Speaker Recognition

Arindam Bhattacharyya, University of California, Davis abhattacharyya@ucdavis.edu
Bhawna Sinha, University of California, Davis bhasinha@ucdavis.edu

Abstract

Speaker recognition, a significant technique in the area of digital signal processing is used in a wide range of applications such as security systems, forensics, and human-computer interaction. This project focuses on implementing speaker recognition using MATLAB, employing Mel-frequency Cepstral Coefficients (MFCC), Mel-filterbank (MELFB), and Linde-Buzo-Gray (LBG) algorithm. The proposed system begins with preprocessing the speech signals to extract audio features using the MFCC technique, which mimics the human auditory system's response to sound. Subsequently, the Mel-filterbank enhances discrimination of the extracted features by modeling the frequency response of the voice speech. The MATLAB implementation provides a user-friendly interface for both training and testing the speaker recognition system. Experimental results demonstrate the effectiveness of the proposed approach in accurately identifying speakers from a given dataset.

Keywords: Speaker recognition, MATLAB, MFCC, Mel-filterbank, LBG algorithm, Feature extraction, Clustering.

1. Introduction

Speaker recognition is the process of finding the identity of an unknown speaker by comparing his/her voice with voices of registered speakers in the database. It's a one-to-many comparison.

Speaker recognition can be classified into identification and verification. Speaker identification is the process of determining which registered speaker provides a given voice sample regardless of what is saying. On the other hand, Speaker verification is the process of accepting or rejecting the identity claim of a speaker. In this paper, we are going to implement speaker identification model. Basic structure of speaker identification is given in the fig. 1.

2. Methodology

2.1. Feature Extraction

Feature extraction is the first step for speaker recognition. In this process a small amount of data is extracted from the voice signal for the identifying a speaker.

2.2. Mel-frequency cepstrum coefficients processor (MFCC)

MFCC is based on the human peripheral auditory system. The human perception of the frequency contents of sounds for speech signals does not follow a linear scale. Thus for each tone with an actual frequency t measured in Hz, a subjective pitch is measured on a scale called the 'Mel Scale' .The mel frequency scale is a linear frequency spacing below 1000 Hz and logarithmic spacing above 1kHz.As a reference point, the pitch of a 1 kHz tone, 40 dB above the perceptual hearing threshold, is defined as 1000 Mels.[1]

2.2.1. Frame Blocking All the recorded audio samples are resampled at 5600Hz. In frame blocking process, each audio file is divided into N short frames of around 20 ms time frame in length with overlap of 10 ms. This allows us to split each 1 second sound file into N individual samples and M overlapping samples. By dividing the signal into such short frames, each section is a relatively constant signal that does not change much.

2.2.2. Windowing Each frame is passed through a windowing function to minimize the discontinuity in the beginning and end of each frame. The concept here is to minimize the spectral distortion by using the window to suppress the signal to zero at the beginning and end of each frame.

Window function for each frame $x_l(n)$ is

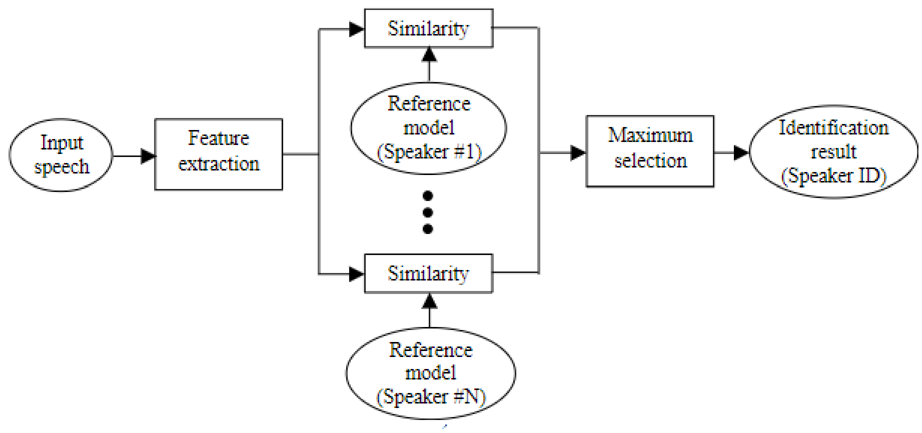

Figure 1. Basic Structure of Speaker Identification

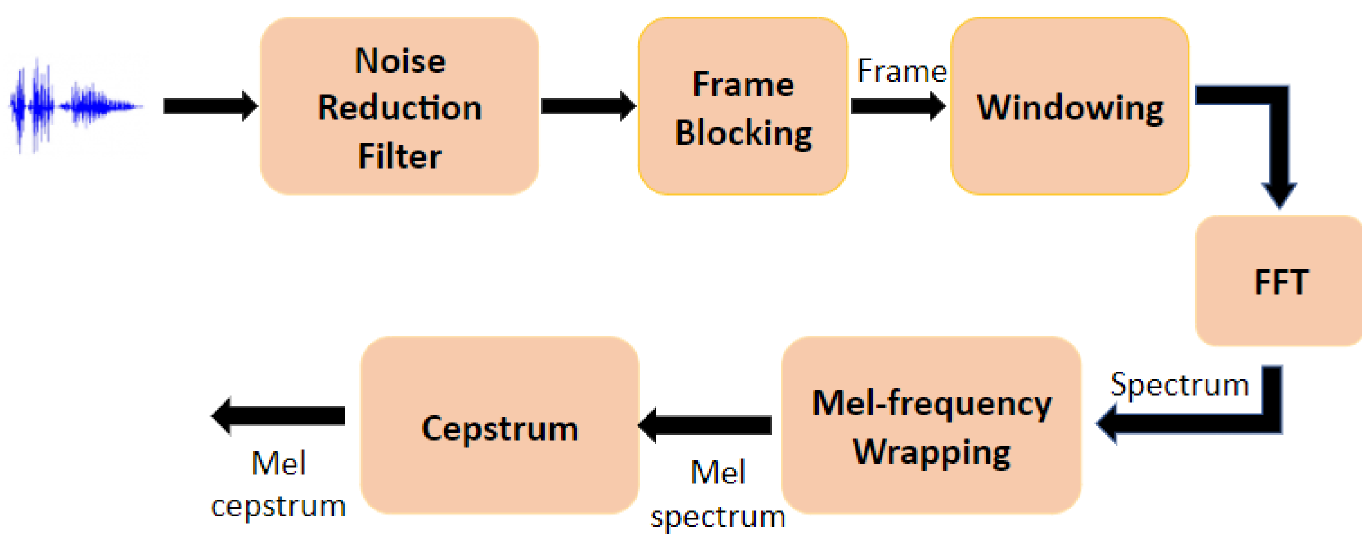

Figure 2. Block diagram of MFCC processor

$$y_l(n) = x_l(n)w_l(n), 0 <= n <= N - 1$$

where N is the number of samples in each frame.

