Turkish Journal of Computer and Mathematics Education *Vol.12 No.3(2021), 5105-5116*

Research Article

Artificial Neural Network Based Amharic Language Speaker Recognition

Gizachew Belayneh Gebre[a], Teklu Urgessa[b], T.GopiKrishna[c]

[a] College of Computing and Informatics, Department of Software Engineering , Haramaya University, Ethiopia
[b,c] Dept of Computer Science and Engineering , Adama Science and Technology University, Adama, Ethiopia
[a]legizachew@gmail.com, [b]tekluurgessa2009@gmail.com, [c] gktiruveedula@gmail.com

Article History: Received: 10 November 2020; Revised 12 January 2021 Accepted: 27 January 2021; Published online: 5 April 2021

Abstract: In this artificial intelligence time, speaker recognition is the most useful biometric recognition technique. Security is a big issue that needs careful attention because of every activities have been becoming automated and internet based. For security purpose, unique features of authorized user are highly needed. Voice is one of the wonderful unique biometric features. So, developing speaker recognition based on scientific research is the most concerned issue. Nowadays, criminal activities are increasing day to day in different clever way. So, every country should have strengthen forensic investigation using such technologies. The study was done by inspiration of contextualizing this concept for our country. In this study, text-independent Amharic language speaker recognition model was developed using Mel-Frequency Cepstral Coefficients to extract features from preprocessed speech signals and Artificial Neural Network to model the feature vector obtained from the Mel-Frequency Cepstral Coefficients and to classify objects while testing. The researcher used 20 sampled speeches of 10 each speaker (total of 200 speech samples) for training and testing separately. By setting the number of hidden neurons to 15, 20, and 25, three different models have been developed and evaluated for accuracy. The fourth-generation high-level programming language and interactive environment MATLAB is used to conduct the overall study implementations. At the end, very promising findings have been obtained. The study achieved better performance than other related researches which used Vector Quantization and Gaussian Mixture Model modelling techniques. Implementable result could obtain for the future by increasing number of speakers and speech samples and including the four Amharic accents.

Keywords: Text-independent Speaker Recognition, ANN, MFCC, Feature Extraction, Amharic Language, MATLAB Neural Network Tool

1. Introduction

Speech is a medium for human to express their thoughts during communication. A speech signal is a complex signal which is packed with several knowledge resources such as acoustic, articulatory, semantics, linguistic and many more. It is the ultimate, universal mode of human communication and it is how a man should be able to interact with computers or machines [1]. Speech interface in the user's own language is in short, an ideal means of communication as being the most natural, flexible, efficient and convenient option allowing to perform hands and eyes free tasks. The study of speech signals and the processing methods of signals is called speech processing. It is the analysis of human speech (using digital signal processing techniques). There are several aspects of speech processing, according to the focus of analysis: speech synthesis, speech recognition, speaker recognition, voice analysis, speech coding and compression, speech enhancement, speaker diarization, speaker classification, language identification, etc. Speaker recognition which is the concern of this study is the process of automatically recognizing who is speaking by using the speaker-specific information included in speech waves to verify identities being claimed by people accessing systems; that is, it enables access control of various services by voice [2]. The aim of speaker recognition is to extract, characterize and recognize the information about speaker identity.

In this digital world, speaker recognition is the most useful biometric recognition technique. Now days many organizations like bank, industries, access control systems, etc. are using this technology for providing greater security to their vast databases [3]. Applicable services include voice dialing, banking over a telephone network, telephone shopping, database access services, information and reservation services, voice mail, security control for confidential information, and remote access to computers. The most essential application of speaker recognition technology is as a forensics tool.

Speaker recognition can be classified into speaker identification and speaker verification. Speaker identification is the process of determining from which of the registered speakers a given utterance comes. Speaker verification is the process of accepting or rejecting the identity claimed by a speaker. Most of the applications in which voice is used to confirm the identity of a speaker are classified as speaker verification. From

a security perspective, identification is different from verification. For example, presenting your passport at border control is a verification process: the agent compares your face to the picture in the document. Conversely, a police officer comparing a sketch of an assailant against a database of previously documented criminals to find the closest match(s) is an identification process.

Speaker recognition systems fall into two categories, text-dependent, and text-independent. Text-dependent uses the same text for enrollment and testing. Text-independent uses different text for enrollment and testing. This study is dealt with text-independent speaker identification in a case of Amharic language speech.

At the highest level, all speaker recognition systems contain two main modules feature extraction and feature modeling and 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 modeling and matching is modeling the extracted features and involves the actual procedure to identify the unknown speaker by comparing extracted features from the input voice with the ones from a set of known speakers.

However, the speech based systems that have been developed so far are meant to serve a specific language and are totally limited to some techno-rich countries of the world. For developing countries like Ethiopia, it is a must, to follow the outfit of those techno-rich countries in relation to such technological advancements to do not lose the opportunities provided by technologies. Based on this fact, speech engineers and language experts invarious countries are making noticeable efforts to develop recognition that works for their own language. In our country, even though it is not enough 40 speech recognition and 3 speaker recognition researcheshad been attempted. As we have seen, there is a shortage of research regarding speaker recognition. The above three speaker recognition concerned researches used Vector Quantization and Gaussian Mixture model for modeling techniques. The goal of this study is exploring the possibility of a state of the art modeling techniques for building Amharic language speaker recognition.

2. Literature Review

S. A. Mahmood and L. E. George, 2007, investigate neural based speaker recognition system. LPC has been used as a feature extraction method. And back propagation neural network has been used for the purpose of speaker modeling and identification. Achieved 90% accuracy for 10 speakers [4].

