EAST WEST UNIVERSITY

Electrical And Electronic Engineering

Course Code and Title:
EEE 309: Digital Signal Processing (DSP)
Semester and Year - Fall 2022
Section - 01

Open-Ended Lab Manual
Experiment #5:

Open-ended lab to investigate characteristics of different commands/spoken words in Bangla and to design a simple Bangla speech recognition system

(Group - 01)
Students Name and ID :

S. M. Abu Nayem	2015-2-80-074
Shamsur Joha Shams	2018-3-80-032
Mubeen Abdullah	2019-1-80-072
Maisha Nawshin	2019-2-80-027
Kazi Sadia	2019-1-80-077

Name of the Lab Instructor :
Dr. Halima Begum
Assistant Professor

Submission Date: 18 December 2022

(i) Objective:

In this era, machines are becoming more user-friendly and pleasant to mankind. The researcher got succeeded in training machines to understand what an image describes, what kind of sentiment a text carries, what a speaker expresses, and many other things like a human being. Recognizing speech is one of the challenging parts. Every giant tech company like Google and Microsoft has its own speech recognition system. In this open-ended design lab, we work on a simple Bangla Speech Recognition system.

(ii) Research on speech recognition:

1. A Digital Personal Assistant using Bangla Voice Command Recognition and Face Detection:

[Research paper done by Dipankar Gupta, Emam Hossain, Mohammad Shahadat Hossain, Karl Andersson, and Sazzad Hossain]

In [1] Bangla short speech commands data set has been reported, where all the samples are taken in a real-life setting. Their Experimental results reveal that the Cross Correlation model shows better accuracy in recognizing Bangla short speech commands. They tested their model on a normal room environment(noiseless), classroom environment, and noisy- city road environment and got an accuracy of 83%, 83%, and 75% in noiseless, moderate, and noisy environments respectively. On average, their proposed Bangla digital personal assistant remarkably recognizes 80.34% of Bangla voice commands and responds within 2-3 seconds. [1]

2. Hidden Markov Model (HMM)

We have found a few works on Bangla speech recognition. To analyze HMM(Hidden Markov Model) we found HMM models become obsolete for large-scale acoustic modeling. We [3] have found some good work by Hasnat, Md, Jabir Mowla, and Mumit Khan. Applying a Hidden Markov Model Based classifier they attempted to recognize both Isolated and Continuous words. They use 100 distinct Bangla words. For isolated word recognition, they achieved 90% accuracy and 70% accuracy on speaker-dependent and speaker-independent testing respectively. Continuous speech recognition accuracy becomes 80% and 60% on speaker-dependent and speaker-independent testing respectively. But we found some disadvantages in this method though its accuracy is very good. The number of parameters to be evaluated is huge. So it needs a large data set for training and it is not practical to represent multiple overlapping features and long-term dependencies.[3]

3. Convolutional Neural Network (CNN) & Recurrent Neural Network (RNN)

This paper contains two methods for Bengali speech recognition systems. The first method is **Convolutional Neural Network (CNN)**. It tried to recognize the spoken Bengali language successfully. In this model, a neural network was implemented which contains 7 levels. The first level was the input level and the last level was the output level. The middle 5 levels were hidden and every layer reduces the weight of the previous layers at a rate of 0.25. CNN model could convert isolated speech signals into texts with an accuracy of 86.058%. The limitations of this method were the lack of recorded speech data and text data. The second method of this paper is **Recurrent Neural Network (RNN)**. For this method, the main challenge was it needed a powerful tool to train itself with thousands of samples. Recognition of the Bengali language is tougher to implement than in other languages such as English for this method. They could train the RNN model with 30,000 Bengali words and tried to build an efficient model which converted the speech signals into text transcriptions. The paper contains systems that can work offline. The paper did not need to identify phonemes and very less preprocessing was needed. [4]

(iii) General observations from the training samples of 'Uthao' and 'Namao'.