Hamming window is used, which has the form:

$$w_l(n) = 0.54 - 0.46 \cos[\frac{2\pi * n}{(N-1)}], 0 <= n <= N - 1$$

2.2.3. Fast Fourier Transform (FFT) Ordinary .wav files store sound by measuring the amplitude of the signal at a certain sampling rate. Each frame of N samples is transformed from the time domain into the frequency domain. FFT of N samples x_n is

$$X_k = \sum x_n e^{-j2kn/N}, k = 0, 1, 2, .., N - 1$$

$$n = 0, 1, 2, .., N - 1$$

We get spectrum or periodogram.

2.2.4. Short Time Fourier Transform The short-time Fourier transform (STFT) is used to analyze how the frequency content of a nonstationary signal changes over time. The magnitude squared of the STFT is known as the spectrogram time-frequency representation of the signal.

The STFT of a signal is computed by sliding an analysis window w(n) of length M over the signal and calculating the discrete Fourier transform (DFT) of each segment of windowed data. The window hops over the original signal at intervals of R samples, equivalent to L = M – R samples of overlap between adjoining segments. Most window functions taper off at the edges to avoid spectral ringing. The DFT of each windowed segment is added to a complex-valued matrix that contains the magnitude and phase for each point in time and frequency. The STFT matrix has

$$k = \frac{(N_x - L)}{(M - L)}$$

columns, where N $_x$ is the length of the signal x(n). The number of rows in the matrix equals NDFT, the number of DFT points, for centered and two-sided transforms and an odd number close to NDFT/2 for one-sided transforms of real-valued signals.[2]

2.2.5. Mel-frequency Wrapping Frequencies from the FFT are passed through the Mel scale filter bank. It is composed of triangular band-pass filters of equal width in the Mel-Scale (used to measure frequencies based on their pitch from people). The number of mel spectrum coefficients, K, is typically chosen as 20.

Formula to calculate mels for a given frequency f in Hz is

$$m = 2595 * \log_{10}(1 + \frac{f}{100})$$

2.2.6. Cepstrum In this step, log mel spectrum is converted into time domain using Discrete Cosine Transform (DCT) to get mel frequency cepstrum coefficients (MFCC).

$$C_n = \sum (log S_k \cos [n(k - \frac{1}{2}) * \frac{\pi}{K}], n = 0, 1, ..., K-1$$

k=1,2,...,K

[1] [1].

3. Vector Quantization

Vector quantization is used to implement of all learning algorithms. The idea behind it is to treat each n-dimensional vector from each frame as a point in n-dimensional space. These points are arranged into k clusters. Linde, Buzo, Gray (LBG) algorithm is used to determine each cluster center.

[3]

For each speaker, take the array of MFFCs. Center of all these points are mapped by taking the mean of all point. This point will be the first cluster-center. We then split this cluster center into two new centers.

Let X be the vector representing the first cluster center. We define

according to the rule

$$X'_n = X(1 - e), X_n = X(1 + e)$$

where n varies from 1 to the current size of the codebook, and e= 0.01 is a splitting parameter.

We then go through all the vectors again and assign each to the cluster center closest to it. Now each vector in the array is assigned to one of these two cluster centers. For each cluster center, we recalculate its position by finding the mean of each vector assigned