M. S. Sinith et al. , 2010, emphasis on text-Independent speaker identification system using Mel-Frequency Cepstral Coefficients (MFCC) as the speaker speech feature parameters in the system and the concept of Gaussian Mixture Modeling (GMM) for modeling the extracted speech feature. And used the Maximum Likelihood Ratio Detector algorithm for the decision making process. The experimental study has been conducted on MATLAB 7. Gaussian mixture speaker model attains high recognition rate for various speech durations. The recognition rate is maximum (98.8 %) when the speech is of 60 seconds duration and the number of Gaussians is 16 [5].

Amr Rashed, 2014, this paper proposed a fast algorithm for speaker recognition. This algorithm first records voice patterns of speakers via noisy channel and use some of noise removal techniques. The feature is extracted by Mel Frequency Cepstral Coefficient (MFCC) and the feature is reduced by Principal component analysis (PCA) technique. Then the result vector is fed to ANN classifier. Experimental results indicates that using ANN with weight/bias training algorithm have better performance. The result shows that the proposed algorithm achieved on average about 99% accuracy rate and higher speed rate in comparison with other methods [7].

A. Azene, 2015, the first attempted Amharic language speaker recognition research. It presents text-independent speaker identification system for the Amharic language. Speech signals are collected from different speakers including both sexes as well as different age groups. MFCC had been used to extract features from the speech signals and to generate feature vector. VQ and GMMs had been used for training and identification purpose. The intention of the researcher was to see which modeling approach is better for text- independent speaker identification. For a total of 50 speakers, 74.2% accuracy was achieved when VQ approach is used where as 84.3% accuracy for the GMMs. The researcher tried to see the speaker identification accuracy based on gender. 25 male and 25 female speakers were considered. From the experiment, 86.2% and 85.9% accuracy was achieved for male and female speakers respectively [5].

A. D. Mengistu, 2017, presents an automatic text-independent speaker identification system for the Amharic language in noisy environments. Speech signals are collected from different 100 speakers including both genders. Each speech has 10 seconds duration from each individual. Combination of MFCC, LPCC, and GFCC had been used for feature extraction purpose. VQ and GMMs had been used for training and identification purpose. The researcher was attempting to see which modeling approach is better for text- independent speaker identification with the combination of the three feature extraction techniques (MFCC, LPCC, and GFCC). The researcher conducted two experiments; One, VQ modeling approach with the combination of the three feature extraction techniques for 30, 60, and 90 speakers. And achieved 77.2%, 70.9%, and 69% accuracy respectively. Two, GMM

modeling approach with the combination of the three feature extraction techniques for 30, 60, and 90 speakers. And achieved 75.2%, 76.9%, and 78% accuracy respectively [6].

M. Islam, F. Khan, and A. M. Haque, 2013, presents the implementation of Text Independent Speaker Identification system. Feature extraction task has been done using Mel Frequency Cepstral Coefficients (MFCC) acquisition algorithm that extracts features from the speech signal, which are actually the vectors of coefficients. The backpropagation algorithm of the artificial neural network stores the extracted features on a database and then identify speaker based on the information. Achieved near 100% accuracy in case of static speech signal and above 90% accuracy in case of real time speech signal [6].

D. Mengistu D. Melesew, 2017, presents a hybrid approach of VQ and GMM have been used for classifying dialects of Amharic language. In speech signals collection, total of 100 speakers from each group of dialects (Gojjam, Wollo, Shewa, and Gonder) are considered. MFCC feature vectors are used to recognize the dialects of speakers. When 25 speakers are considered from areas, 85.9% accuracy had been achieved. When the number of speakers are increased to100, which is the maximum number of dialect speakers of the experiment, 92.7% accuracy had been achieved [7].

A. Antony and R. Gopikakumari, 2018, introduces an isolated word speaker identification system based on a new feature extractor and using ANN. The system is designed for both text independent and text dependent speaker identification system for English words. The speech is recorded using audio wave recorder. Then the preprocessing is applied for the given speech signals. UMRT is a transform which has been used for image compression. Combinations of MFCC and UMRT are taken and are used as a feature extractor. The classification of the features is done using Multi-layer perceptron with back propagation algorithm. The accuracy is taken using confusion matrix. The accuracy achieved is around 97.91% for speech dependent systems while for speech independent system the accuracy is around 94.44% [8].

3. Methodology

Activities involved in this study framework of methodology are speech collection, preprocessing of speech signals, feature extraction of preprocessed data, feature modeling of extracted features, feature matching during identification, and performance evaluation. All activities has been done using an appropriate techniques and tools. Figure 1. shows the overall detailed framework of the methodology.