We have considered 5 train samples for "Uthao" and 5 train samples for "Namao". Then we got different values after doing cross-correlation with one "Uthao" test sample and one "Namao" test sample. The Cross-correlation values are inserted in the table below,

Table:01 (Cross-correlation Values of Different "Uthao" & "Namao" Samples

Serial	Sample Number	Test Data	Max Cross-correlation with Uthao (T-10)	Max Cross-correlation with Namao (T-1)
1	U-3	Uthao	6.04	4.47
2	U-5	Uthao	49.89	30.86
3	U-8	Uthao	22.52	12.43
4	U-9	Uthao	15.56	12.32
5	U-15	Uthao	21.56	13.57
6	D-1	Namao	32.09	36.61
7	**D-2**	**Namao**	**75.83**	**55.96**
8	D-3	Namao	47.94	56.88
9	D-11	Namao	39.32	43.37
10	**D-12**	**Namao**	**36.62**	**24.41**

From the data **Table: 01** we can observe that the maximum cross-correlation value between two “Uthao” samples is greater than the maximum cross-correlation value between “Uthao” and “Namao” samples. Moreover, the maximum cross-correlation value between two “Namao” samples is greater than the maximum cross-correlation value between the “Namao” and “Uthao” samples.

There is two error in **row # 7** and **row # 10**. In both cases, the maximum cross-correlation value between the “Namao” and “Uthao” samples is greater than the maximum cross-correlation value between two “Namao” samples.