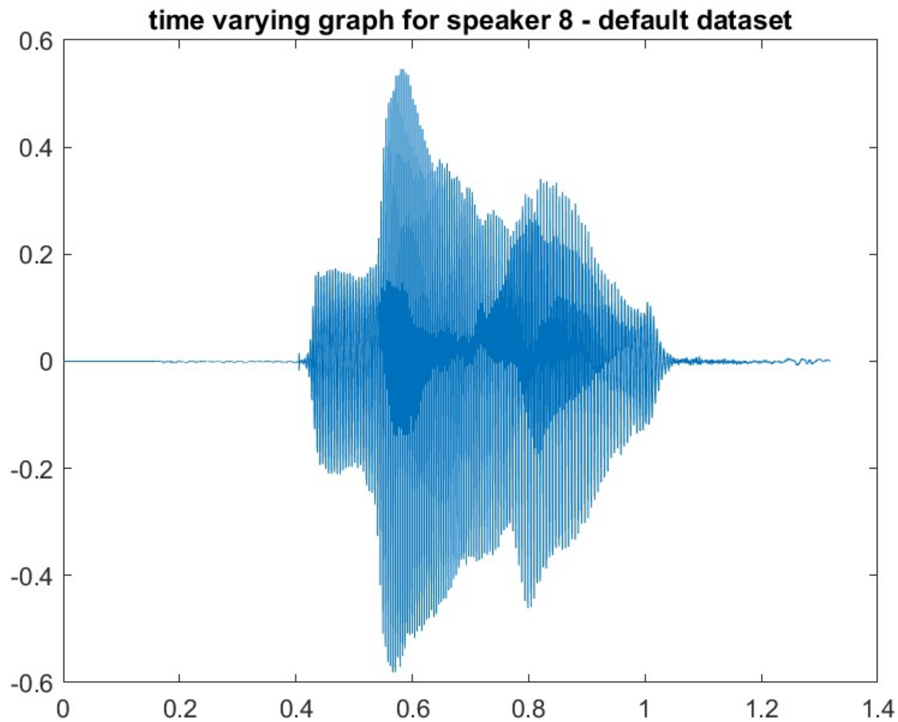

Figure 3. Speech Waveform

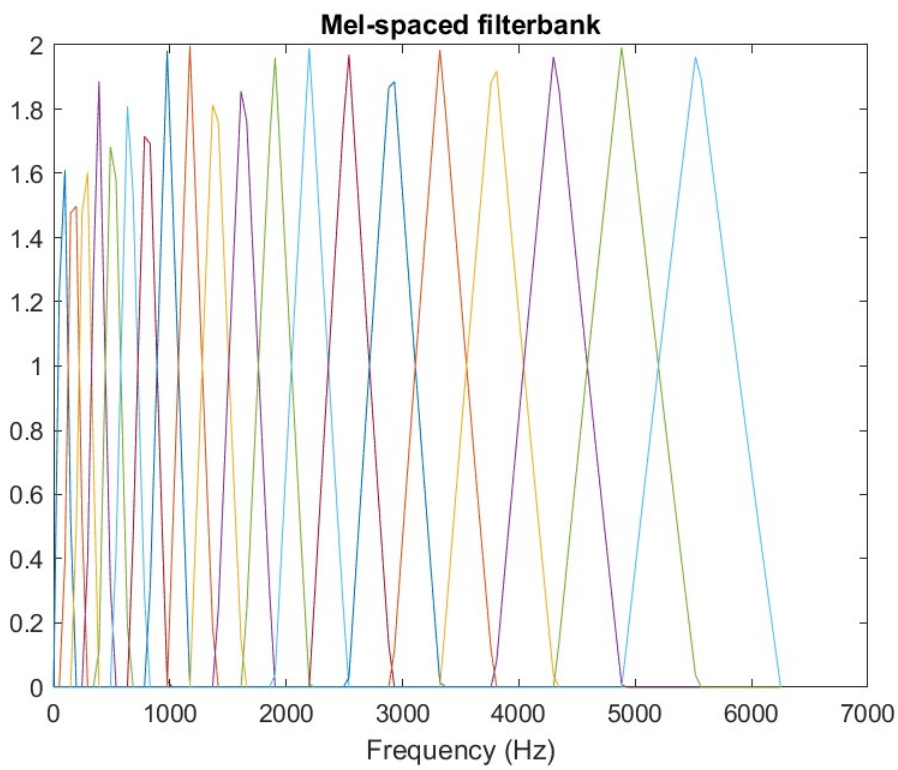

Figure 4. Mel spaced filterbank of Speaker 8

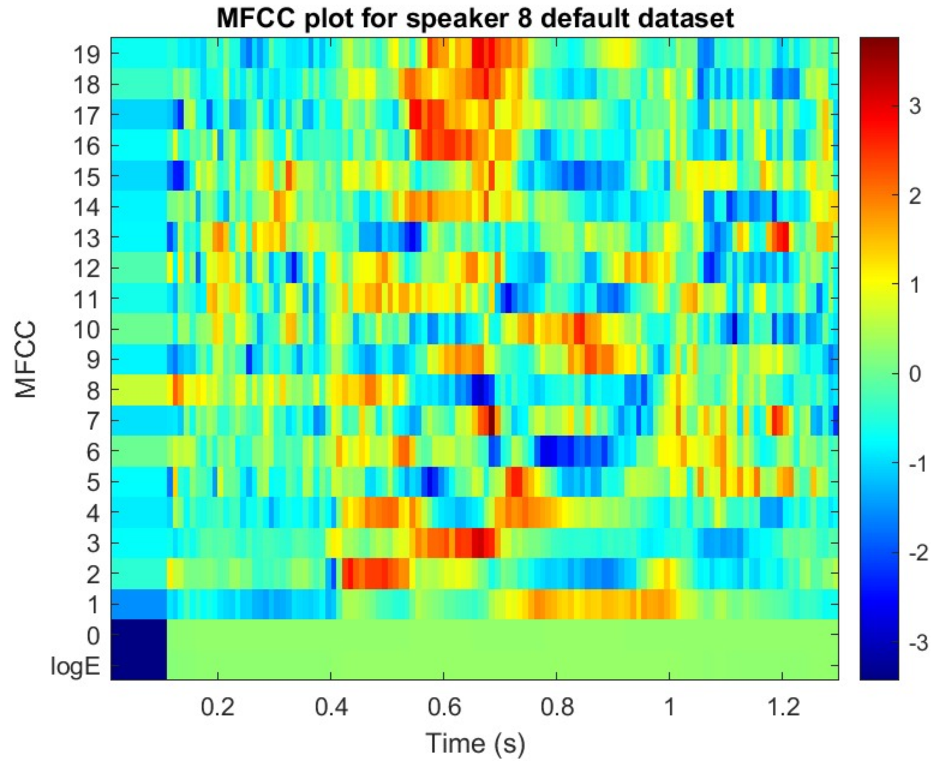

Figure 5. Spectrogram

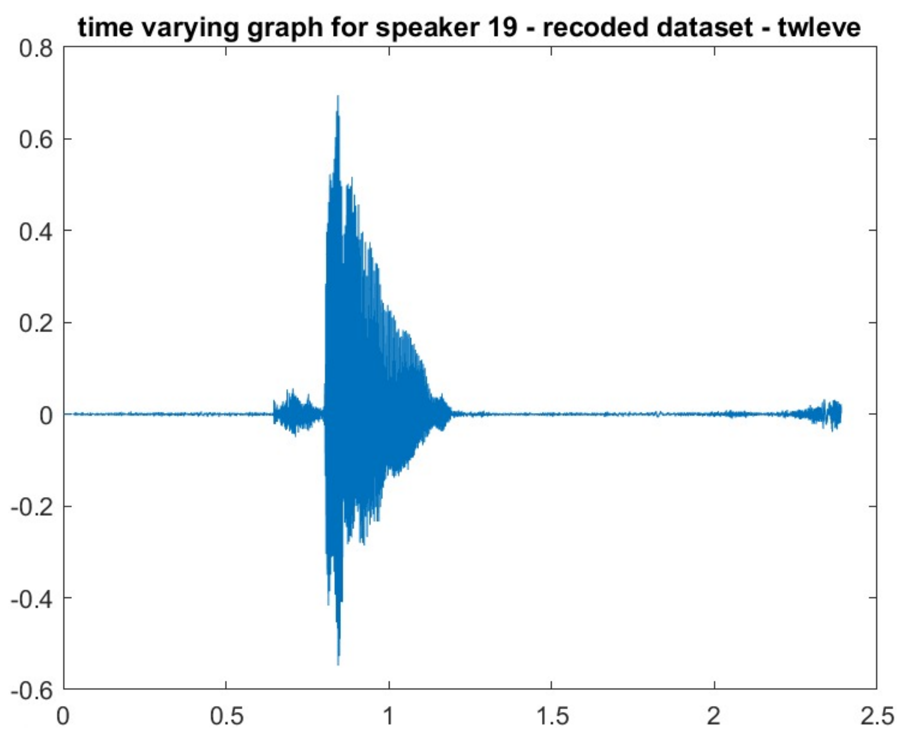

Figure 6. Speech Signal of Speaker 19(twelve)