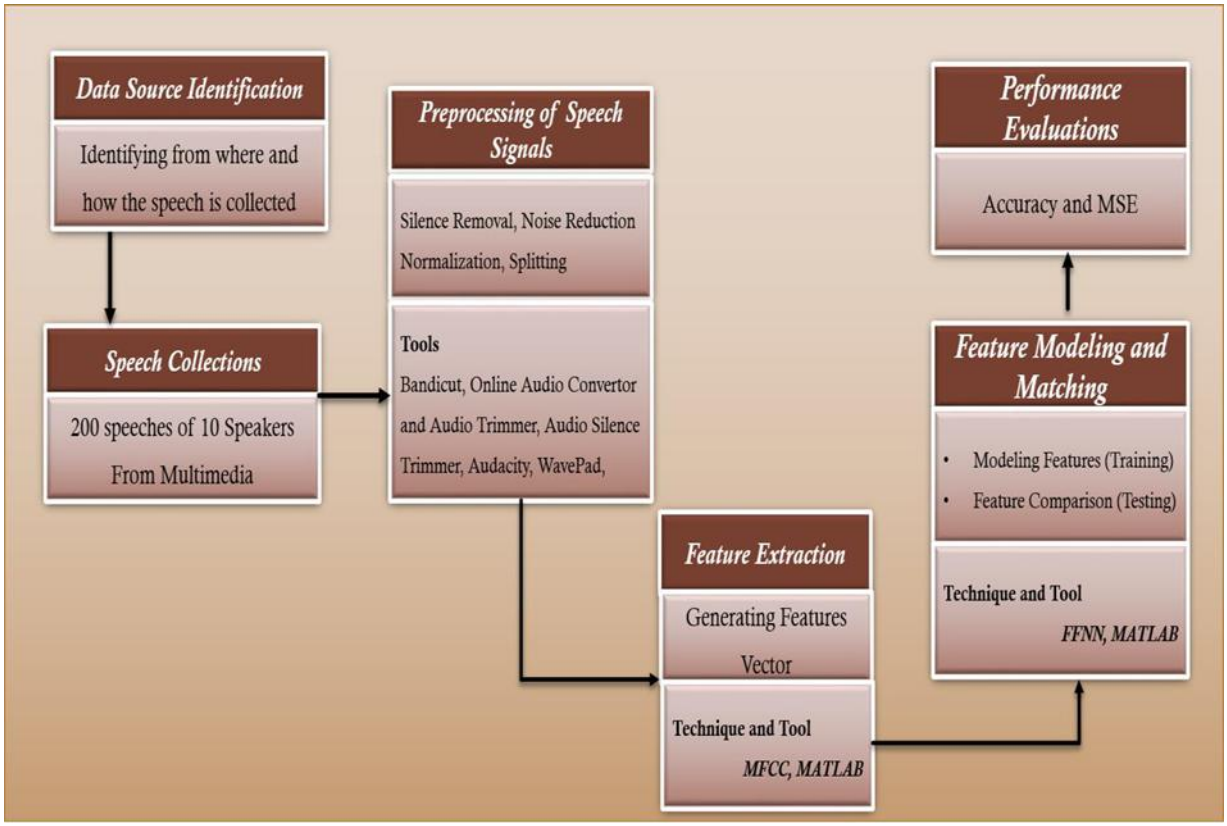

Figure 1. Conceptual Framework of the Methodology

4. Experiments, Findings

Proposed Speaker Recognition Prototype

In order to have simple interaction in demonstration, we developed graphical user interface prototype. So, in this study, the overall processes are conducted using GUI speaker recognition prototype.

Figure 2. depicts the general architecture of the proposed speaker recognition prototype which shows the overall process flow.

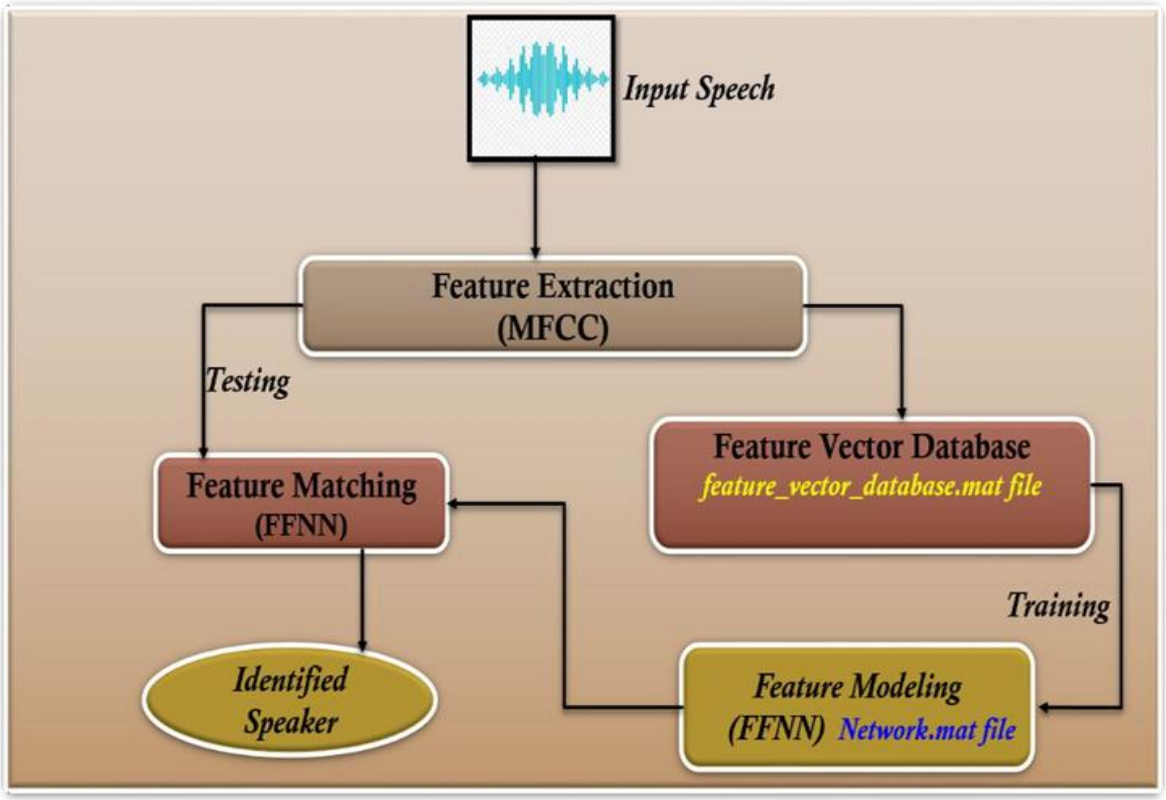

Figure 2. Architecture of Proposed Speaker Recognition Prototype

Input speech is preprocessed speech signal. The preprocessed input speech always pass through feature extraction technique (MFCC). After feature extraction, there are two paths; training and testing. During training, feature vector of speech signal is created and stored in to database. Then, all feature vectors of speech signals which have been stored in the database will be trained for the feedforward neural network and 'Network.mat' file which contain the trained neural network is created. During testing, a comparison between feature vector of the input speech and feature vector of speech signals in the trained neural network (in Network.mat) will be done. Finally, after comparison has been made, the neural network make a decision to identify the speaker.

