De La Salle University - Manila

**Final Project**

# Audio Identification using Pitch pattern recognition through Digital Signal Processing

Date Performed: March 28, 2023

Date Submitted: April 17, 2023

Submitted to:

**Ruiz, Ramon Stephen L.**

SIGNALS, SPECTRA, AND SIGNAL PROCESSING LABORATORY

**LBYEC4A-EK1**

Submitted by:

**De Vera, Renato Angelo DL.**

**Leoncio, Abigail Joy T.**

**Lucmayon, Ethan James T.**

## I. Abstract

Electronic devices have played a crucial role in bringing people together around the world as technology has advanced in modern civilization. With this interconnectedness, technological communication has become mundane in our society through these electronic devices and equipment. With its integration with technology, although humans have little to no trouble distinguishing one voice from another, technology in solitary will find the process of voice recognition as speech has different tones, intonations, and pitch. This study aims to apply pitch pattern recognition in a speech type of audial identification through digital signal processing. Discrete cosine transform will be the main focus through the application of Matlab as a function initialized for the audio recordings. The project aims to train technology in recognizing voice by sound and pattern as one of the few ways machines can analyze the information omitted by humans.


## II. Introduction

Machines and technologies have developed in today's technologically advanced world to acquire and evaluate data entered by humans through a variety of sources, including aural and visual methods. By identifying voices by sound and pattern, machines can interpret information from humans in a variety of methods. In recent years, there has been a growing demand for reliable and efficient methods of audio identification, with applications in fields such as music, speech recognition, and security.

Creating a voice recognition system that can identify particular speech will be an implementation that can be adapted to numerous machines and devices. Methods for this system which can be beneficial in terms of efficiency include comparing the amplitudes entered by the user into a certain database to any audio that is to be succeedingly inputted. By doing so, it can identify the source of the audio, whether it be a specific musical instrument, a human voice, or even background noise. In this paper, we present the results of our research, detailing the development and implementation of our program, as well as its performance in identifying a variety of audio signals.

## III. Problem Statement

As the world grows more connected, audial transmission has grown to be an essential component of communications in the twenty-first century. With the development of technology in society, humans have become more reliant on the devices that the voices that humans omit.

Along with its development is technology's adaptation analyzes the information of and from these voices which are set and initialized by humans as well. With that being said, voice recognition has multiple applications yet can lack efficiency. Certain methods such as the discrete cosine transform can be applied for a more accurate representation in terms of pitch pattern matching and comparison.

## IV. Objective

To transmit audio information, audio signals have become a crucial component of communications equipment. Audio signals tend to become tarnished with various sorts of noise that degrade the sound quality. With the use of MATLAB, the project aims to identify specific speech through pitch pattern recognition, by developing a voice recognition system that can identify and train various speech patterns involving discrete cosine transform functions.

## V. Theoretical Consideration

Digital signal processing and pitch pattern recognition for audio identification includes several theoretical concerns. The comprehensive process by which the human voice produces various sounds in various environments, having a deep understanding of the use of machine learning and pattern recognition techniques, and understanding discrete cosine transform and fast Fourier transform are just a few of these. The methods DCT and FFT use to convert audio signals from the time domain to the frequency domain enable the identification of pitch patterns in audio streams [1]. Another thing to think about is the different voices people can make in different locations, which can be identified by pitch (sound's frequency) and volume (sound's loudness). Specific speech patterns can be recognized using these characteristics of the human voice. Additionally, a successful audio identification system requires the application of machine learning and pattern recognition techniques to acquaint the program or system with distinct speech patterns, found in the database, through the comparison of amplitude and pitch patterns of the audio stream [2].

### Discrete Cosine Transform

Similar to the discrete Fourier transform, the discrete cosine transform manipulates a signal to be converted into the frequency domain. The discrete cosine transform is represented by the equation [1]:

$$F(u) = \left(\frac{2}{N}\right)^{\frac{1}{2}} \sum_{k=0}^{N-1} \Delta(k) cos\left[\frac{\pi * u}{2N}(2k + 1)\right] f(i)$$

The discrete cosine transform utilizes cosines in which are components for real vectors. Discrete cosine transforms are classified as sinusoidal unitary transforms. Such transforms are linear and defined by discrete cosine functions [2].