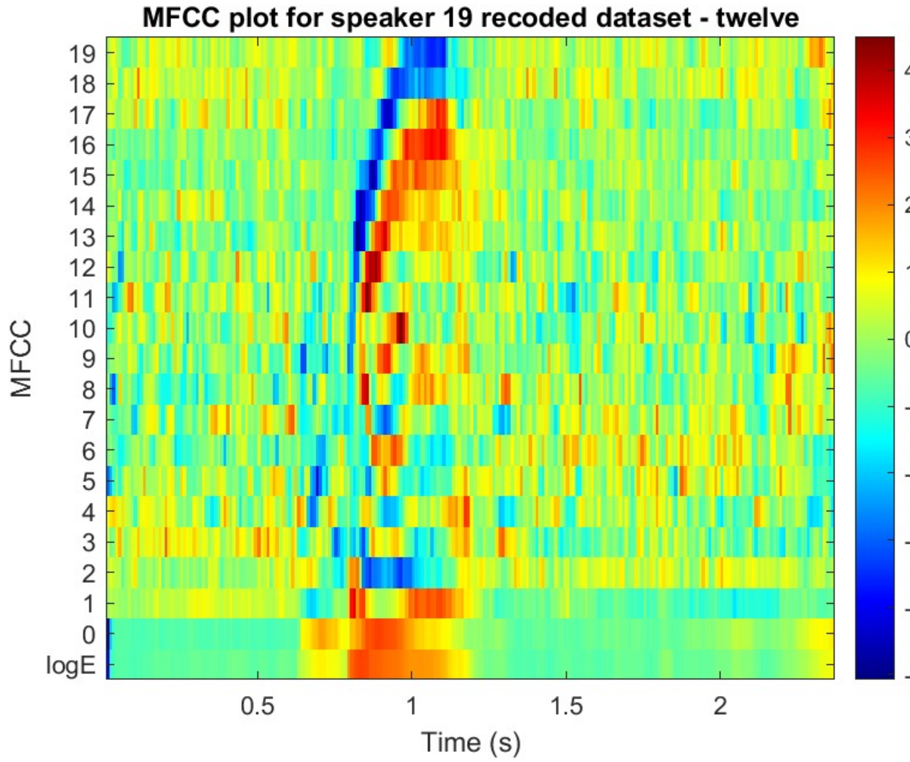

Figure 7. Spectrogram of Speaker 19 (twelve)

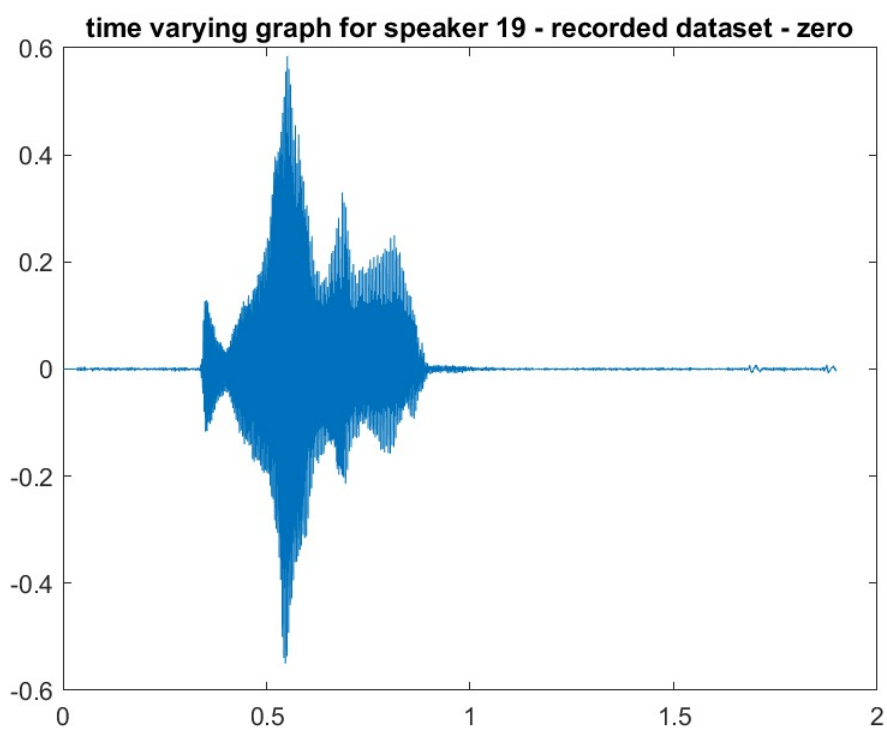

Figure 8. Speech Signal of Speaker 19 (zero)

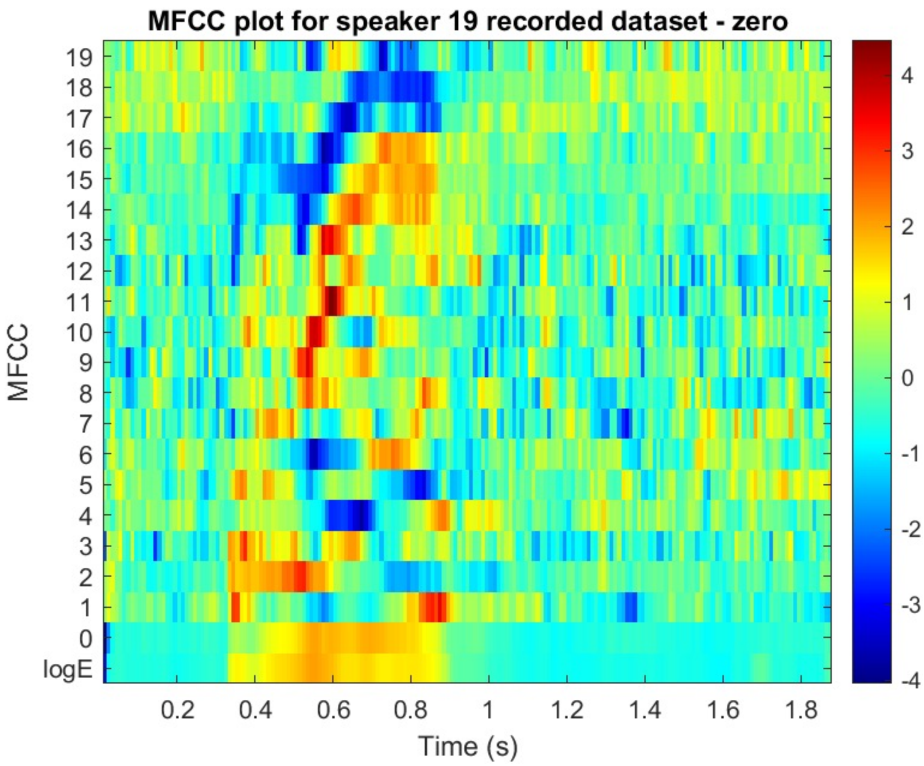

Figure 9. Spectrogram of Speaker 19 (zero)