Feature Extraction

Since the prototype is GUI, the preprocessed input speech is loaded one by one to the system for feature extraction. The class ID is given when the input speech is loaded for feature extraction. 200 speech samples of 10 speaker has been loaded to the system for feature extraction and stored in database (speech_feature_vector_database).

features_data(14, 1)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	2.8699	1.1832	-0.1893	-0.0076	-0.5339	-0.4064	-0.5626	-0.6033	-0.2004	-0.8687	-0.5650	-0.1731	0.3479
2	4.2790	-0.2889	0.0301	0.4275	-0.3201	-0.3705	-0.8595	-0.6920	0.1130	-1.5117	-0.3485	-0.3310	0.8741
3	4.5847	-0.8375	-0.0405	1.1251	-0.0074	-0.1313	-0.8534	-0.6244	0.2521	-1.4954	-0.0995	0.0944	0.9358
4	4.6909	-1.2093	-0.0015	1.4364	-0.3671	0.6279	-0.6668	-0.2150	0.1471	-1.6593	0.2909	0.4088	0.9851
5	4.9409	-1.6793	-0.2533	1.5911	-0.6909	1.0484	-0.5419	0.3667	0.4952	-1.3673	0.8776	0.2083	0.9911
6	5.1179	-1.6914	-0.1011	2.0299	-0.6555	0.7897	-0.6613	0.4115	0.6063	-1.2128	0.6449	0.1331	1.0622
7	5.5996	-1.5133	-0.3935	1.9989	-0.9916	0.7663	-0.6968	0.0636	0.6775	-1.2701	0.4354	0.5996	1.1428
8	6.0111	-1.2313	-0.5928	1.4986	-1.3707	0.9227	-0.5487	-0.1562	0.5108	-1.2437	0.6453	0.8413	1.1721
9	5.7036	-1.4455	-0.2355	1.8168	-1.3632	1.0735	-0.6215	-0.3802	0.2836	-0.9968	0.6953	0.8106	1.0788
10	4.7554	-1.7645	0.1953	1.8278	-1.0372	1.1053	-0.4247	-0.1545	0.4770	-0.7717	0.9953	0.2042	0.7368
11	4.8013	-2.2152	-0.0225	1.5327	-1.1544	1.0819	-0.6811	-0.1482	0.2963	-0.7788	1.2454	0.2078	0.9213
12	4.6836	-2.0491	0.3261	1.4392	-0.8911	1.2143	-0.9335	-0.0871	0.0290	-1.2212	1.2117	0.2859	1.1183
13	5.1574	-1.3266	0.3223	1.0983	-0.7902	1.2110	-1.1000	0.1735	0.1053	-1.5509	0.9598	0.0539	1.1092
14	5.4253	-1.3542	0.1687	1.4409	-0.6338	1.0628	-0.9483	0.1725	-0.1955	-1.6081	0.6595	-0.0656	1.0060
15	5.3100	-1.2870	0.3225	1.6961	-0.5680	0.8675	-1.1794	-0.2543	-0.0200	-1.4683	0.6605	-0.1907	1.0504
16	4.8300	-1.0823	0.6564	1.6126	-0.6194	1.1056	-1.0781	-0.3998	0.0179	-1.0867	0.9029	-0.3970	0.9124
17	4.6591	-0.9536	0.6255	1.4791	-0.3694	1.3657	-0.8097	-0.3308	-0.3412	-1.2785	1.0447	-0.3298	0.9709
18	4.5483	-1.1791	0.5441	1.5260	-0.2804	1.3737	-0.8132	-0.1179	-0.2073	-1.4184	1.2200	-0.2572	0.9381

Figure 3. MFCC Feature Vector Matrix of AA_00.wav

Finally, the class ID is given based on the alphabetical order from 1 up to 10.

Table 1. Training dataset and their corresponding class ID after feature extraction

No.	Speaker Name	Speaker Code	Wav files'Name (20 perSpeaker)	Class ID
1	Dr. Abiy Ahmed	AA	AA_00 to AA_19	1
2	Biniam Belete	BB	BB_00 to BB_19	2
3	Birtukan Mideksa	BM	BM_00 to BM_19	3
4	Fikadu T/mariam	FT	FT_00 to FT_19	4
5	Meaza Biru	MB	MB_00 to MB_19	5
6	Dr. Mihret Debebe	MD	MD_00 to MD_19	6
7	Muferiat Kamil	MK	MK_00 to MK_19	7
8	Sahilework Zewde	SZ	SZ_00 to SZ_19	8
9	Ustaz Abubeker Ahmed	UAA	UAA_00 to UAA_19	9
10	Yetnebersh Nigussie	YN	YN_00 to YN_19	10

The code given for the speakers' is assigned using the two capital letters of first and last name of the speaker. Each speaker has 20 speeches, and the wav files are represented using the speaker code with numbers from 00-19 (i.e. AA_00.wav, AA_01.wav, AA_02.wav……. AA_19.wav).

Training of Neural Network

During the feature extraction stage, the speech signals are transformed in to feature vector in the way it will be suitable for training the neural network. The next step is training and testing the neural network. As tried to mention in chapter 3, in this study, three experiments have been conducted. After setting the training parameters, the set of inputs with respective target outputs fed for the neural network sequentially.

The training parameters used for all experiments are listed below

- Number of Input Neurons=26
- Number of Hidden Neurons=15, 20, 25 (Expt.1, Expt.2, and Exp. 3 respectively).
- Number of Output Neurons=10
- Train Function= trainlm (Levenberg-Marquardt)
- Performance Function= Mean Squared Error (MSE)
- Divide Function= dividerand
- Training Ratio=70%, Validation Ratio=15%, Testing Ratio=15%
- Epochs=100

Figure 4. is the plot of the performance of experiment one for the training, validation, and testing versus the epochs.

