

Speech Recognition Using MATLAB and Cross-Correlation Technique

Ledisi G. Kabari, Marcus B. Chigoziri

Abstract—Speech is a prominent communication method among humans, whereas the communication between human and computers were based on text user interface and graphic user interface. Speech recognition is used in almost every security project where you need to speak and tell your password to computer and is also used for automation. This paper demonstrates a model that enhances technological advancement where humans and computers interact via voice user interface. In developing the model, cross correlation was implemented in MATLAB to compare two or more signals and detect the most accurate one of the all. We are actually used cross correlation to find similarity between our recorded signal files and the testing signal. Thus we were able to develop a model where machines can differentiate between commands and act upon them.

Index Terms—Cross Correlation, MATLAB, Recorded Signal, Signal Processing, Speech Recognition, Testing Signal.

I. INTRODUCTION

Speech is the most prominent means of communication amongst humans. Human-to-human interaction is based on speech, emotion and gestures, thereby making it a lot easier to understand one another. On the other hand, the communication between humans and computers is based on either Text User Interface (TUI) or Graphic User Interface(GUI). It is a lot easier for us humans to recognize a person's voice than computers. Hence, speech recognition in machine learning is a game changer as developing machines that can understand and uniquely identify a person's voice would make Human-Computer interaction more intriguing.

Speech recognition is one of the next generation technologies for human-computer interaction. Speech recognition has been researched since the late 1950s but due to its computational complexity and limited computing capabilities of the last few decades, its progress has been impeded. In laboratory settings automatic speech recognition systems (ASR) have achieved high levels of recognition accuracies, which tend to degrade in real world environment [1].

In today era speech technologies play an important role. This technology is commercially and easily available for a different uses. These technologies make machines respond correctly and it provides valuable services. In modern era, no one wants to reveal his identity due to security purposes. So, Speech can be used for the identification of person

because every person has different speech characteristic. Thus with the different information in speech waves we can easily identify the speaker [2].

II. LITERATURE REVIEW

A. Speech Recognition

Speech recognition methods can be divided into text-independent and text dependent methods. In a text independent system, speaker models capture characteristics of somebody's speech, which show up irrespective of what one is saying. In a text-dependent system, on the other hand, the recognition of the speaker's identity is based on his or her speaking one or more specific phrases, like passwords, card numbers, PIN codes, etc. [3].

Every speech recognition application is designed to accomplish a specific task. Examples include: to recognize the digits zero through nine and the words “yes” and “no” over the telephone, to enable bedridden patients to control the positioning of their beds, or to implement a VAT (voice-activated typewriter). Once a task is defined, a speech recognizer is chosen or designed for the task [2]

The task of speech recognition is to convert speech into a sequence of words by a computer program. Speech recognition applications enable people to use speech as another input mode to interact with applications with ease and effectively. Speech recognition interfaces in native language will enable the illiterate/semi-literate people to use the technology to greater extent without the knowledge of operating with computer keyboard or stylus [4].

There are different modes available for Speech Recognition System: Speaker Dependent / Independent System: It must be trained in order to recognize accurately what has been said. To train a system, Speaker is asked to record predefined words or sentences that will be analyzed and that results will be stored.

Isolated Word Recognition: It is Simplest mode and less greedy in terms of CPU requirement. Word is surrounded by silence so that boundaries are well known.

Continuous Speech Recognition: It assumes that system is able to recognize a sequence of words in a sentence.

Keyword Spotting: It is able to identify in a sentence a word corresponding to a particular command. Created to cover the gap between isolated and continuous System.

Vocabulary Size: Larger the vocabulary the system can make more errors. So vocabulary size matters.

B. Cross Correlation

Cross correlation is a standard method of measuring the similarities/relationships between two signals. It is a measure of similarity of two series as a function of the

Published on August 6, 2019.

L. G. Kabari is with Ken Saro-Wiwa Polytechnic, Bori, Nigeria. (e-mail: ledisiogiokkabari@yahoo.com).

M. B. Chigoziri is with Ignatius Ajuru University of Education, Port Harcourt, Nigeria. (e-mail: chigozirimarcus@yahoo.com).

displacement of one relative to the other. There are some cases where it is necessary to compare one reference signal with one or more signals to determine the similarities between signals and to determine additional information based on their relationships. This is also known as a sliding dot product or sliding inner-product. It is commonly used for searching a long signal for a shorter, known feature. It has applications in pattern recognition, single particle analysis, electron tomography, averaging, cryptanalysis, and neurophysiology [5].

The term cross-correlation is utilized for alluding to the relationships between the sections of two arbitrary vectors X and Y, while the connections of an irregular vector X are thought to be simply the connections between simply the passages of X, those shaping the connection lattice (network of connections) of X.

