Application of FFT and KNN Methods for the Process of Identifying Sound Signals

Yessi Yunitasari [1*], Saifulloh [2], Daniswara Andhika Putra Harly [3]

[1,2,3]*Informatics Engineering Department, Engineering Faculty, PGRI Madiun University, Jl. Auri No. 14-16 Kota Madiun 6317, Indonesia*

Article Information
Received: 21-11-2024
Revised: 28-11-2024
Published: 5-12-2024

Keywords
Sound Signals, FFT, KNN, Machine Learning

****Correspondence Email:***
yessi@unipma.ac.id

Abstract
Everyone has a different kind of voice. Sound is a unique thing and has a certain range of frequencies and intensity of sound that can and cannot be heard by humans. we can detect important characteristics of the sound. The Fast Fourier Transform (FFT) algorithm is an algorithm for calculating Discrete Fourier Transform (DFT). Process of Identifying Sound Signals beginning with the sound data was preprocessed, feature extraction using FFT, classification using KNN and finding the nearest distance using the Euclidean distance method, an accuracy result of 79% of the tested data was obtained.

1. Introduction

Sound is a unique thing and has a certain range of frequencies and intensity of sound that can and cannot be heard by humans. The unit for measuring the intensity of the sound is the decibels (dB). The development of Natural Language Processing (NLP) for speech processing has seen rapid progress in recent decades, mainly thanks to the contributions of machine learning and deep learning techniques. Many benefits can be obtained from the process of voice detection and voice signal processing, among which voice detection can be used to analyze the emotions of a person's voice, which is relevant for applications such as automated customer service or psychological analysis, Automated voice technology can be integrated with IoT devices to create smart homes, smart vehicles, and other devices that respond to voice commands, Helping users with special needs, such as voice recognition for the visually impaired or voice control for people with limited mobility, In the manufacturing or health sectors, automatic voice detection can be used to analyze voice-based alerts or detect problems through vibration or machine noise.

Sound is a signal that is influenced by time. To change the signal from the time domain to the frequency domain, signal transformation is needed, one of which can be used is Fast Fourier Transform (FFT) method. The process for recognizing the speaker's voice can be define as voice recognition.Fast Fourier Transform (FFT) is one of the techniques that is often used in the analysis of sound signals, including for sound detection. FFT is an algorithm that accelerates the process of discrete Fourier transform (DFT), allowing us to break down sound signals into frequency components. By understanding the frequency of sound, we can detect important characteristics of the sound.The Fast Fourier Transform (FFT) algorithm is an algorithm for calculating Discrete Fourier Transform (DFT) which is used for two calculations of the signal frequency spectrum and FFT is an efficient DFT calculation procedure so that it will speed up the DFT calculation process which can substantially save more time than conventional methods. The fast fourier transfrom algorithm divides the frequency per period, so it can work well so that it produces accuracy quickly and efficiently(Riyani et al., 2019).

1.1 Literature Review

Everyone has a different kind of voice. Based on gender, voice type is divided into six parts, namely soprano, mezzo soprano, and alto for women; and tenor, baritone, and bass in men. Each type of sound has a different range and with different frequencies. This study classified the type of voice in women using the Fast Fourier Transform (FFT) method by recording the voices of each user which would then be processed using the FFT method to obtain the appropriate sound range. This research got results with an accuracy of up to 80%.The results obtained from this study are quite appropriate and it is proven that the FFT method can be used in digital signal processing (Apsari & Widiartha, 2021).

Voice-based systems permit users access to info on the net over a voice interface. Previous studies on examination systems that create use of voice interface don't sufficiently exhibit intelligent variety of assessment that diminishes the rigor of examination. the target of the system is to boost on the achievements of previous studies by providing a framework that can guide the development of a voice-based examination like government exams like TNPSC cluster II, cluster IV here we tend to develop the project for blind students United Nations agency all have an interest to participate the examination, so case knowledgeable system for the visually impaired students therein suggests that the queries are prepared by the language as they chose within the examination portal. So, the question are on scan mode, the scan mode can enhance the speech so queries are scan and therefore the choices also will admit defeat the scan mode in order that they will simply perceive the question and choices they'll answer it manner. The study employs a mixture of technologies like system style, server-side scripting, voice-based system development, knowledge management and rule-based reasoning in developing the system (Manikandan et al., 2022).

Biometric identification is one of the media in the operation of home automation systems that is in great demand. This is because conventional identification such as the use of PINs, passwords, cards and keys is not reliable enough, in terms of security and how it operates. The method used in this study is the Fast Fourier Transform (FFT) method which consists of system analysis, design, implementation and testing. This study uses 3 (three) words used for voice commands, namely "Open" and "Key" or "Close". The word "Open" is used to open the solenoid door lock, the word "Key" is used to lock the solenoid door lock, while the word "Close" is only an alternative when the word "Key" is not detected by the sound sensor due to various factors such as noise around the device, poor pronunciation. Where in the future it can use another version of the voice sensor to be applied to the voice recognition-based door lock system prototype (Tobing & Agung, 2020).

In the world of engineering, signals is a magnitude that changed by the time and space as well as bring some information. According to the ITU (International Telecommunication Union), the signal is a physical phenomenon in which one or more of its characteristics typify information. With a sound signal we can deliver an information easily. However, In this study will be conducted randomization of a sound signal that we want to convey this information only be accepted by the people we want. The process of voice recording is record by using the internal microphone of the PC and for signal processing I used MATLAB program . The method used is the Fast Fourier Transform to randomize and transform voice signals in the time domain into voice signals in the frequency domain. After recording the voice signals, the randomization step will be done to make the information that we send can not be heard by the others except our destiny. Based on the testing, this system can randomize all kinds of voices that recorded in intervals of 3 seconds . This study is only a simulated system through matlab (Yohanes et al., 2014).