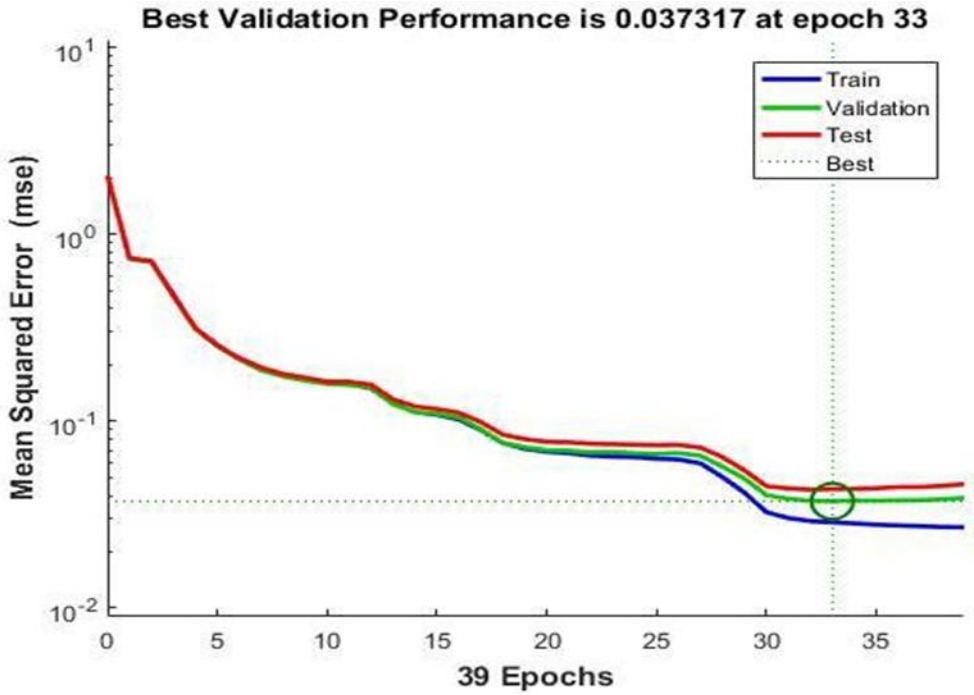

Figure 4. Performance Plot of Experiment One [Hidden Neurons=15]

Figure 5. is the plot of the performance of experiment two for the training, validation, and testing versus the epochs.

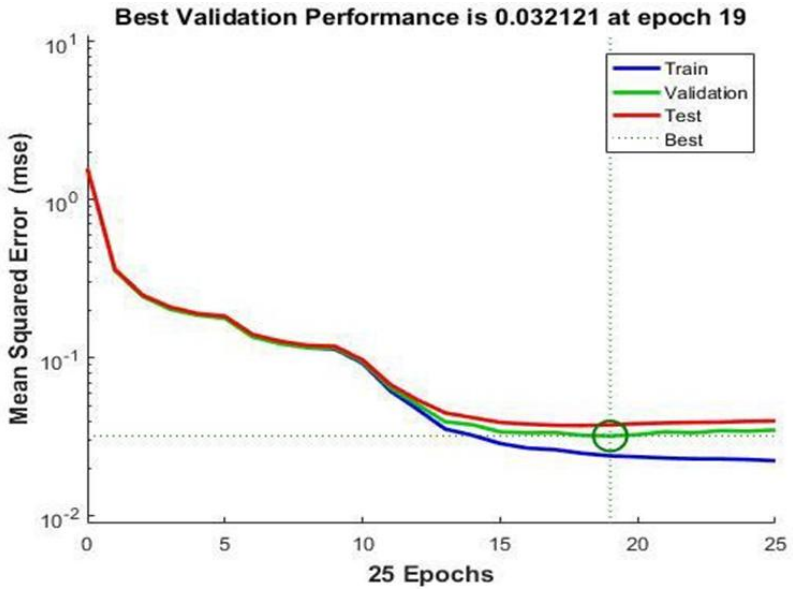

Figure 5. Performance Plot of Experiment Two [Hidden Neurons=20]

Figure 6. is the plot of the performance of experiment three for the training, validation, and testing versus the epochs

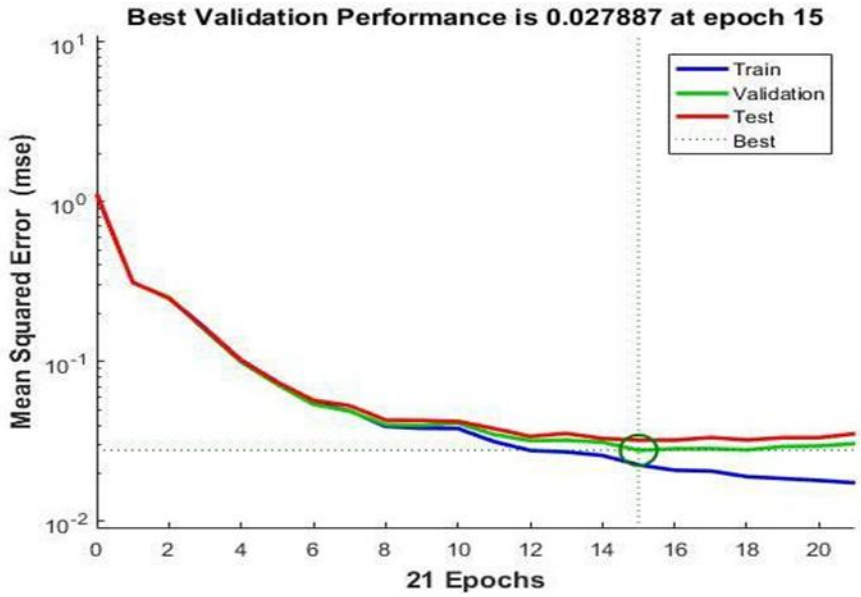

Figure 6. Performance Plot of Experiment Three [Hidden Neurons=25]

Performance Evaluation

Here, the evaluation has been performed in two ways:

- Trained neural network model evaluation based on confusion matrix.
- Real time test using speeches that are prepared for testing.