MATLAB plots of Different Cross-correlation Values

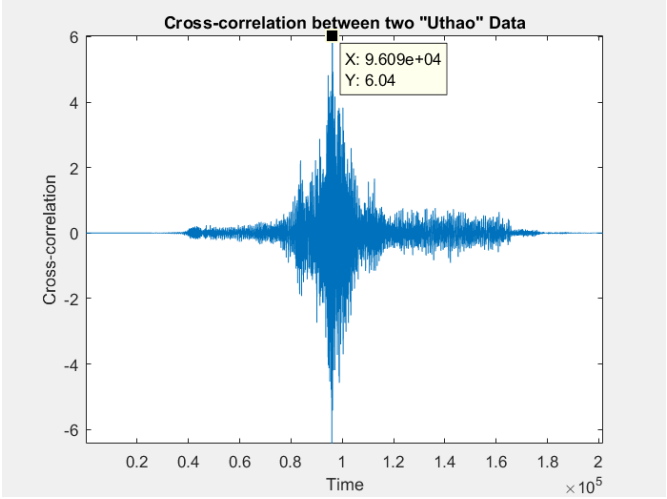

Fig 01_U3 : Cross-correlation between two "Uthao" Data

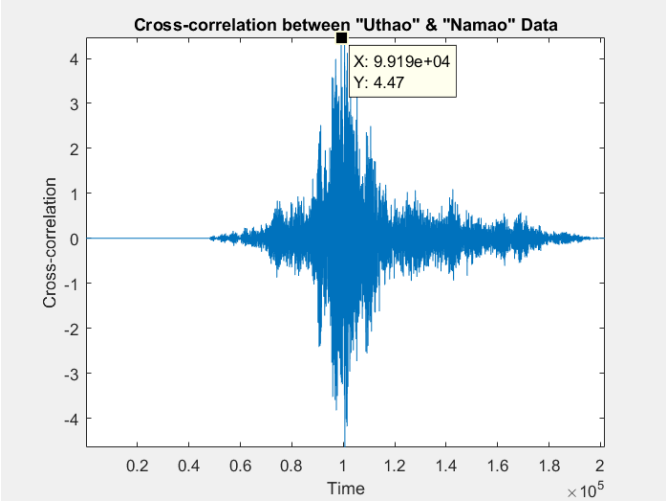

Fig 01_U3 :Cross-correlation between "Uthao" & "Namao" Data

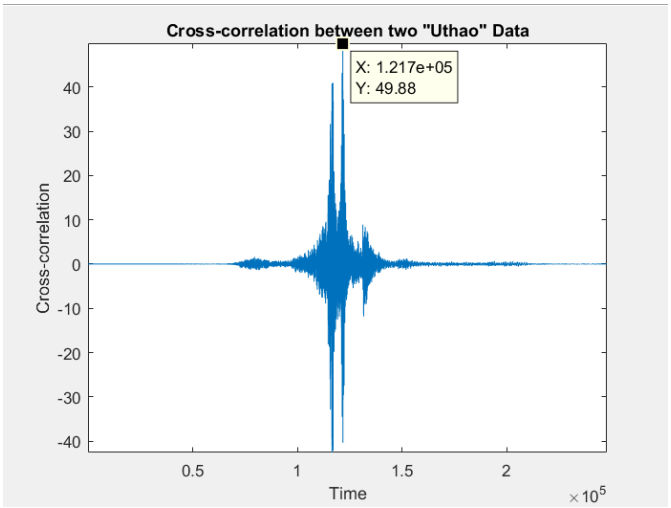

Fig 02_U5 : Cross-correlation between two "Uthao" Data

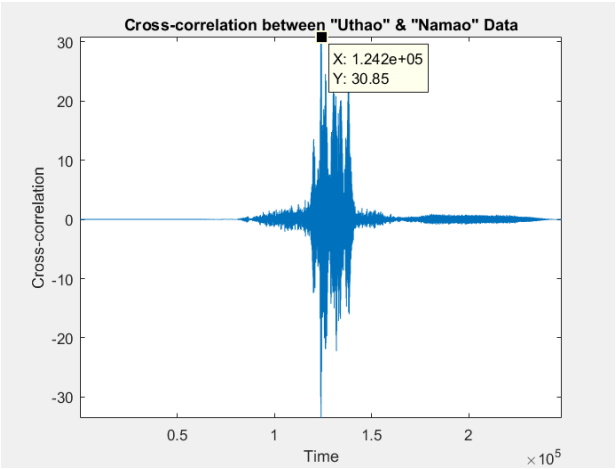

Fig 02_U5 :Cross-correlation between "Uthao" & "Namao" Data