In MATLAB the cross correlation function is `xcorr` for sequence for a random process which includes autocorrelation. Thus the Syntax for Correlation in MATLAB is derived as `r = xcorr(x,y)`. `r = xcorr(x,y)` returns the cross-correlation of two discrete-time sequences, x and y. Cross-correlation measures the closeness amongst x and moved (slacked) duplicates of y as a component of the slack. In the event that x and y have diverse lengths, the capacity annexes zeros toward the finish of the shorter vector so it has a similar length, N, as the other [5].

III. METHODOLOGY

Five recorded wav sound samples were stored in the database and we wish to recognize them using a correlation test.wav in MATLAB.

STEP1: We create a test file command as shown in figure 1.

```

1 - function speechrecognition(filename)
2 -     voice=wavread(filename);
3 -     x=voice;
4 -     x=x';
5 -     x=x(1,:);
6 -     x=x';

```

Fig. 1: Test Function Code

STEP2: The code shown in figure 2 represents the sample one.wav, do this repeatedly for as many samples you want to be present. In this case, we have five samples. Therefore, we repeated this code five times with every place that say 1 replaced with 2, 3, 4 and 5 respectively.

STEP3: We note that the test has x value, it is this value that is used to compare the y values of the sample. Hence, `Z1=xcorr(x,y)`;

STEP4: We create a conditional statement where `m6=300`; and if `m=max` the machine will sound the allowed

signal. This simply means, if `m<=M1`, sound one.wav speech else, it continues accordingly until it finds a perfect signal else, it sounds the denied.wav signal.

```

7 - y1=wavread('one.wav');
8 - y1=y1';
9 - y1=y1(1,:);
10 - y1=y1';
11 - z1=xcorr(x,y1);
12 - m1=max(z1);
13 - l1=length(z1);
14 - t1=-((l1-1)/2):((l1-1)/2);
15 - t1=t1';
16 - %subplot(3,2,1);
17 - plot(t1,z1);

```

Fig. 2: Typical Code for sample file

```

66 - m6=300;
67 - a=[m1 m2 m3 m4 m5 m6];
68 - m=max(a);
69 - h=wavread('allow.wav');
70 - if m<=m1
71 -     soundsc(wavread('one.wav'),50000)
72 -     soundsc(h,50000)
73 - elseif m<=m2
74 -     soundsc(wavread('two.wav'),50000)
75 -     soundsc(h,50000)
76 - elseif m<=m3
77 -     soundsc(wavread('three.wav'),50000)
78 -     soundsc(h,50000)
79 - elseif m<=m4
80 -     soundsc(wavread('four.wav'),50000)
81 -     soundsc(h,50000)
82 - elseif m<=m5
83 -     soundsc(wavread('five.wav'),50000)
84 -     soundsc(h,50000)
85 - else
86 -     soundsc(wavread('denied.wav'),50000)
87 -
88 - end

```

Figure 3: Condition Statement

IV. SPEECH RECOGNITION TEST AND RESULTS

In the command window, we typed the command "speechrecognition('test.wav') and hit the enter button to get the spectrum graphs. In this particular case, test.wav and test2.wav represents the number 2 and 3 respectively.

So we run a test for test2.wav. Now because test2.wav sounds the number 3, figure 3 as shown in figure 4 is the most accurate spectrum.

We repeated the step for test.wav which represents number 2. From figure 5 we can see that graph 2 is more accurate as test.wav sound is two.