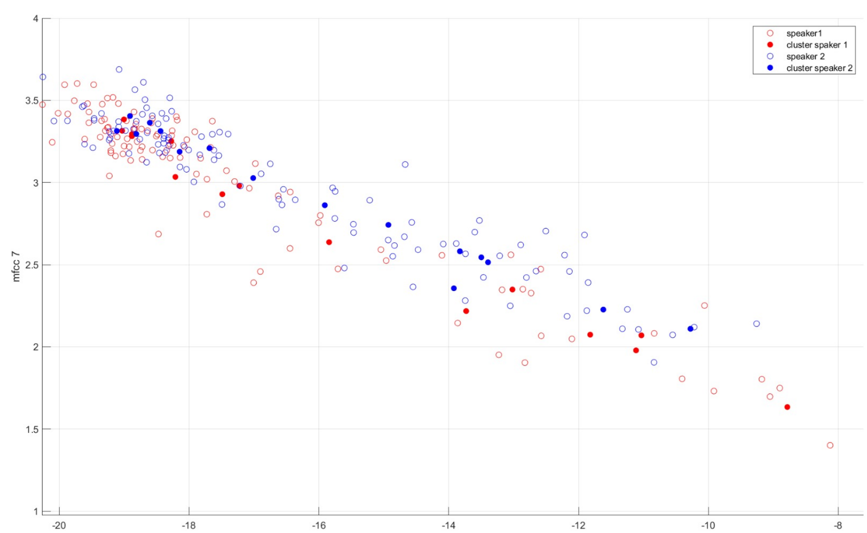

Figure 10. Example of Clustering for speaker dataset

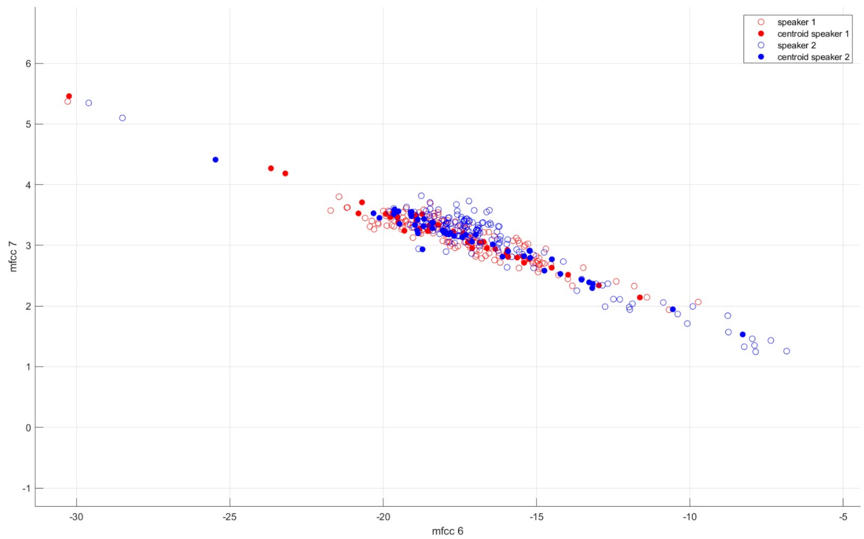

Figure 11. Example of Clustering for zero dataset

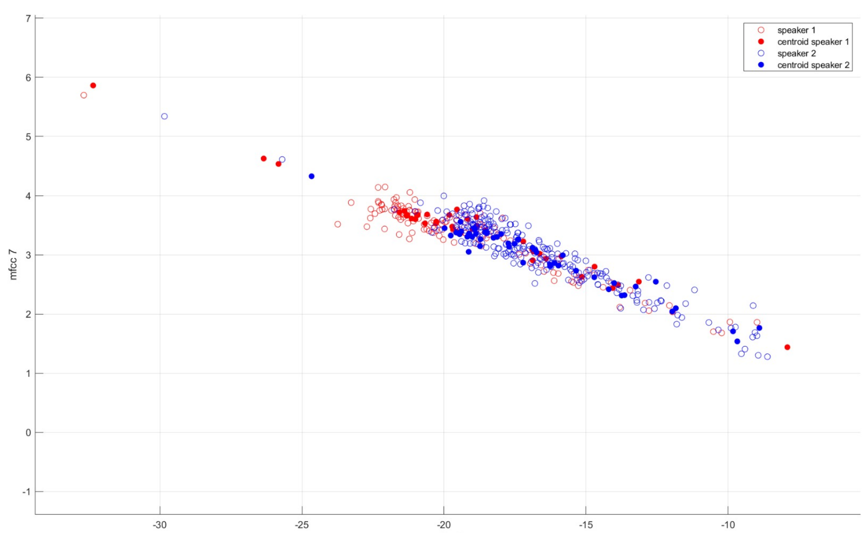

Figure 12. Example of Clustering for twelve dataset

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	predictio
S1	**309.4239**	471.2618	652.3464	456.5511	442.4466	516.5233	439.6235	435.4831	906.9319	927.4457	757.8407	1
S2	615.4884	**313.3701**	572.2096	497.7617	582.4276	647.5451	542.8439	661.3545	1043.769	815.8202	725.6593	2
S3	674.9368	548.4313	**304.4654**	474.8432	584.0589	551.5619	522.0755	579.6971	1017.005	895.7641	829.8869	3
S4	537.2585	519.3353	568.2811	**324.6698**	597.7618	610.8655	488.9124	514.3132	1063.034	974.5910	915.4328	4
S5	734.8321	863.3431	969.0913	866.2182	**450.3871**	713.8499	788.5320	654.9727	1293.068	1437.280	1207.433	5
S6	692.9904	790.3493	674.0464	652.0235	631.4952	**352.1661**	548.8813	604.6485	1153.518	1309.908	1146.086	6
S7	657.2307	614.3812	619.3748	548.6506	605.9586	568.1600	**375.3274**	615.8514	1121.843	1056.443	996.7251	7
S8	529.0112	644.3822	675.0564	553.4255	527.2045	526.7260	531.1531	**361.9045**	1014.884	1153.233	941.1077	8
S9	3152.721	3362.721	3809.680	3568.224	3351.848	3541.282	3204.793	3314.731	**607.9941**	1043.879	1013.220	9
S10	3641.901	3690.457	4174.772	3961.764	3790.783	4022.819	3589.497	3792.017	1044.684	**646.5196**	861.6779	10
S11	3380.912	3395.014	3789.410	3653.937	3519.743	3781.649	3309.046	3502.080	1113.302	879.4938	**707.0457**	11
											Acc	100%