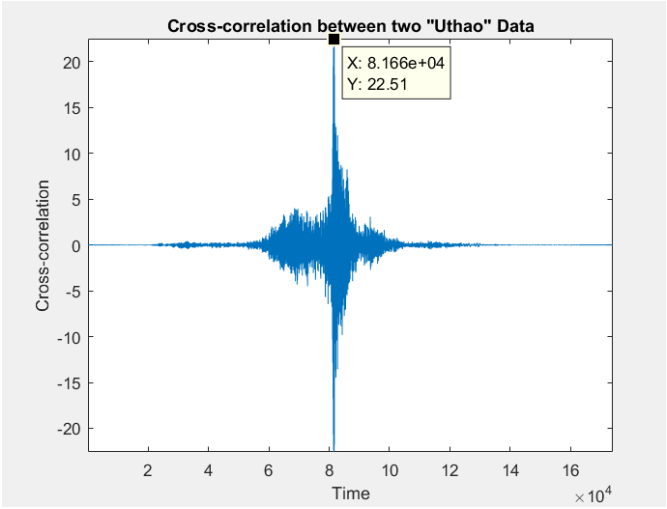

Fig 03_U8 :Cross-correlation between two "Uthao" Data

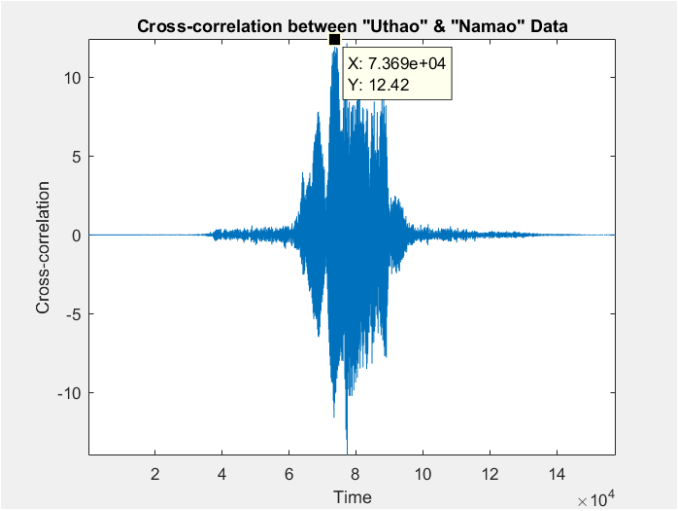

Fig 03_U8 :Cross-correlation between "Uthao" & "Namao" Data

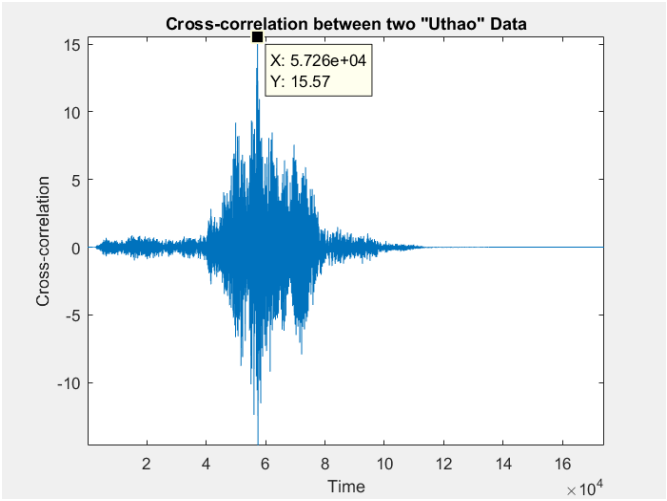

Fig 04__U9 :Cross-correlation between two "Uthao" Data

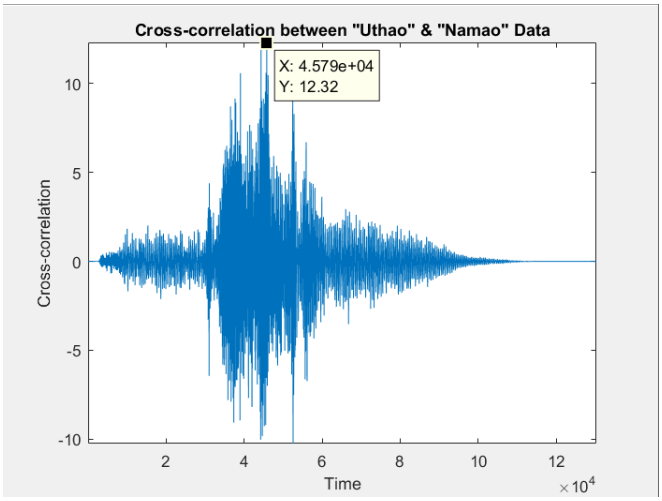

Fig 04_U9 :Cross-correlation between "Uthao" & "Namao" Data

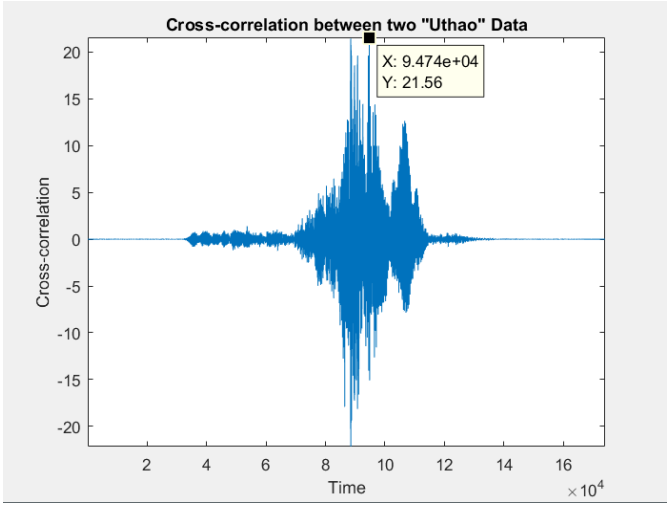

Fig 05_U15 :Cross-correlation between two "Uthao" Data

