Update

For each weight-synapse follow the following steps:

1. Multiply its output delta and input activation to get the gradient of the weight.
2. Subtract a ratio (percentage) of the gradient from the weight.

This ratio (percentage) influences the speed and quality of learning; it is called the learning rate. The greater the ratio, the faster the neuron trains; the lower the ratio, the more accurate the training is. The sign of the gradient of a weight indicates where the error is increasing; this is why the weight must be updated in the opposite direction.

Repeat phase 1 and 2 until the performance of the network is satisfactory.

Let I be the Number of input nodes, J be of Number of hidden nodes and K be the Number of output nodes for the neural network as shown in fig 3. Consider V be the weight vector for hidden layer and W be the weight vector for output layer. The size of W matrix is K X J. The size of V matrix is J X I.

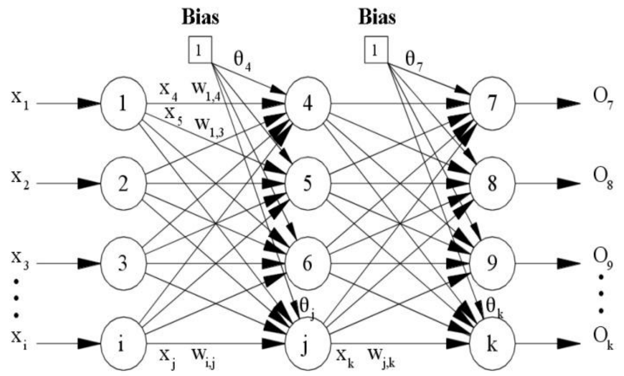

Fig 3 artificial neural network with single hidden layer

3.2 The Steps for the Training Cycle

1. By applying the feature vectors one by one to the input layer the output of hidden layer is computed as

$$y_j = f_1\left(v_j^t z\right) for\ j = 1,2, \ldots, \mathrm{K} \quad (7)$$

2. The function $f1(.)$ is unipolar Sigmoid Function. Output of output layer is computed by using (8)

$$O_k = f_2(w_k^t y)\ for\ k = 1,2, \ldots K \quad (8)$$

The function *f2*(.) is Generalized Sigmoid Function defined by using (8)

3. The error value is computed by using (9)

$$E = \frac{1}{2}(d_k - o_k)^2 + E, for\ k = 1,2, \ldots., K \quad (9)$$

4. Error signal vectors δ_o and δ_{oy} of both layers are computed. Dimension of Vector δ_o is (K X 1) and Dimension of δ_y is (J X 1). The error signal terms of output layer is given by using

$$\delta_{ok} = (d_k - o_k)(1 - o_k)o_k, \quad (10)$$

$$for\ k = 1,2, \ldots. K$$

5. The error signal terms of hidden layer is given by using (11)

$$\delta_{yj} = y_j\left(1 - y_j\right)\sum_{k=1}^{K} \delta_{ok} w_{kj},$$
$$for\ j = 1,2, \ldots J \quad (11)$$

6. The output layer weights are adjusted as

$$w_{kj} = w_{kj} + \eta\delta_{ok} y_j, \quad (12)$$
$$for\ k = 1,2, \ldots K\ and j = 1,2, \ldots. J$$

7. The hidden layer weights are adjusted by using (13)

$$v_{ji} = v_{ji} + \eta\delta_{yj} z_i,$$
$$for\ j = 1,2, \ldots J\ and \quad (13)$$
$$i = 1,2, \ldots. I$$

8. Repeat the steps 1 to 5 for all the feature vectors

9. The training cycle is repeated for 1000 epochs.

4. PROPOSED METHOD

The LPC and MFCC are both very useful methods in speech and speaker recognition process. And they have almost 100 percentage of correct classification in text-dependent speaker recognition. But in case of text-independent speaker recognition these methods are failed to produce good results. So in this paper the combined LPC and MFC methods are implemented, and this method yields at least 98 percent of correct classification.

The steps for proposed method is as follows

1. Consider input speech signal.
2. Find LPC coefficients.
3. Find MFC coefficients.
4. Apply these LPC and MFC coefficients to the back propagation network individually and parallely.
5. The corresponding results of these two are averaged to get single result.
6. This result in the step 5 declares the best match.