Figure 13. Evaluation of training of speaker dataset

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	prediction
S1	**415.9359**	456.6084	610.9102	462.7926	481.6099	493.8928	455.0355	454.5815	962.9422	874.3072	823.6148	1
S2	658.1387	**401.9343**	572.7095	496.6401	579.2625	634.4541	553.1293	659.6643	1033.760	863.8140	741.3590	2
S3	753.1501	654.5344	**593.9419**	622.3336	701.8784	721.4873	633.4239	687.4091	1323.374	1201.361	1028.938	3
S4	634.7553	616.6956	613.5429	**508.5170**	621.6639	605.5353	604.6348	552.1956	1043.210	1168.047	1030.379	4
S5	618.5611	720.8562	843.1579	732.8956	**516.0775**	598.8361	662.9925	587.0220	1077.521	1258.206	971.2697	5
S6	767.7032	866.1255	797.9517	748.2084	766.4343	**535.5116**	644.3914	714.1726	1444.094	1514.945	1284.596	6
S7	536.4928	538.3147	622.2033	511.1779	546.5394	553.3675	**462.3987**	514.3715	986.2482	998.4427	831.0911	7
S8	829.6057	930.5336	1030.092	933.8729	799.3786	833.5728	846.2914	**782.7571**	979.5740	1258.141	995.8536	8
											Acc	100

Figure 14. Evaluation of testing of speaker dataset

S1	**360.01**	582.167	650.347	654.604	658.153	670.815	626.823	614.836	474.628	626.390	640.592	578.478	588.905	532.420	662.931	703.192	681.8523	672.1283	1
S2	809.819	**492.642**	789.210	770.993	736.425	934.896	867.874	727.930	757.531	818.016	826.294	726.294	754.082	766.646	739.675	911.479	742.5946	771.8739	2
S3	938.79	816.868	**527.866**	813.821	806.549	1245.19	1028.37	793.788	834.188	894.496	887.208	760.956	820.176	809.417	875.967	845.679	765.0070	831.0320	3
S4	887.87	861.87	873.598	**498.477**	732.497	1150.36	912.298	782.902	799.421	852.191	1017.46	795.398	779.498	781.684	820.206	1003.50	867.3067	783.7209	4
S6	1132.3	985.394	1041.05	977.690	**617.684**	1500.29	1332.87	950.755	977.243	1088.45	1280.56	851.444	919.522	1008.88	952.532	1322.21	1121.031	1014.467	6
S7	758.73	766.414	990.848	771.821	799.515	**481.279**	725.160	828.133	690.545	813.460	861.790	794.852	836.266	762.948	813.242	933.900	836.3100	775.5504	7
S8	931.98	813.196	1011.84	833.557	871.576	805.579	**429.794**	829.210	865.844	834.381	947.658	815.876	945.952	767.051	808.230	954.447	822.2154	851.8408	8
S9	565.29	581.856	572.504	519.052	501.435	696.500	569.217	**326.904**	530.088	546.440	611.873	510.353	476.707	513.373	535.963	670.594	527.2217	542.3989	9
S10	1031.2	1140.06	1255.62	1074.53	1080.47	1283.40	1248.17	1113.61	**625.802**	1115.94	1236.73	1045.00	945.433	1017.76	1188.19	1289.18	1212.247	1164.793	10
S11	659.079	705.650	756.016	627.881	577.310	840.334	693.803	563.820	605.953	**457.050**	722.136	577.372	599.294	608.190	691.460	834.369	648.8249	682.4700	11
S12	546.490	530.322	567.816	555.262	526.784	634.136	599.517	475.965	508.979	528.701	**292.174**	468.993	480.947	485.611	601.442	695.496	490.7930	543.8374	12
S13	950.49	837.015	847.730	846.371	749.452	1080.94	977.391	764.810	804.114	830.420	887.942	**541.349**	784.334	750.345	829.273	978.896	763.0503	787.4105	13
S14	1030.0	1007.1	1083.09	957.833	867.263	1368.09	1222.46	897.667	858.305	1004.47	1071.54	904.146	**606.507**	940.449	974.519	1244.21	981.3080	998.9486	14
S15	602.98	619.33	676.149	677.697	633.684	718.276	671.811	585.557	563.345	671.270	687.972	507.242	602.370	**387.620**	625.763	765.336	645.0556	634.0365	15
S16	970.319	786.757	914.697	866.499	753.678	1105.41	980.760	812.245	856.534	964.432	1019.40	799.192	845.468	928.976	**543.093**	977.677	821.4157	890.5612	16
S17	814.264	737.794	698.053	752.472	731.336	1187.10	843.706	762.079	743.282	770.447	851.841	756.677	681.749	850.925	748.203	**501.693**	727.8030	816.5467	17
S18	831.70	776.844	759.618	765.492	723.118	948.799	838.112	672.930	731.017	747.950	812.194	682.688	663.097	720.543	789.429	835.552	**520.2541**	724.8072	18
S19	788.99	699.388	742.698	630.282	651.358	788.481	734.398	634.847	703.374	704.586	783.282	632.673	654.452	640.003	724.325	813.414	674.1612	**444.9618**	19