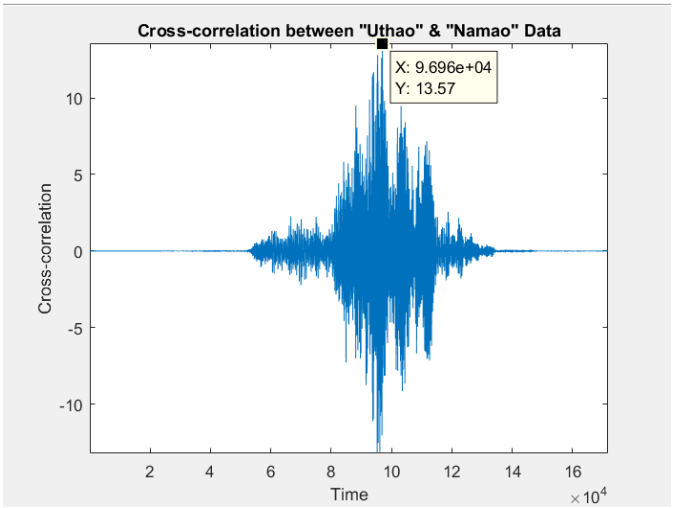

Fig 05_U15 :Cross-correlation between "Uthao" & "Namao" Data

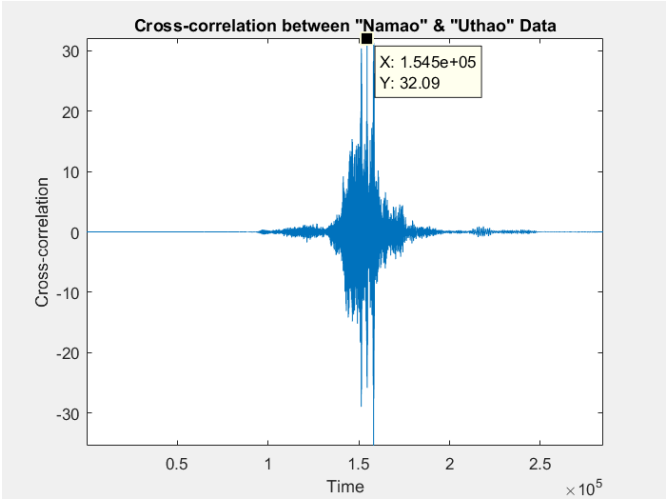

Fig 06_D1 :Cross-correlation between "Namao" & “Uthao” Data

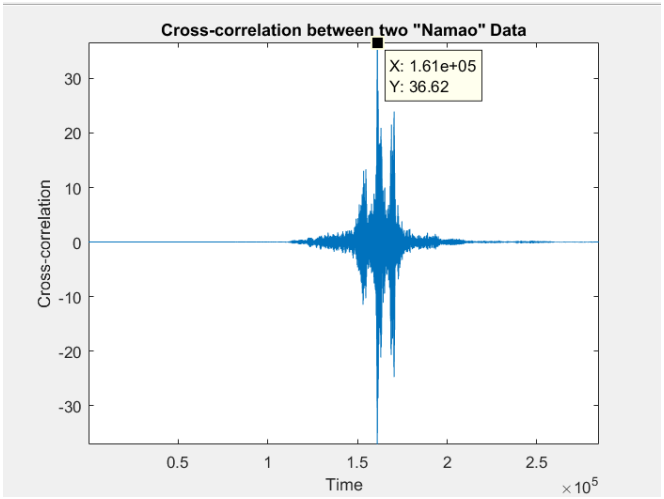

Fig 06_D1 :Cross-correlation between two "Namao" Data

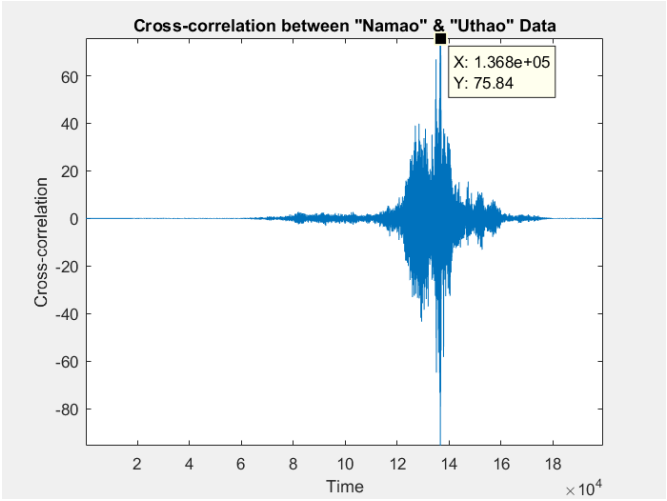

Fig 07_D2 :Cross-correlation between "Namao" & “Uthao” Data

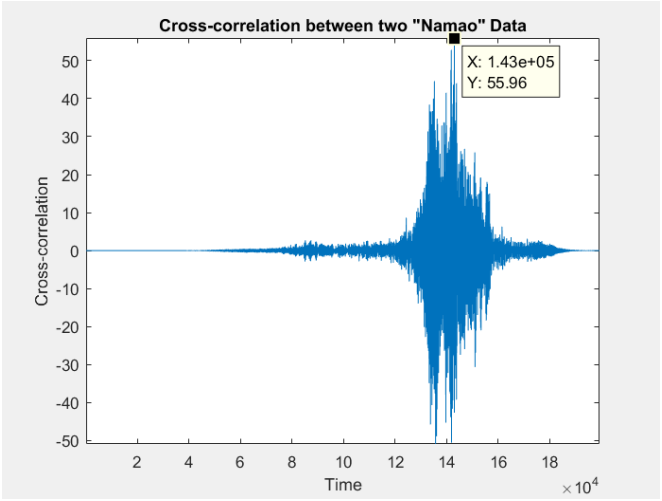