5. EXPERIMENTAL RESULTS

The proposed method is implemented with the10 speakers and each speaker was given the 5 utterances each. In these 5 utterances 3 utterances are used for training and remaining 2 for testing.

So here the neural network is designed to classify 10 speakers. So it has 10 input nodes and 10 output nodes. This neural network is trained with LPC and MFC coefficients using back propagation algorithm separately and parallely. The training data and the desired output data for the first 5 speakers is mentioned in the following Table 1 and Table 2.

In the following table the speakers are represented in the format **Si_j.wav** here i represents the speaker and j represents the speaker utterance.

Table 1 Training data (LPC coefficients) for first 5 speakers

Speaker	**LPC coefficients for training**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
S1_1.wav	1	-1.227	1.291	-1.284	1.409	-1.069	0.806	-0.532	0.470	-0.100
S1_2.wav	1	-1.349	1.436	-1.434	1.525	-1.159	0.886	-0.615	0.507	-0.121
S1_3.wav	1	-1.308	1.362	-1.381	1.477	-1.071	0.740	-0.470	0.382	-0.038
S2_1.wav	1	-1.648	1.591	-1.761	1.623	-1.039	0.632	-0.252	0.069	0.022
S2_2.wav	1	-1.776	1.971	-2.135	2.035	-1.566	1.062	-0.504	0.219	-0.042
S2_3.wav	1	-1.660	1.798	-2.056	1.944	-1.466	1.056	-0.514	0.218	-0.047
S3_1.wav	1	-0.849	1.009	-1.074	0.762	-0.466	0.480	-0.258	0.285	0.019
S3_2.wav	1	-0.903	1.035	-1.114	0.743	-0.379	0.412	-0.154	0.199	0.010
S3_3.wav	1	-0.978	1.105	-1.157	0.735	-0.343	0.347	-0.148	0.191	0.003
S4_1.wav	1	-1.878	2.521	-2.999	3.037	-2.482	1.911	-1.163	0.526	-0.031
S4_2.wav	1	-1.934	2.603	-3.052	3.025	-2.334	1.656	-0.902	0.349	0.046
S4_3.wav	1	-2.079	2.748	-3.093	2.886	-2.088	1.415	-0.761	0.286	0.017
S5_1.wav	1	-2.196	2.948	-2.814	1.884	-0.808	0.224	0.018	0.047	-0.032
S5_2.wav	1	-2.263	3.150	-3.258	2.540	-1.444	0.653	-0.166	0.056	-0.009
S5_3.wav	1	-2.286	3.206	-3.340	2.678	-1.580	0.687	-0.134	0.0246	0.011

Table 2 Training data (MFC coefficients) for first 5 speakers

speaker	**MFC coefficients for training**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
S1_1.wav	4.191	-1.602	1.046	-1.805	0.573	-0.392	-0.381	-0.568	0.000	-0.116
S1_2.wav	4.762	-1.600	1.166	-1.802	0.313	-0.319	-0.284	-0.682	0.046	-0.064
S1_3.wav	4.595	-1.423	1.257	-1.879	0.457	-0.136	-0.384	-0.564	-0.141	0.031
S2_1.wav	6.393	-0.012	3.229	-0.876	0.205	-0.199	-0.496	0.096	-0.167	-0.229
S2_2.wav	6.666	-0.978	2.586	-1.035	0.466	-0.425	-0.553	0.117	-0.069	0.011
S2_3.wav	6.083	-0.851	3.101	-0.974	0.334	-0.599	-0.489	0.099	-0.069	0.103
S3_1.wav	2.228	-1.884	2.734	-0.612	-0.735	0.432	-0.180	-1.028	-0.065	0.020
S3_2.wav	2.635	-1.785	2.800	-0.560	-1.019	0.264	-0.280	-0.688	-0.078	-0.002
S3_3.wav	2.898	-2.026	2.883	-0.342	-1.214	0.587	-0.141	-0.820	-0.039	-0.136
S4_1.wav	6.873	-2.597	3.424	-1.584	-0.047	-0.841	0.308	-0.458	-0.630	-0.272
S4_2.wav	7.432	-2.416	3.36	-1.428	-0.277	-0.535	-0.064	-0.557	-0.686	-0.302
S4_3.wav	8.307	-2.089	3.130	-0.744	-0.638	-0.579	-0.042	-0.466	-0.463	-0.064
S5_1.wav	8.799	-3.247	1.220	0.885	-0.836	-0.432	-0.133	-0.560	0.384	0.029
S5_2.wav	8.902	-3.145	1.617	0.416	-0.866	-0.360	-0.086	-0.526	0.376	-0.121
S5_3.wav	9.005	-3.162	1.566	0.074	-0.606	-0.048	-0.236	-0.587	0.327	-0.102