Figure 15. Evaluation of training of zero dataset

S1	**462.50**	585.74	633.44	681.68	680.02	555.77	576.735	613.71	493.630	616.731	602.96	551.799	600.896	512.432	657.150	706.826	667.816	625.439	1
S2	1042.9	**835.03**	998.31	963.06	902.95	1257.8	1073.59	996.067	904.664	999.963	1152.64	1002.05	1000.69	1084.67	940.578	979.959	1013.974	1021.04	2
S3	784.79	723.32	**604.85**	751.53	687.62	1028.7	984.416	743.50	740.720	760.733	836.17	635.382	724.866	721.427	827.506	814.175	690.3107	710.621	3
S4	959.84	914.61	943.88	**755.95**	815.63	1079.0	931.990	841.06	842.458	889.872	1008.63	826.559	845.062	829.666	882.893	1044.38	870.3378	836.056	4
S6	1288.0	1080.8	1057.9	1038.9	**903.08**	1845.6	1345.66	1091.34	1082.71	1144.09	1391.74	1048.62	1003.69	1160.38	1052.19	1145.51	1137.95	1134.41	6
S7	619.07	603.16	584.34	551.95	555.24	686.29	698.364	580.122	**524.157**	616.111	643.35	532.394	536.925	560.365	619.170	692.889	587.7579	571.114	10
S8	915.64	744.05	950.30	733.23	699.83	811.73	**668.160**	728.554	764.401	790.259	923.490	739.146	826.020	774.110	733.480	958.287	789.8295	774.107	8
S9	644.41	605.93	648.34	584.52	532.82	707.97	601.495	**489.894**	595.704	579.138	622.79	541.401	545.800	557.056	582.034	706.351	542.2447	570.284	9
S10	755.57	797.48	870.04	738.83	764.19	1045.0	951.050	769.609	**628.955**	860.253	996.26	759.206	737.968	701.965	832.569	1013.93	954.0967	844.374	10
S11	476.59	575.54	525.44	497.28	518.32	638.55	637.986	520.887	479.246	499.730	551.79	**448.793**	514.423	507.205	617.476	683.390	580.088	557.470	13
S12	740.67	661.92	637.41	633.74	634.06	881.02	830.196	614.719	646.780	697.161	**584.65**	590.768	617.597	619.869	736.845	821.524	635.1322	659.235	12
S13	958.97	942.67	957.41	888.54	842.88	1165.8	1011.69	873.009	799.756	829.031	876.303	**780.179**	804.063	866.572	975.490	1044.26	894.7651	895.051	13
S14	671.60	622.73	620.72	608.11	582.23	747.07	697.757	579.416	586.296	646.752	710.063	553.694	**537.821**	583.706	596.313	734.477	580.7379	579.306	14
S15	1020.2	895.33	979.26	945.23	904.29	1283.8	1114.54	960.288	919.937	932.284	1121.77	900.484	**889.083**	980.974	1008.96	1071.89	983.014	973.153	14
S16	615.96	481.24	525.91	476.94	481.44	659.94	513.165	462.567	538.517	505.966	652.628	508.860	524.419	514.912	510.413	563.500	493.5507	**449.821**	19
S17	695.86	608.04	545.67	585.77	556.20	1080.5	740.771	636.854	587.329	645.490	727.978	578.742	578.961	673.544	601.398	**501.470**	583.4479	687.807	17
S18	751.28	653.39	**592.97**	658.47	652.00	980.74	771.731	668.566	638.027	667.394	743.515	626.041	650.863	668.257	696.062	593.929	621.9897	656.961	3
S19	944.72	877.85	942.13	780.47	788.58	923.46	845.287	764.802	812.151	831.160	907.986	780.897	828.359	795.924	841.668	1057.40	842.7546	**698.425**	19
																			13/18

Figure 16. Evaluation of testing of zero dataset

S1	**354.76**	635.98	446.38	501.36	509.46	539.76	636.441	494.04	524.961	571.731	480.085	497.717	491.332	475.235	604.211	636.342	708.0787	566.2674	1
S2	774.04	**469.02**	778.72	845.19	964.21	726.49	793.871	818.01	734.451	934.195	825.331	787.466	865.995	719.165	818.926	916.844	852.9191	720.6624	2
S3	580.25	777.85	**398.24**	642.61	631.17	653.05	775.137	612.59	683.020	759.947	636.657	601.731	643.141	584.415	712.513	737.009	825.1187	663.991	3
S4	778.67	1105.2	799.40	**558.14**	818.60	861.21	1078.73	802.907	920.236	985.83	812.884	777.437	770.964	799.082	925.367	1083.62	1167.581	921.055	4
S6	943.35	1524.1	913.54	917.74	**625.85**	1063.61	1538.50	1056.5	1200.46	1114.2	1028.72	953.230	865.267	938.833	1138.87	1278.99	1429.20	1146.16	6
S7	451.55	471.33	423.67	441.79	483.33	**280.72**	511.427	441.58	407.924	505.213	451.938	451.205	427.762	414.943	442.922	526.932	592.526	431.458	7
S8	848.59	783.75	814.87	793.35	1225.5	840.85	**448.272**	857.21	739.762	980.223	1047.02	844.023	977.396	727.676	911.215	1075.91	1105.154	753.327	8
S9	627.40	843.71	552.78	574.60	628.12	655.61	843.852	**378.02**	723.754	770.965	610.938	592.015	612.995	626.248	751.773	815.917	979.6827	700.558	9
S10	667.44	708.64	686.77	664.16	807.34	638.22	721.739	706.363	**370.692**	702.061	749.026	601.498	709.005	614.986	707.776	893.516	980.968	661.316	10
S11	624.09	690.06	582.42	579.28	600.09	596.14	663.725	610.293	539.909	**317.228**	628.038	536.364	649.822	549.901	575.573	768.223	881.294	561.161	11
S12	804.28	884.60	818.66	841.93	843.44	885.65	1047.57	826.314	879.320	967.380	**543.668**	785.629	758.546	803.594	936.984	885.025	925.669	886.364	12
S13	744.68	996.58	733.37	756.57	775.66	855.57	946.293	766.932	812.872	869.296	751.143	**517.530**	758.919	769.327	816.778	894.137	1031.90	847.7594	13
S14	624.98	867.43	582.78	584.08	577.96	666.98	839.143	619.881	732.594	777.009	576.160	573.426	**410.419**	615.322	682.022	766.021	884.912	689.467	14
S15	791.27	996.73	855.00	824.28	1088.0	906.76	932.779	905.304	852.912	934.422	999.987	813.224	959.537	**557.655**	990.771	1162.81	1193.79	866.986	15
S16	695.75	736.06	690.75	692.46	721.11	652.31	761.605	721.286	672.007	786.587	705.119	618.580	668.612	667.940	**393.302**	759.451	808.390	646.9751	16
S17	512.80	526.80	463.06	586.15	542.27	546.15	562.880	533.282	590.865	633.659	476.455	522.676	522.562	498.328	581.014	**326.594**	432.0787	535.583	17
S18	542.47	547.61	536.17	627.32	590.31	582.24	677.344	593.628	635.341	723.673	525.420	570.141	544.878	572.667	630.143	504.227	**363.9058**	579.960	18
S19	883.62	838.26	790.72	853.00	969.59	828.60	866.411	864.258	861.916	929.554	873.879	798.956	904.378	780.099	824.942	975.104	1043.645	**500.6695**	19