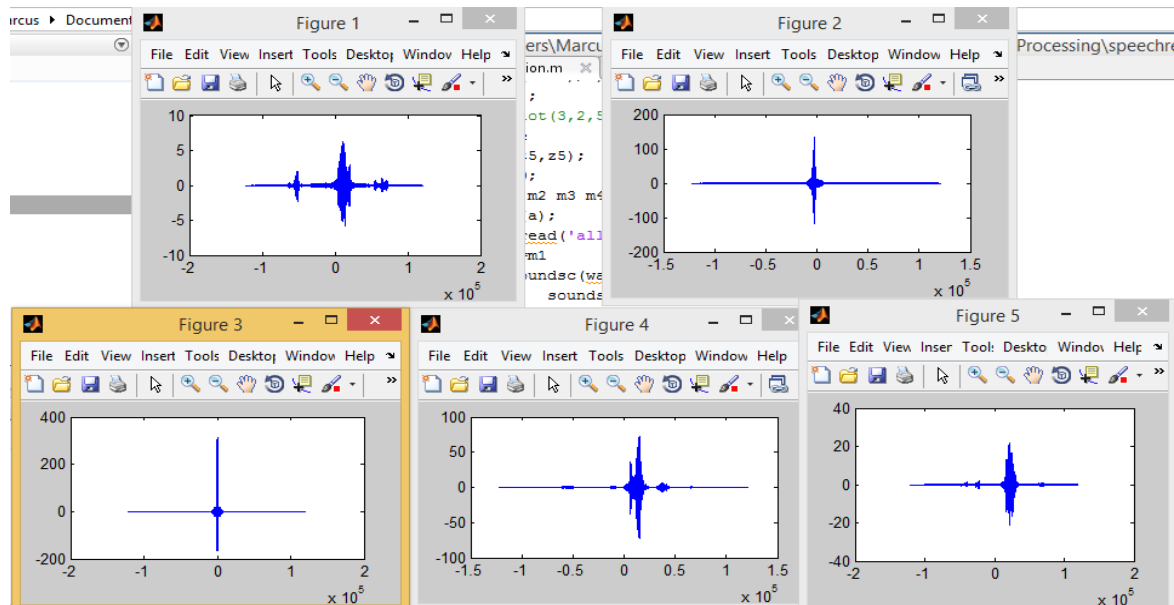


Figure 4: Spectrum Graphs

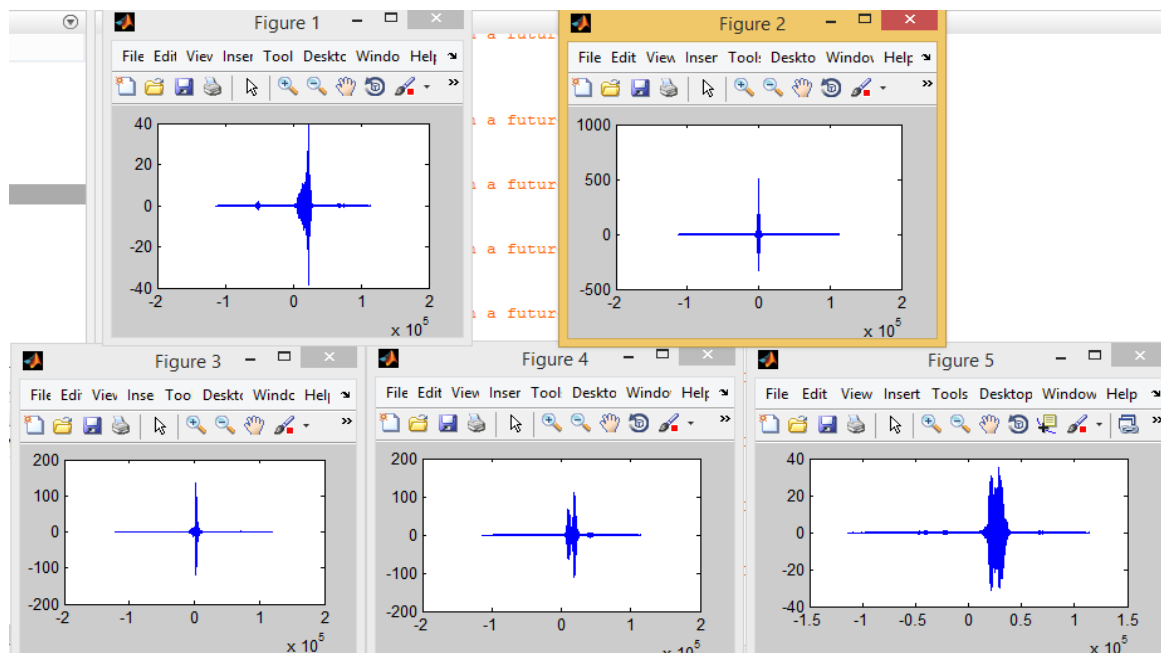


Figure 5: Spectrum Graphs

V. CONCLUSION

In this work, the five recorded wav audio files were used to demonstrate the speech recognition. Using the cross correlation method to find the similarities between the recorded audio files, we were able to develop a model where machines can differentiate between commands and act upon them. It clearly shows that machines can understand and interact with humans fluently, although they are very sensitive to noise and pronunciations. Figure 4 and 5 demonstrates a typical scenario where the machine recognizes one out of the audio files.

REFERENCES

- [1] P. Ranjeet, T. Prakash, S. Amruta and S. Monali, "Automatic Speech Recognition System – A review, Imperial Journal of Interdisciplinary Research (IJIR) Vol-2, Issue-3, 2016, 165-169.
- [2] R. Paul, R. Beniwal, R. Kumar and R. Saini, "A Review on Speech Recognition Methods", International Journal on Future Revolution in Computer Science & Communication Engineering, Volume: 4 Issue 2, 292– 298, 2018.
- [3] M. Saundade and P. Kurlle, "Speech Recognition using Digital Signal Processing", International Journal of Electronics, Communication & Soft Computing Science and Engineering ISSN: 2277-9477, Volume 2, Issue 6, 2013.
- [4] S. Karpagavalli and E. Chandra, "Review on Automatic Speech Recognition Architecture and Approaches", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.9, No.4, (2016), pp.393–404, <http://dx.doi.org/10.14257/ijsp.2016.9.4.34>.
- [5] A. Gupta, P. Raibagkar, A. Palsokar, "Speech Recognition Using Correlation Technique", International Journal of Current Trends in Engineering & Research (IJCTER)e-ISSN 2455–1392 Volume 3 Issue 6, June 2017 pp. 82–89.