Table 3 desired data for both LPC and MFC methods

Speaker	**Desired output at output node of a neural network**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
S1_1.wav	1	0	0	0	0	0	0	0	0	0
S1_2.wav	1	0	0	0	0	0	0	0	0	0
S1_3.wav	1	0	0	0	0	0	0	0	0	0
S2_1.wav	0	1	0	0	0	0	0	0	0	0
S2_2.wav	0	1	0	0	0	0	0	0	0	0
S2_3.wav	0	1	0	0	0	0	0	0	0	0
S3_1.wav	0	0	1	0	0	0	0	0	0	0
S3_2.wav	0	0	1	0	0	0	0	0	0	0
S3_3.wav	0	0	1	0	0	0	0	0	0	0
S4_1.wav	0	0	0	1	0	0	0	0	0	0
S4_2.wav	0	0	0	1	0	0	0	0	0	0
S4_3.wav	0	0	0	1	0	0	0	0	0	0
S5_1.wav	0	0	0	0	1	0	0	0	0	0
S5_2.wav	0	0	0	0	1	0	0	0	0	0
S5_3.wav	0	0	0	0	1	0	0	0	0	0

After training the neural network with the training data(LPC and MFC coefficients) towards the desired vector, this trained neural network is used for testing. But in the text-independent speaker recognition the testing is performed with the utterances which are not there in training. The untrained speaker utterances for both LPC and MFC methods are mentioned in the Table 4 and Table 5.

Table 4 Testing data (LPC coefficients) for first 5 speakers

Speaker	**LPC Testing data**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
S1_4.wav	1	-1.286	1.122	-1.015	1.0753	-0.840	0.695	-0.502	0.422	-0.080
S1_5.wav	1	-1.264	1.299	-1.243	1.312	-0.973	0.764	-0.614	0.495	-0.093
S2_4.wav	1	-1.727	1.827	-2.036	1.932	-1.387	0.863	-0.374	0.161	-0.027
S2_5.wav	1	-1.759	1.873	-2.046	1.912	-1.383	0.855	-0.345	0.150	-0.033
S3_4.wav	1	-0.908	1.097	-1.193	0.779	-0.415	0.404	-0.092	0.193	0.004
S3_5.wav	1	-0.594	0.920	-1.174	0.453	-0.468	0.521	-0.020	0.290	-0.017
S4_4.wav	1	-1.829	2.323	-2.512	2.411	-1.908	1.457	-0.868	0.401	0.003
S4_5.wav	1	-1.752	2.413	-2.807	2.851	-2.419	1.984	-1.180	0.584	-0.065
S5_4.wav	1	-2.291	3.139	-3.297	2.716	-1.683	0.839	-0.281	0.072	0.012
S5_5.wav	1	-2.265	3.148	-3.253	2.573	-1.481	0.660	-0.121	0.008	0.003

Table 5 Testing data (MFC coefficients) for first 5 speakers.

Speaker	**MFCC Testing data**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
S1_4.wav	4.834	-0.520	0.861	-1.158	0.284	-0.487	-0.105	-0.603	-0.007	0.007
S1_5.wav	4.326	-1.510	0.947	-1.654	0.327	0.183	0.056	-0.815	0.136	0.312
S2_4.wav	6.481	-0.610	3.084	-1.113	0.450	-0.190	-0.647	0.212	0.035	-0.048
S2_5.wav	6.683	-0.632	2.985	-0.912	0.485	-0.206	-0.753	0.237	0.101	0.047
S3_4.wav	2.645	-1.961	3.022	-0.479	-0.986	0.227	-0.494	-0.678	0.025	0.009
S3_5.wav	1.085	-1.817	4.401	0.231	-0.884	0.227	-0.345	-0.875	-0.184	-0.176
S4_4.wav	6.944	-2.364	2.263	-1.174	0.090	-0.836	0.248	-0.623	-0.393	-0.229
S4_5.wav	6.250	-3.051	2.897	-1.599	0.062	-1.423	0.006	-0.457	-0.509	0.098
S5_4.wav	9.064	-2.622	1.924	-0.060	-0.624	-0.145	-0.011	-0.451	0.222	-0.115
S5_5.wav	8.986	-3.061	1.561	0.139	-0.862	-0.450	-0.260	-0.499	0.386	-0.112