Fig 07_D2 :Cross-correlation between two "Namao" Data

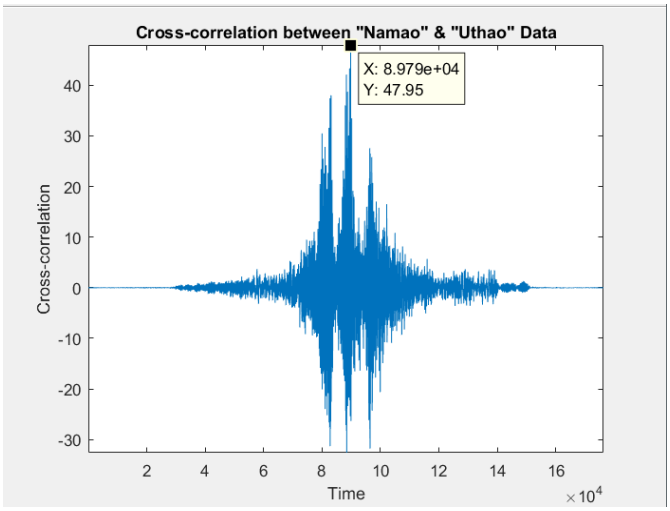

Fig 08_D3 :Cross-correlation between "Namao" & “Uthao” Data

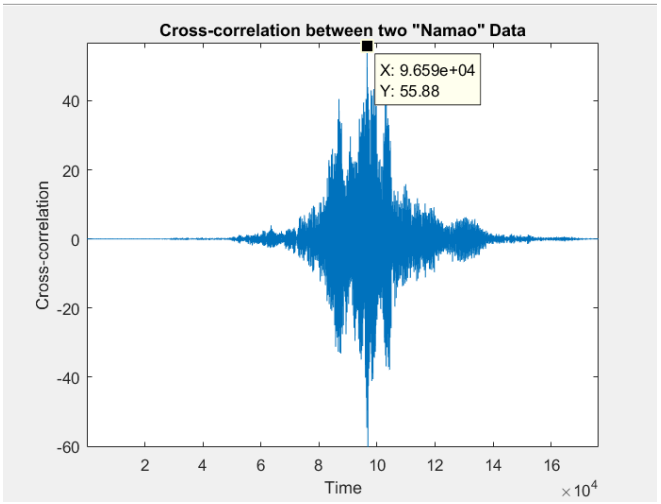

Fig 08_D3 :Cross-correlation between two "Namao" Data

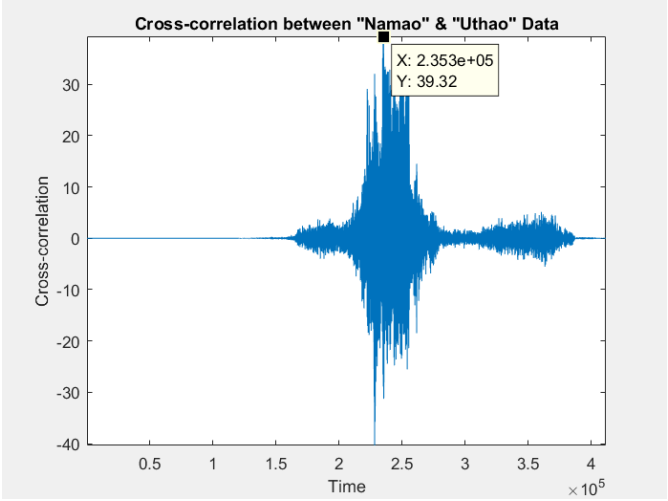

Fig 09_D11 :Cross-correlation between "Namao" & “Uthao” Data

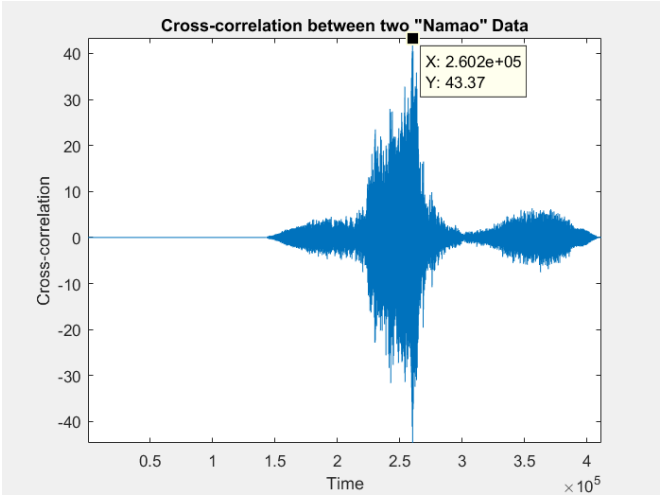

Fig 09_D11 :Cross-correlation between two "Namao" Data