## VI. Methodology

The program will function on executing the voice recognition by relying on the amplitude obtained from each audio file. Since the group will have a directory for the audio, they will be loading the input audio and loading the directory from its file location. They will then plan on differentiating the amplitude of the audio using the discrete cosine transform function. The maxima of the signals are to be acquired and to be matched [6].

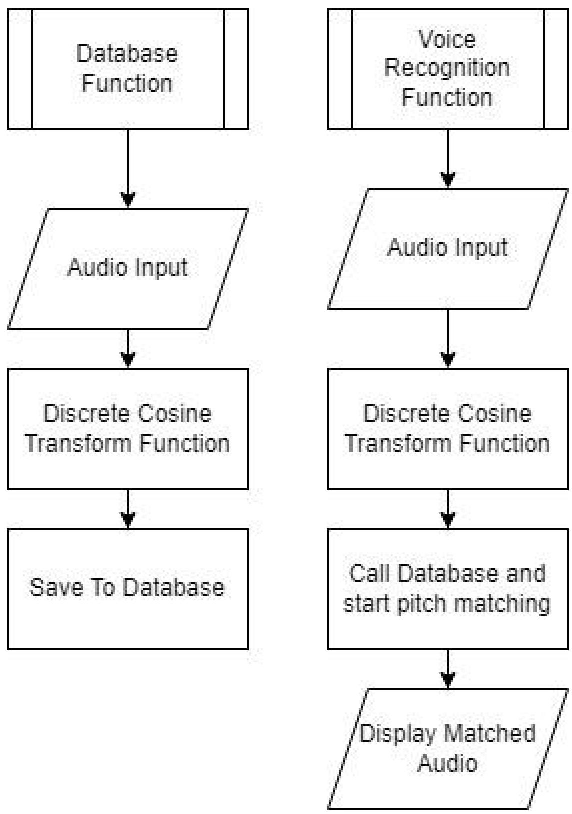


**Figure 1. Method Flowchart**

### Training

In training, a set of voice recordings from each member are run into the database function and saved into the database to have reference audios for the test input. By saving multiple reference data, the program is trained to have a variety of choices when the operation eventually runs the actual feature matching.

**Matching or Testing**

Matching or testing is the actual comparison of the input test data to the multiple reference test data from the database. The program should be able to identify through visual means the similarity between the test input data from the counterpart reference data.

**Codes**

```
function PitchMatch=AudioTransform(aud)
F=dct(aud(:,1));
m=max(real (F));
PitchMatch = find(real(F) == m,1);
end
```

**Code 1.**

The code in Code 1 represents the main discrete cosine function that will be used throughout the program. The function finds the discrete cosine function of a certain audio recording using the "dct()" command, as well as the maxima frequency in the real output.

```
for a=[1 2 3]
filename=strcat('C:\Users\ethan\Documents\DLSU\WAVDATABASE\A',num2str(a),'.wav');
    aud=audioread(filename);
    FE=AudioTransform(aud);

try
    load database
    F=[F;FE];
    FN=[FN;a];
    database=[zeros;F;FN];
    save database.mat database F FN
catch
    F=FE;
    FN=a;
    save database F FN
end
end
```

**Code 2.**

| Name | # | Title | Contributing artists | Album |
|---|---|---|---|---|
| A1 | | | | |
| A2 | | | | |
| A3 | | | | |

**Folder 1.**

Code 2 performs the construction of the matching database with which the test audio recordings will be matched. A for loop is constructed for the continuous saving of the three audio recordings found in Folder 1 to the database as matching data. The 'AudioTransform' function in Code 1 is already utilized in such a way that the discrete cosine transform of the three matching audio files is to be computed before being saved in the database.

```
namefile=input('Input filename: ','s');
takefile=strcat('C:\Users\ethan\Documents\DLSU\WAVTEST\B',namefile,'.wav');
readfile=audioread(takefile);
FE=AudioTransform(readfile);
load database
D=[];
for a=1:size(F,1)
    d=sum(abs(F(a)-FE));
    D=[D d];
end
small=Inf;
id=-1;
for a=1:length(D)
    if (D(a)<small)
        small=D(a);
        id=a;
    end
end
```