Based on all confusion matrixes, the below result summary has been summarized.

Table 2. Result Summary Based on Gender

No. Hidden Neurons	15	20	25
Overall Accuracy	96.0 %	96.7%	97.3%
Male	49.5%	49.8%	49.9%
Female	50.5%	50.2%	50.1%

Table 3. Confusion Matrix Metrics Result of each Class of all Experiments

Classes	Hidden Neuron =15					Hidden Neuron =20					Hidden Neuron =25				
	# Samples of Class	TP	FN	FP	TN	# Samples of Class	TP	FN	FP	TN	# Samples of Class	TP	FN	FP	TN
Class 1	2730	2558	172	83	23214	2730	2567	163	64	23206	2730	2579	151	50	23229
Class 2	2600	2435	165	243	23184	2600	2497	103	148	23252	2599	2506	93	103	23307
Class 3	2600	2496	104	129	23298	2600	2497	103	81	23319	2600	2525	75	68	23341
Class 4	2470	2458	12	160	23397	2470	2455	15	178	23352	2470	2461	9	141	23398
Class 5	2627	2596	31	21	23379	2600	2578	22	43	23357	2600	2571	29	16	23393
Class 6	2600	2544	56	37	23390	2600	2545	55	18	23382	2600	2560	40	23	23386
Class 7	2600	2321	279	102	23325	2600	2381	219	109	23291	2610	2438	172	123	23276
Class 8	2600	2541	59	70	23357	2600	2552	48	46	23354	2600	2555	45	35	23374
Class 9	2600	2495	105	136	23291	2600	2493	107	140	23260	2600	2534	66	85	23324
Class 10	2600	2551	49	51	23376	2600	2570	30	28	23372	2600	2577	23	50	23359
Total	**26027**					**26000**					**26009**				

Experiment One: [Hidden Neuron=15]

Table 4 shows how to classify the samples sound frames versus with four confusion matrix parameters. For example, Class 1, has total 2730 sound sample frames stored in sound database extracted from his voice. From those, the truly classified samples are the actual Class 1 predicted correctly 2558 (TP). The 83 (FP) frame samples are classified incorrectly. The 172 (FN) frame samples of actual Class 1 are predicted incorrectly by otherclasses.The23214(TN)framesampledneitheractualclassnorpredictedclasswhich means truly predicted incorrectly from the total samples**26027**.

Experiment Two: [Hidden Neuron=20]

Table 4 shows how to classify the samples sound frames versus with four confusion matrix parameters. For example, Class 1, has total 2730 sound sample frames stored in sound database extracted from his voice. From those, the truly classified samples are the actual Class 1 predicted correctly 2567 (TP). The 64 (FP) frame samples are classified incorrectly.

The 163 (FN) frame samples of actual Class 1 are predicted incorrectly by otherclasses.The23206(TN)framesampledneitheractualclassnorpredictedclasswhich means truly predicted incorrectly from the total samples**26000**.

Table 4. Computed Model and Classification Accuracy of all Experiments

No. of Hidden Nuerons	Metrics	Classes									
		1	2	3	4	5	6	7	8	9	10
	Model	9.8	9.4	9.6	9.5	9.9	9.8	8.9	9.8	9.6	9.8
	Classification Accuracy (%)	99	98.4	99.1	99.3	99.8	100	98.5	99.5	99.1	99.6
15	Precision (%)	96.9	90.9	95.1	93.9	99.2	99	95.8	97.3	94.8	98
	Recall (%)	93.7	93.7	96	99.5	98.8	98	89.3	97.7	96	98.1
	F1-score (%)	95.3	92.3	95.6	96.6	99	98	92.4	97.5	95.4	98.1
	Model	9.9	9.6	9.6	9.4	9.9	9.8	9.2	9.8	9.6	9.9
	Classification Accuracy (%)	99.1	99	99.3	99.3	99.8	100	98.7	99.6	99	99.8
20	Precision (%)	97.6	94.4	96.9	93.2	98.4	99	95.6	98.2	94.7	98.5
	Recall (%)	94	96	96	99.4	99.2	98	91.6	98.2	95.9	98.8
	F1-score (%)	95.8	95.2	96.5	96.2	98.8	99	93.6	98.2	95.3	98.7
	Model	9.9	9.6	9.7	9.5	9.9	9.8	9.4	9.8	9.7	9.9
	Classification Accuracy (%)	99.2	99.2	94.5	99.4	99.8	100	98.9	99.7	99.4	99.7
25	Precision (%)	98.1	96.1	97.4	94.6	99.4	99	95.2	98.6	96.8	98.1
	Recall (%)	94.5	96.4	97.1	99.6	98.9	99	93.8	98.3	97.5	99.1
	F1-score (%)	96.3	96.3	97.3	97.1	99.2	99	94.5	98.5	97.2	98.6

Experiment Three: [Hidden Neuron=25]

Table 4 shows how to classify the samples sound frames versus with four confusion matrix parameters. For example, Class 1, has total 2730 sound sample frames stored in sound database extracted from his voice. From those, the truly classified samples are the actual Class 1 predicted correctly 2579 (TP). The 50 (FP) frame samples are classified incorrectly. The 151 (FN) frame samples of actual Class 1 are predicted incorrectly by otherclasses.The23229(TN)framesampledneitheractualclassnorpredictedclasswhich means truly predicted incorrectly from the total samples**26009**.