The output for the corresponding LPC and MFC methods are collected at the final stage and averaged to get the new output, as shown in Table 6

Table 6 Outputs of trained network for first 5 speakers

Speaker	**output**	**Result at output nodes for the test data**										**Estimation**
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	
S1_4.wav	LPC	**0.725**	0.020	-.382	0.015	-0.00	-0.26	0.023	0.412	0.582	0.036	Correct
	MFC	**0.634**	0.101	0.220	0.202	0.134	0.145	0.224	0.101	0.389	0.242	Correct
	COMBINED	**0.679**	0.061	-0.08	0.109	0.066	-0.05	0.123	0.256	0.486	0.139	Correct
S1_5.wav	LPC	0.613	-0.166	-0.20	0.101	0.106	-0.24	0.004	0.235	**0.725**	-0.01	**Wrong**
	MFC	**0.978**	-0.642	-0.46	-0.50	0.139	0.418	-0.33	-0.20	0.135	-0.19	Correct
	COMBINED	**0.795**	-0.404	-0.33	-0.20	0.122	0.084	-0.16	0.015	0.430	-0.10	Correct
S2_4.wav	LPC	-0.01	**0.838**	0.107	-0.08	0.063	0.030	0.007	0.013	-0.17	0.073	Correct
	MFC	-0.17	**0.982**	-0.26	-0.28	0.054	-0.18	-0.01	0.002	0.416	-0.28	Correct
	COMBINED	-0.09	**0.910**	-0.08	-0.18	0.058	-0.07	-0.00	0.007	0.118	-0.10	Correct
S2_5.wav	LPC	-0.06	0.620	**0.653**	-0.15	0.102	0.057	0.005	0.043	-0.27	0.138	**Wrong**
	MFC	-0.04	**0.979**	-0.49	-0.46	0.061	-0.43	0.266	0.009	0.538	-0.39	Correct
	COMBINED	-0.05	**0.797**	0.081	-0.31	0.081	-0.18	0.136	0.026	0.133	-0.12	Correct
S3_4.wav	LPC	0.180	0.014	**0.864**	-0.03	-0.04	0.173	-0.03	-0.18	-0.21	0.025	Correct
	MFC	0.228	0.238	**0.947**	-0.04	0.063	0.112	-0.01	-0.05	-0.13	0.045	Correct
	COMBINED	0.204	0.126	**0.905**	-0.03	0.010	0.143	-0.02	-0.11	-0.17	0.035	Correct
S3_5.wav	LPC	-0.07	-0.156	**0.995**	-0.06	-0.17	-0.02	0.078	-0.14	-0.78	0.302	Correct
	MFC	0.850	0.0078	0.847	-0.24	-0.26	-0.68	-0.50	**0.87**	-0.36	-0.28	**Wrong**
	COMBINED	0.387	-0.074	**0.921**	-0.15	-0.22	-0.35	-0.21	0.36	-0.57	0.008	Correct
S4_4.wav	LPC	0.168	0.061	-0.27	**0.838**	-0.01	-0.00	0.00	-0.13	0.369	-0.11	Correct