**Code 3.**

| Name | # | Title | Contributing artists | Album |
|---|---|---|---|---|
| B1 | | | | |
| B2 | | | | |
| B3 | | | | |

**Folder 2.**

Code 3 is where the audio files to be tested for matching are going to be inputted. The files for said audio files are located in Folder 2. To make the program more simple and seamless the same audio files from the database are also to be utilized in the input audio files. The computation of the discrete cosine transform will also be performed for the inputted audio files. The way the matching works is that in Code 1 ‘FN’ is established to be a 29x1 matrix with the elements 0,3,1,2,3,1,2,3,1,2…Then, ‘F’ is the 29x1 matrix of the values of the database recordings itself from the calculated discrete cosine transform. The length of F is returned in the first dimension. The sum of the elements of ‘F’ and the value of the inputted signal ‘FE’ is determined and stored in ‘D’. Since the inputted recording is subtracted from ‘F’ the matched element in ‘D’ should have a value of ‘0’. Therefore, a for loop is created to detect the ‘0’ in which the inputted audio recording is matched with the matching recording from the database.

```
file_number=FN(id);
indicate=strcat('File No. of' , namefile , 'is: ');
disp(indicate)
file_number

[j,k]=audioread(takefile);
subplot(2,1,1);
plot(abs(j(:,1)));
xlabel('Time');
ylabel('Amplitude');
title('TEST');

subplot(2,1,2);
original=strcat('C:\Users\ethan\Documents\DLSU\WAVDATABASE\A',num2str(file_number),'.wav');
[n,m]=audioread(original);
plot(abs(n(:,1)))
xlabel('Time')
ylabel('Amplitude')
title('MATCH')
```

**Code 4.**

Code 4 is simply indicating the file number the inputted audio recording is matched with as well as the amplitude plot of the inputted audio recording and the matched audio recording from the database to verify the matching.