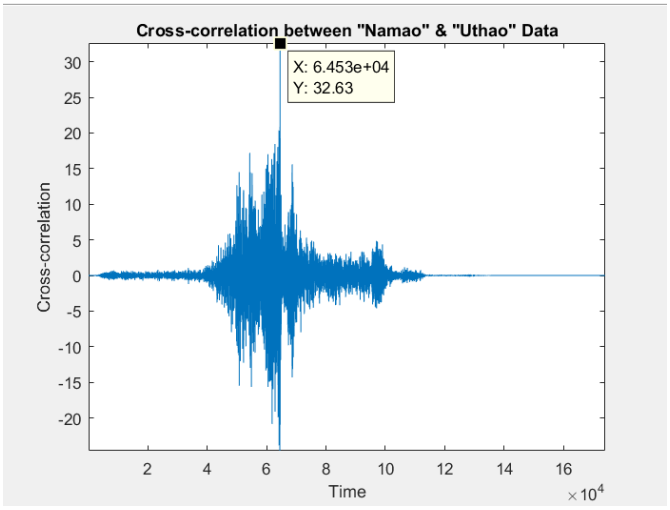

Fig 10_D12 :Cross-correlation between "Namao" & “Uthao” Data

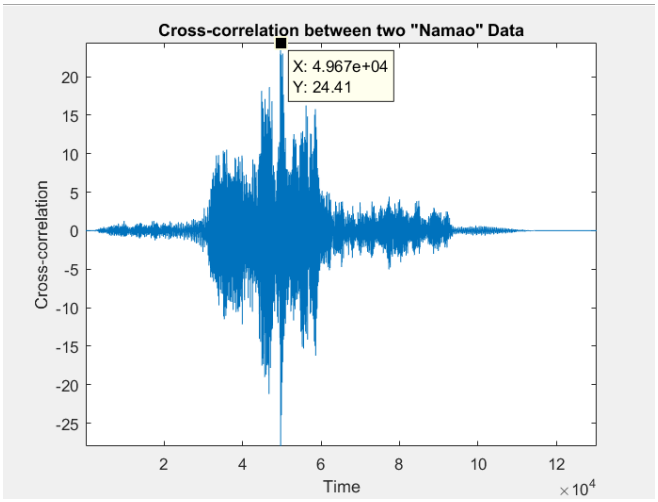

Fig 10_D12 :Cross-correlation between two "Namao" Data

(iv) Mention multiple feature extraction processes based on the observation and then your chosen feature extraction technique between these multiple techniques along with proper justification. You should also mention any required preprocessing of the signals.