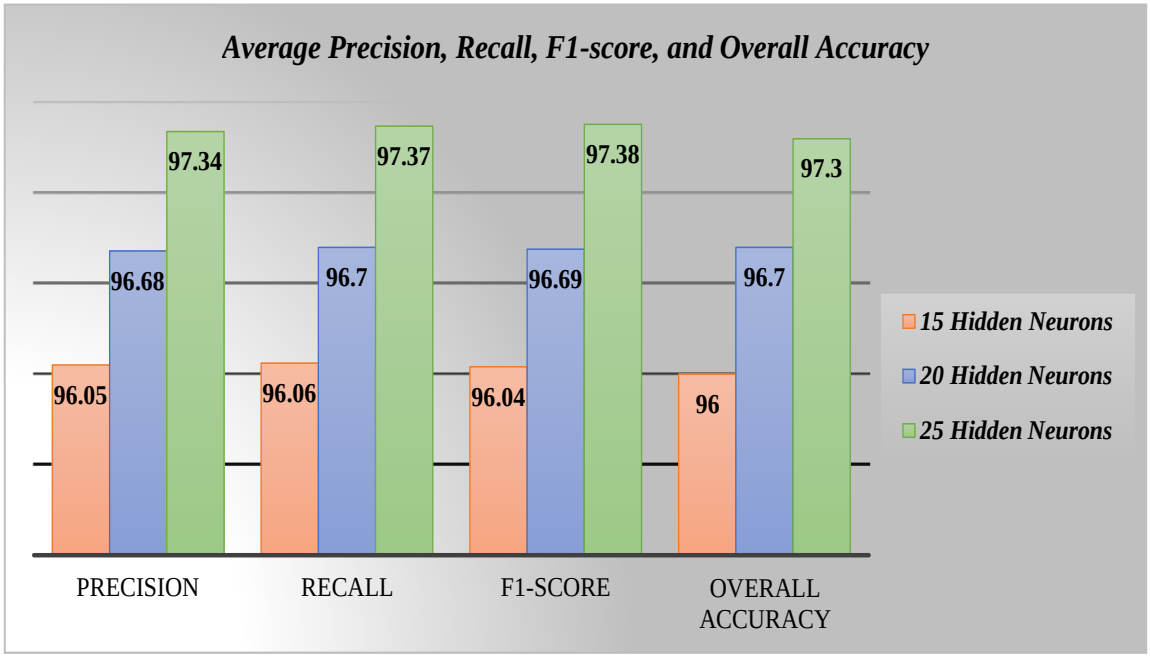

Figure 7. Average Precision, Recall, F1-score, and overall Accuracy of all Experiments

Here, the prototype has been tested using testing speeches that are preprocessed separately. 20 testing speeches are prepared for each speaker. So, one speaker has been tested 20 times with his/her respective prepared test speeches.

Table 5. Real Time Testing Recognition Results

Speakers Experiments	Result	AA	BB	BM	FT	MB	MD	MK	SZ	UAA	YN
Experiment 1	Correct	17	17	18	17	18	18	17	18	18	18
	Incorrect	3	3	2	3	2	2	3	2	2	2
Experiment 2	Correct	18	17	18	19	19	18	17	18	19	19
	Incorrect	2	3	2	1	1	2	3	2	1	1
Experiment 3	Correct	18	18	19	20	20	19	18	20	18	19
	Incorrect	2	2	1	0	0	1	2	0	2	1

5. Result Discussions

In the experiments and findings section, various findings are presented. The findings have addressed the problems in research question. The first research question was realizing whether ANN has promising performance for speaker recognition. Definitely, ANN has promising performance for speaker recognition. The indication is separately discussed later on. First, based on the confusion matrix on the overall accuracy is 96.0%, 96.7%, and 97.3% respectively for the three experiments. This shows how often the classifier is correct. Second, based on table 4 which presented the performance in terms of TP, TN FP, and FN, the approach showed good results in classifying objects correctly and incorrectly. They were a few objects that are classified wrongly. Third, based on table 5 which presented the performance in terms of precision, recall, and F1-score, the approach showed good results. Finally, based on the real time testing on table 5, the trained neural network provides a promising result in recognition of untrained speeches of registered speakers.

Later exploring internal factors that could improve the performance of the model and recognition. Selecting appropriate number of hidden layers and neurons is still a gap. There is no common standard approach, various approaches are listed in[12], from them rule of thumb is used in this study. So, used three different number of hidden neurons (15, 20, and 25) randomly. The findings obtained using this three experiments are presented

separately. Basically, the response times also increases during training the networks when increasing the number of the hidden neurons.

It is not appropriate to compare and contrast this study with the three previously done papers because to compare and contrast, researches should be in common background or equivalent environments. [5] and [6] used 50 and 100 speakers directly recorded speeches respectively, in this study the speeches are collected from multimedia and conducted using 10 speakers. But, the good thing is in [5], the researcher presented the performance per number of speakers (10, 20, 30, 40, 50). For 10 speaker, he achieved 83.2% and 87% accuracy using VQ and GMM respectively. In the same manner in [6], the researcher presented the performance per number of speakers (30, 60, 90). For 30 speakers and he achieved 70% and 66% accuracy using VQ and GMM respectively. Therefore, based on this analysis, our study achieved better performance than the two previously worked papers. However, it couldn't mean that this modeling technique has better performance than others for large number of speakers because this study is conducted for only 10 speakers.

At the end, the study introduce a novel attempt of modeling techniques for Amharic language speaker recognition research. And showed that ANN performs better than others on this context.

Table 6. Comparison of this research with other related researches

Author (s) and Year	Title	Feature Extraction Technique	Modeling Technique, Tools	Findings
Aykefam Azene, 2015	Text-independent speaker identification system for Amharic language.	MFCC	VQ & GMM, MATLAB	74.2% & 84.3% accuracy had been achieved for 50 speakers using VQ & GMMs respectively.
				For 10 speaker, he achieved 83.2% and 87% accuracy using VQ and GMM respectively.
Abrham Debasu, 2017	Automatic Text Independent Amharic Language Speaker Recognition in Noisy Environment Using Hybrid Approaches of LPCC, MFCC and GFCC.	LPCC,	VQ & GMM, MATLAB	[77.2%, 70.9%, 69%] and [75.2%, 76.9%, 78%]
		MFCC and GFCC.		accuracy had been achieved for 30, 60, 90 speakers using VQ and GMMs respectively.
Abrham Debasu, 2017	Text Independent Amharic Language Dialect Recognition: A Hybrid Approach of VQ and GMM.	MFCC	VQ & GMM, MATLAB	85.9% and 92.7% accuracy had been achieved for 25 & 100 speakers respectively.
This Study	ANN Based Amharic Language Speaker Recognition	MFCC	ANN, MATLAB	96%, 96.7%, 97.3% using 15, 20, and 25 number of hidden neurons respectively for 10 speakers.