## VII. Results

```
File No. of1is:
file_number = 1
```

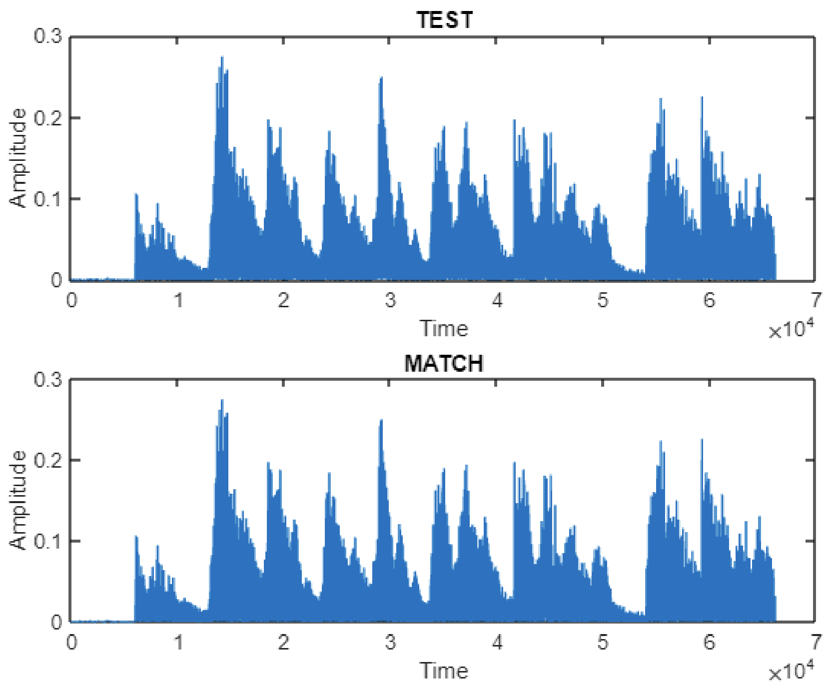


**Figure 1. Plot of file B1 matched with File A1**

```
File No. of2is:
file_number = 2
```

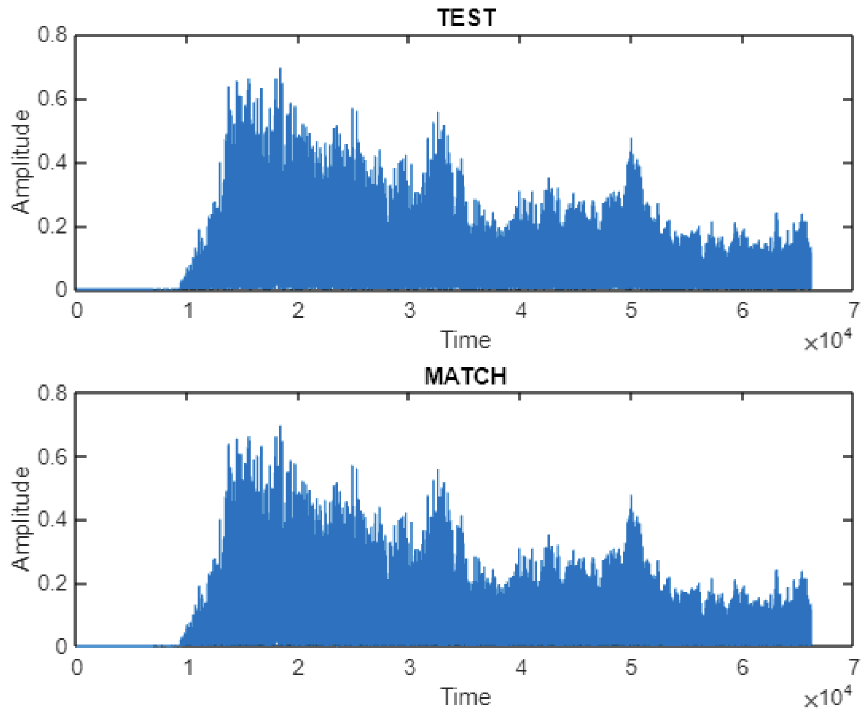


**Figure 2. Plot of file B2 matched with file A2**

```
File No. of3is:
file_number = 3
```

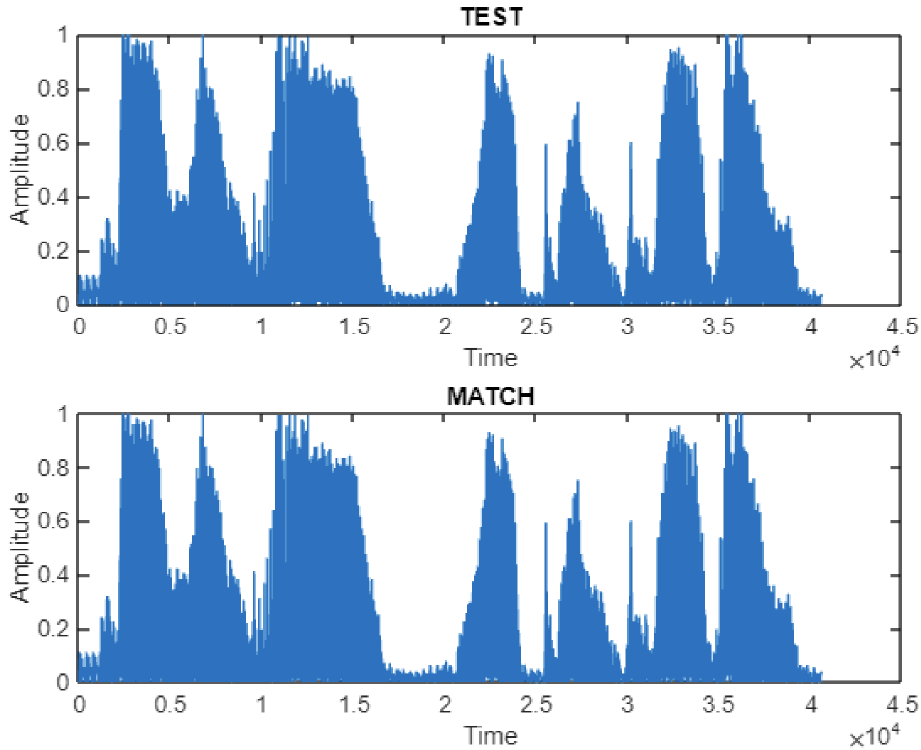


**Figure 3. Plot of file B3 matched with file A3**

## VIII. Discussion

As seen from the results from Figures 1 to 3, the program's matching is very accurate. Showing visual confirmation that both the test input recording and the matching recording are completely identical. Array creation has proven to be critical in the success of the program, specifically in the creation of the 'F' and 'FN' arrays wherein the position of the audio recording and its respective discrete cosine transform is stored. The discrete cosine transforms itself has proven to be an efficient medium for the necessary techniques in matching the audio recordings and has made the technique simple since the discrete cosine transform utilizes the real part of numbers as opposed to a discrete Fourier transform which utilizes complex numbers.

## Conclusion

The potential of the discrete cosine transform (DCT) as a medium for audio identification and matching has been proven effective by this study. Through the use of

DCT, audio-matching techniques through MATLAB were formulated and exhibited accurate results. Wherein, all tested audio recordings were successfully matched with the counterparts located in the database. With that said, the limitation of such a method is that the tested audio recordings are similar to that of the ones in the database. Further studies should be conducted when noise or distortion is introduced to the inputted signal to further understand the efficiency of utilizing discrete cosine transforms and pitch matching in audio recognition. Furthermore, Due to the limitations of this project, the program can further be improved such that this code can be integrated into a multitude of devices that have the capabilities to acquire audial information such as cellular phones, desktop computer recorders, etc. With the objectives of the program achieved, its potential can further be utilized in a variety of applications thereby offering valuable contributions to the field of audio identification.