Figure 17. Evaluation of training of twelve dataset

S1	**512.32**	710.64	532.02	571.65	574.63	611.321	750.450	554.482	594.310	639.92	552.82	538.699	560.116	575.139	675.790	673.553	750.4401	670.0037	1
S2	632.60	**486.81**	618.82	641.321	696.04	576.35	683.364	645.59	599.380	718.35	613.85	620.716	634.901	601.841	590.521	697.813	669.593	575.160	2
S3	584.31	816.27	**524.52**	651.54	618.82	650.98	799.594	631.71	678.929	701.557	634.607	585.138	651.070	599.966	707.128	739.201	842.865	645.247	3
S4	707.37	883.45	680.35	**642.821**	709.37	716.29	844.967	703.231	746.237	847.34	689.05	660.321	705.117	724.625	749.196	846.964	944.817	781.979	4
S6	875.65	1475.4	871.44	810.69	**658.93**	1009.5	1459.86	910.88	1167.92	928.81	1001.0	908.110	768.764	828.678	1191.48	1365.87	1502.52	1100.08	6
S7	1121.1	1151.3	1194.3	1147.6	1297.3	1138.1	1371.34	1233.12	1069.66	1151.4	1180.2	1086.04	1203.35	**1051.70**	1257.01	1454.40	1427.137	1154.60	15
S8	951.19	923.31	946.11	852.45	1316.2	914.141	**686.014**	941.97	811.10	1061.5	1095.84	870.518	1019.73	810.024	975.222	1187.44	1245.16	878.793	8
S9	799.42	1106.1	772.64	804.65	779.33	855.41	1088.59	754.70	879.064	855.137	827.61	**742.581**	812.102	816.564	881.368	1000.00	1132.437	887.978	13
S10	861.68	874.79	853.47	779.66	929.68	833.74	947.100	791.71	**688.598**	880.51	821.56	783.217	893.161	774.362	879.259	1231.63	1285.35	812.187	10
S11	637.04	651.83	579.22	611.94	632.07	594.00	653.542	575.814	554.432	**483.69**	605.79	536.793	626.736	587.208	593.925	764.595	880.359	565.538	11
S12	674.49	815.42	682.83	696.78	716.67	770.16	882.095	785.551	723.318	823.89	702.32	**647.736**	686.679	683.550	736.973	805.090	761.9084	744.705	13
S13	771.87	1005.9	721.33	755.61	808.34	872.567	939.555	761.09	805.587	852.66	752.681	**664.128**	784.662	741.717	863.790	931.992	1066.11	825.142	13
S14	512.79	836.36	492.88	533.56	521.14	578.76	789.094	535.16	650.443	625.15	540.59	510.837	**481.702**	528.406	659.135	692.270	846.717	643.706	14
S15	**383.78**	488.00	413.34	445.65	495.58	484.31	477.081	459.77	454.150	515.597	426.09	429.167	451.872	399.807	526.910	457.733	464.3041	499.7367	1
S16	699.81	717.67	697.26	700.38	726.01	663.31	748.837	734.25	676.829	718.447	680.03	685.970	653.192	653.041	**620.823**	764.259	833.609	676.007	16
S17	480.26	530.81	449.14	541.82	489.84	534.85	543.231	522.93	568.698	588.331	435.89	474.608	486.782	479.718	545.581	**396.176**	469.9417	528.811	17
S18	557.02	600.56	546.741	602.81	612.93	612.95	675.191	649.457	618.509	746.80	563.49	559.437	565.795	559.912	583.519	562.877	**524.4574**	593.580	18
S19	1013.3	983.64	998.91	1059.3	1157.0	1001.7	1053.88	1089.8	965.702	1050.41	1049.1	995.148	1055.95	937.802	1055.70	1120.16	1241.31	**850.446**	19
																		acc	14/18

Figure 18. Evaluation of testing of zero dataset

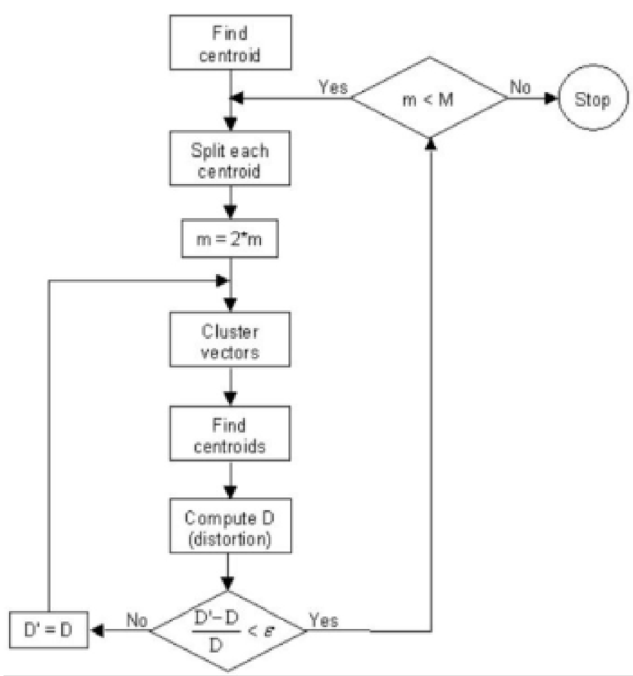

Figure 19. Flowchart of LBG

to it. These new cluster centers are then split again into 2 cluster centers. This process of splitting and recalculating means is repeated until 16 number of cluster centers for given speech data and 64 number of cluster centers for zero and twelve voice recording are found. The result is a collection of cluster centers called a “codebook”. Examples of centroids and their clustering is shown in Fig.
This codebook will represent the way a speaker “sounds” and is ultimately the tool to classify which speaker is assigned to a new speech file.

4. Result

The training sample data and given speaker data was able to achieve a 100% accuracy rate while the overall accuracy rate of the model is 91%. Performance of the speaker recognition system against the training set is in Figure 13-18. The column corresponding to the minimum distance in each row is re-coded as the predicted user. All the testing datasets of given speech data are predicted accurately.
In testing dataset of zero 13 out of 18 are correctly predicted whereas in testing dataset of twelve 14 out of 18 are accurately predicted. This discrepancy can most likely be attributed to sampling frequency or number of centroids. Modifying the appropriate matlab code should greatly increase the models accuracy.

5. Conclusion

In this project, we created a model using digital signal processing and voice recognition pattern to identify the user of a given speech signal. We have used a very simple and efficient approach of vector quantization and Euclidean minimum distance on the given datasets. It can achieve the accuracy with with no computation complexity, less execution time.

6. Appendix

Github Link :-https://github.com/ari1idont/eec201$dspabbs$
https://tinyurl.com/3yk6ykwa

Contributions:-
Arindam- vector quantization, LBG Algorithm and testing
Bhawna- STFT, report, video and presentation

References

[1] V. Tiwari, “Mfcc and its applications in speaker recognition,” *International Journal on Emerging Technologies*, vol. 1, pp. 19–22, 2010.

[2] “Short-time Fourier transform - MATLAB stft — mathworks.com.” `https://www.mathworks.com/help/signal/ref/stft.html`. [Accessed 19-03-2024].

[3] R. G. Yoseph Linde, Andres Buzo, “An algorithm for vector quantizer desig,” *IEEE Transactions on Information Theory*, vol. 28, no. 1, 1980.