In[2] our experiment we mentioned 4 different types of methods and analyzed those types of methods as we need to choose a method that is suitable for us. To analyze HMM (Hidden Markov Model) we found HMM models become obsolete large-scale acoustic modeling so it needs large-scale data sheets for training. Then we work on Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). Since convolutional neural networks are typically used for image classification. Therefore, this

method generally deals with high-dimensional data (images). We generally need specialized hardware (like a GPU) to expedite the training process. So we used the Cross-correlation method implemented in MATLAB to compare two or more signals and detect the most accurate one of them all. We have actually used cross-correlation to find similarities between our recorded Signal files and the testing signal. The cross-correlation method is comparatively easier than the other 3 methods mentioned above. Thus, we were able to develop a model where machines can differentiate between commands and act upon them. In communication and signal processing, cross-correlation is a measure of the similarity of two waveforms as a function of a time lag applied to one of them. It is commonly used for searching a long signal for a shorter, known feature. It also has applications in pattern recognition, single particle analysis, electron tomographic averaging, cryptanalysis, and neurophysiology. In our experiment, we used this for matching the test data in the easiest way.[2]

(v) Verification Algorithm/Flowchart :

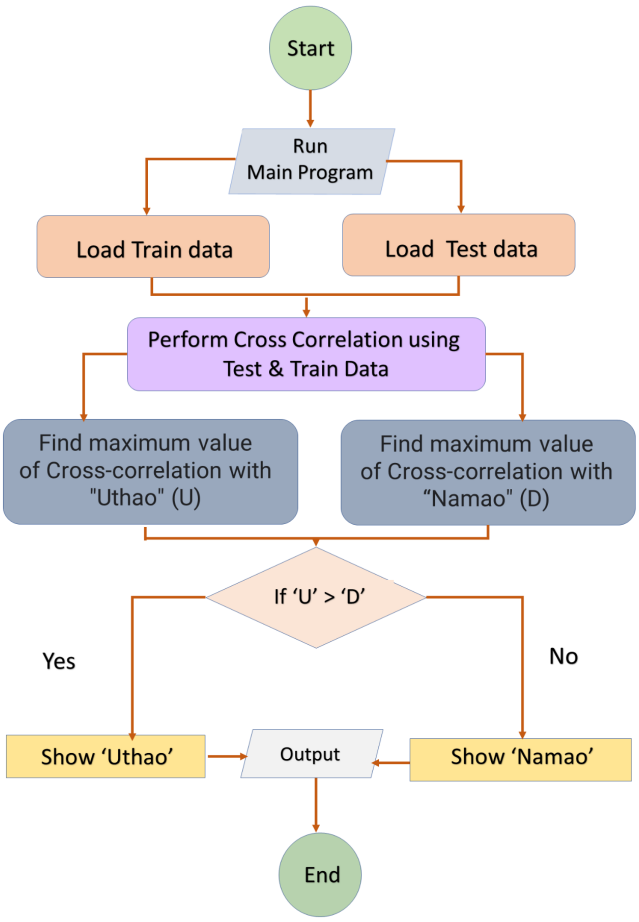

Figure-11: Verification Flow Chart

References:

[1] D. Gupta, E. Hossain, M. S. Hossain, K. Andersson, and S. Hossain, “A Digital Personal Assistant using Bangla Voice Command Recognition and Face Detection,” *2019 IEEE International Conference on Robotics, Automation, Artificial-intelligence and Internet-of-Things (RAAICON)*, Nov. 2019, doi: 10.1109/raaicon48939.2019.47.

[2] Muhammad, Ghulam & Alotaibi, Yousef & Huda, Mohammad. (2010). Automatic speech recognition for Bangla words. International Journal of Advanced Science and Technology, Vol.50,January,2010. 379 - 383. 10.1109/ICCIT.2009.5407267.

[3] M. Hasnat, J. Mowla, M. Khan et al., “Isolated and continuous Bangla speech recognition: implementation, performance and application perspective,” 2007.

[4] J. Islam, M. Mubassira, Md. R. Islam, and A. K. Das, “A Speech Recognition System for Bengali Language using Recurrent Neural Network,” 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Feb. 2019, doi: 10.1109/ccoms.2019.8821629.