	MFC	0.047	0.110	-0.18	**0.788**	0.523	-0.07	0.149	0.128	0.508	0.170	Correct
	COMBINED	0.107	0.085	-0.22	**0.813**	0.255	-0.03	0.078	-0.00	0.438	0.029	Correct
S4_5.wav	LPC	0.451	0.162	0.048	**0.837**	-0.07	-0.21	-0.11	-0.05	-0.07	0.041	Correct
	MFC	**0.814**	0.226	-0.04	0.699	-0.20	0.39	-0.50	-0.20	-0.15	-0.20	**Wrong**
	COMBINED	0.632	0.194	0.001	**0.768**	-0.13	0.090	-0.306	-0.128	-0.116	-0.081	Correct
S5_4.wav	LPC	-0.037	0.117	0.016	0.271	0.613	0.003	-0.039	0.065	-0.087	**0.732**	**Wrong**
	MFC	-0.397	-0.002	-0.302	0.227	**0.932**	0.491	0.032	0.004	0.008	-0.058	Correct
	COMBINED	-0.217	0.057	-0.142	0.249	**0.772**	0.247	-0.003	0.035	-0.039	0.337	Correct
S5_5.wav	LPC	-0.014	0.083	-0.000	-0.047	**0.894**	-0.073	0.044	0.032	0.080	0.014	Correct
	MFC	0.236	0.035	-0.067	0.059	**0.945**	0.221	-0.137	-0.058	-0.131	0.001	Correct
	COMBINED	0.111	0.059	-0.033	0.005	**0.919**	0.073	-0.04	-0.019	-0.025	0.007	Correct

In Table 6 the bold numericals represents the maximum values in the ouput of neural network, which are used in the selection of a speaker. it was observed that the LPC and MFC methods fails to produce the 100 percent correct recognition but the combined method produced the 100 percent correct estimation.

The proposed method is implemented with the speech database that is collected from telephone based speaker identification dataset from Indians.[4]

The percentage of correct classification is calculated by varying number of hidden nodes in the network and these observations are tabulated in Table 7.

Table 7 percentage of correct classification

No. of hidden nodes	Percentage of correct classification		
	LPC	MFCC	Combined LPC and MFCC method
5	2.5	62.5	82.5
10	7.5	87.5	93.5
15	85	92.5	95
20	7.5	95	97.5
21	7.5	95	97.5
22	90	97.5	100
23	3.5	97.5	100
24	97.5	97.5	100
25	95	87.5	100

The untrained speaker utterances are applied in random to the trained network and the corresponding output is observed and noticed that this method gives 100 percentage of correct classification by choosing proper number of hidden nodes in the neural network.

Table 8 percentage of wrong classification

No. of hidden nodes	Percentage of wrong classification		
	LPC	MFCC	Combined LPC and MFCC method
5	27.5	37.5	17.5
10	12.5	12.5	7.5
15	15	7.5	5
20	12.5	5	2.5
21	12.5	5	2.5
22	10	2.5	0
23	7.5	2.5	0
24	2.5	2.5	0
25	5	12.5	0

6. CONCLUSIONS

The text-independent speaker recognition is very difficult compared with the text- dependent speaker recognition because here the testing is performed with the new inputs which are not there in training. So the new methods are necessary and the present study is still on-going. The MFC method is widely used in text-dependent speaker recognition due to its mel-scale coefficients. But the MFC alone is failed to give the good result in case of text-independent speaker recognition. The use of these basic techniques(LPC, MFCC) is very easy so the modifications to the basic techniques(LPC, MFCC) along with some new classification algorithms is needed which may give the 100 percent of correct classification.

REFERENCES

[1] Rabiner LR, juang BH, “Fundamentals of speech recognition”, Prentice Hall India

[2] Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman “SPEAKER IDENTIFICATION USING MEL FREQUENCY CEPSTRAL COEFFICIENTS” 3rd International

Conference on Electrical & Computer Engineering ICECE, December 2004

[3] R. B. Shinde and V. P. Pawar “Fusion of MFCC & LPC Feature Sets for Accurate Speaker Identification” International Journal of Current Engineering and Technology ISSN 2277 - 4106

[4] Azarias Reda, Saurabh Panjwani and Edward Cutrell “A Low-cost Remote Attendance Tracking System for Developing Regions”

[5] Tomi Kinnunen, Haizhou Lib “ An Overview of Text-Independent Speaker Recognition: from Features to Supervectors ”

[6] Sajid A. Marhon “Speaker Recognition Based On Neural Networks”

[7] Srinivasan, "Speaker Identification and verification using Vector Quantization and Mel frequency Cepstral Coefficients", Research Journal of Applied Sciences, Engineering and Technology 4(I):33-40, 2012.

[8] Vibha Tiwari, "MFCC and its applications in speaker recognition", International Journal on Emerging Technologies1(I):19-22(2010)