6. Conclusion

Biometric techniques are one of the modern advances in security systems. No longer requires of entering a password or a PIN which is difficult to remember. Physical characters of the person are used instead. This thesis has presented voiceprint as one of the most promising and useful technologies fitting to the biometric security.

The general issues and applications of speaker recognition is described in the introduction of this thesis in well manner. The goal of the thesis was enabling the environment to develop text-independent speaker recognition for Amharic language using a novel modeling technique which is not attempted for Amharic previously.

Total of 200 sample speeches dataset has been prepared from 10 famous people public speeches. MFCC feature extraction and ANN modeling technique is used to meet the goal. MFCC transformed the preprocessed speech signals in to feature vectors so that it will be appropriate for training the neural network. The feedforward neural network trained the feature vectors using Levenberg Marquadrt training algorithm with training epoch parameters of 100. Training speech dataset has been divided into 70%, 15%, 15% for training, validation and testing respectively using dividerand function, and tansig activation function has been used to convert input signal of a neuron in to output signal.

The study is conducted based on three experiments in changing the number of hidden neurons in to 15, 20, and 25. The experiments have been conducted and the maximum promising findings have been obtained. The

proposed modeling technique showed better performance than other techniques which are used in previously done researches.

Findings addressed the whole research questions raised in the study. Findings have been discussed in discussion section of chapter four. Since the third research question is general question, the answer is found after completion of the whole research process. Its' implication was exploring external cause that could reduce the performance of the model and recognition accuracy. For the performance improvements of model as well as recognition accuracy, there are various external factors.

From speech perspective;

- Way of speech collection (direct recorded or from speech database)
- Environment on which the speech is recorded (Noisy or Noise free)
- Emotion and health condition of the speaker.
- Accent of speaker.
- *From techniques perspective;*
- Appropriateness of preprocessing steps
- Selecting of best feature extraction technique
- Modeling using a technique that is best in discriminating.
- Criteria to evaluate the performance.

Every research context has its' own fitting requirements. It needs scientific searching in order to get appropriate requirements for specific research context. Still it is a big gap for the researcher to standardize common appropriate requirements that works for any kind of research context. Mimicking the nature is so difficult.

At the end, we witnessed that the architecture selected for the neural network in this thesis, which is a pattern recognition feed-forward neural network with one hidden layer containing 15, 20, and 25 neurons and an output layer containing 10 neurons, inputted 26 coefficients vector was effective and suitable for the identification process.

The study contributes a novel attempt of modeling technique for Amharic language speaker recognition. Since the technique achieved better performance than other modeling techniques that are used in previously worked papers, it gives a clue which modeling technique is best, when anyone who needs to implement practically. The availability of this study will create a chance for the coming researchers who have willingness in this research area in order to conduct comparative analysis on modeling techniques especially for Amharic language speaker recognition. Finally, even though it is not the objective of this study, after all it is the product of this study; the GUI prototype developed by the researcher could be used as a tool for the coming researchers who have willingness in this field of research based on MFCC and ANN. It simplifies searching for different source codes and incorporating all in one.

References

Z. Seifu, "Hidden Markov Model Based Large Vocabulary, Speaker Independent, Continuous Amharic Speech Recognition, Unpublished", MSc. Thesis sumitted to School of Graduate Studies Faculty of Informatics, Addis Ababa University", 2003.

S. Furui, "Pattern Recognition Letters Recent advances in speaker recognition", Pattern Recognit. Lett., vol. 18, pp. 859–872, 1997.

N. Singh and A. Agrawal, "Principle and Applications of Speaker Recognition Security System", no. June 2018.

S. A. Mahmood and L. E. George, "Speaker Identification Using Backpropagation Neural Network," Journal of Zankoy Sulaimani, December 2008

M. S. Sinith, et al. "A novel method for text-independent speaker identification using MFCC and GMM", International Conference Audio, Language Image Processing, IEEE, 2010.

M. Islam, et al. "A novel Approach for Text-Independent Speaker Identification Using Artificial Neural Network", International Journal of Innovative Research and Computer Communication Engineering, vol. 1, no. 4, pp. 838–844, 2013.

M. Islam, et al. "A novel Approach for Text-Independent Speaker Identification Using Artificial Neural Network", International Journal of Innovative Research and Computer Communication Engineering, vol. 1, no. 4, pp. 838–844, 2013.

A. Antony and R. Gopikakumari, "Speaker identification based on combination of MFCC and UMRT based features", 8th international conference in adnaces in computing and communication, 2018.

A. Azene, "Text-independent speaker identification for the amharic language ", MSc Thesis submitted to Bahirdar University, no. January, p. 117, 2015.

A. D. Mengistu, “Automatic Text Independent Amharic Language Speaker Recognition in Noisy Environment Using Hybrid Approaches of LPCC, MFCC, and GFCC”, International Journal of Advanced Studies in Computer Science and Engineering IJASCSE Volume 6 Issue 5, 2017

A. D. Mengistu and D. Melesew, “Text Independent Amharic Language Dialect Recognition: A Hybrid Approach of VQ and GM”, International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 10, no. 1, pp. 215– 222, 2017.

F. S. Panchal and M. Panchal, “Review on Methods of Selecting Number of Hidden Nodes in Artificial Neural Network”, International Journal of Computer Science and Mobile Computing, vol. 3, no. 11, pp. 455–464, 2014.