## IX. Authors Contribution

DE VERA, Renato Angelo D.L.

- Abstract
- Introduction
- Programming/Coding
- Conclusion

LEONCIO, Abigail Joy T.

- Introduction
- Objective
- Theoretical Considerations
- Methodology

LUCMAYON, Ethan James T.

- Programming/Coding
- Methodology
- Results
- Discussion
- Conclusion

## X. References

[1] “Frequency Domain and Fourier Transforms,” Princeton.edu. [Online]. Available: https://www.princeton.edu/~cuff/ele201/kulkarni_text/frequency.pdf. [Accessed 11-April-2023].

[2] “Understanding Audio data, Fourier Transform, FFT and Spectrogram features for a Speech Recognition System,” Towards Data Science. [Online]. Available: https://towardsdatascience.com/understanding-audio-data-fourier-transform-fft-spectrogram-and-speech-recognition-a4072d228520. [Accessed 12-April-2023].

[1] D. Marshall, “The Discrete Cosine Transform (DCT),” Cf.ac.uk. [Online]. Available: https://users.cs.cf.ac.uk/Dave.Marshall/Multimedia/node231.html. [Accessed: 13-Mar-2023].

[2] V. Britanak, P. C. Yip, and K. R. Rao, “Fast DCT/DST Algorithms,” in Discrete Cosine and Sine Transforms, Elsevier, 2007, pp. 73–140.

[3] S. Tallat, “Voice identification and recognition system,” 2015.

[4] IMPLEMENTATION OF A VOICE-BASED BIOMETRIC SYSTEM, “PROJECT REPORT,” Cornell.edu. [Online]. Available: http://chenlab.ece.cornell.edu/people/adarsh/research/speech_project.pdf?fbclid=IwAR1loJWjuS5pfHklVtfRtyb2cmOqctexdCxfXX0A9Mn8chex12fIupUCYHs. [Accessed: 13-Mar-2023].

[5] S. Bunrit, The authors are with the School of Computer Engineering, SUT, 111 University Avenue, Muang, Nakhon Ratchasima 30000, Thailand, T. Inkian, N. Kerdprasop, and K. Kerdprasop, “Text-independent speaker identification using deep learning model of convolution neural network,” Int. J. Mach. Learn. Comput., vol. 9, no. 2, pp. 143–148, 2019.

[6] T. Siddiqui, “Voice Recognition System,” Jul. 2020.

[7] R. A, R. Agustina, and Hidayatulloh, “Comparison of discrete cosine transforms (DCT), discrete Fourier transforms (DFT), and discrete wavelet transforms (DWT) in digital image watermarking,” Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 2, 2017.