2. Research Methods

The steps to carry out the process of identifying the sound system start from the stage of collecting sound set data, then the preprocessing process to standardize the data size, followed by the process of extracting sound features, the classification process using machine learning algorithms and the last is to obtain the results of voice signal classification. The machine learning algorithm used is the KNN (K-Nearest Neighbor) algorithm. The flow of research methods can be seen in Figure 1.

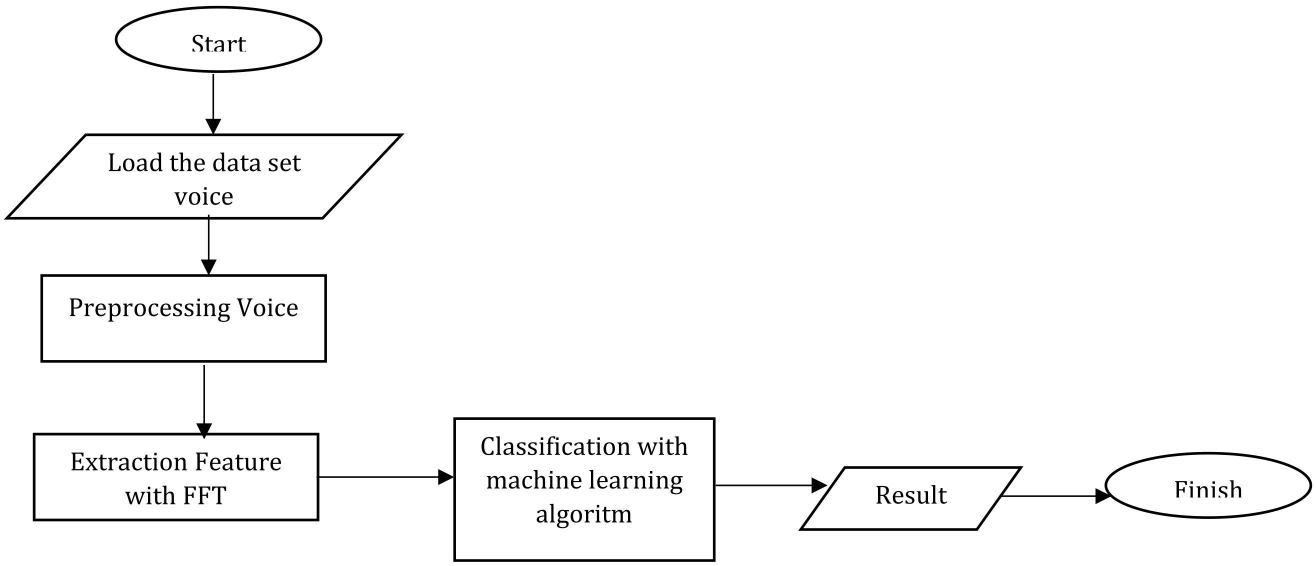

Fig 1 Flowhart Research Methods

2.1 Load the data set voice

At this stage, the vote data has been prepared. The sound set data used was 300 data with 3 labels. Each label consists of 100 data.

2.2 Preprocessing Voice

The preprocessing process aims to improve the sound recording so that it produces a good sound recording signal and the effects of noise on the recording can be reduced(Heriyanto, 2013). At this stage, segmentation of voices is carried out, voices are taken in the range of 2 seconds for each training data. To make this process easier, the sound has been edited first so that the sound immediately starts from the amplitude index to 0. The training data provided has the same bit rate of 16 bits and the same sample frequency of 44100 Hz. All training data is converted into digital discrete data that represents amplitude in the time domain. The voice data is cut to a duration of 2 seconds so that each data has the same time duration. The sound data that has been preprocessed can be seen in figure 2.

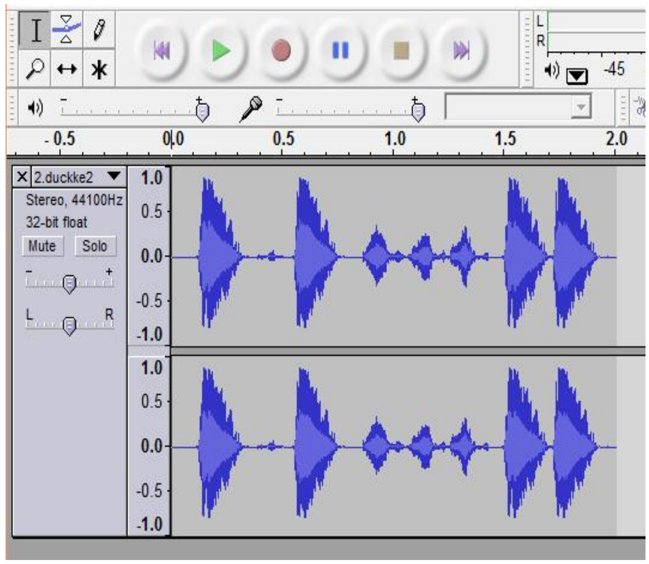

Fig 2 Sound Sample Preprocessing Results 1

2.3 Time-Frequency Domain Feature Extraction

At the time domain frame capture stage, the audio is converted into digital data and the amplitude is plotted on the time domain. The plot description for the training data sounds is as follows. An example of one of the sound plots can be seen in Figure 3.

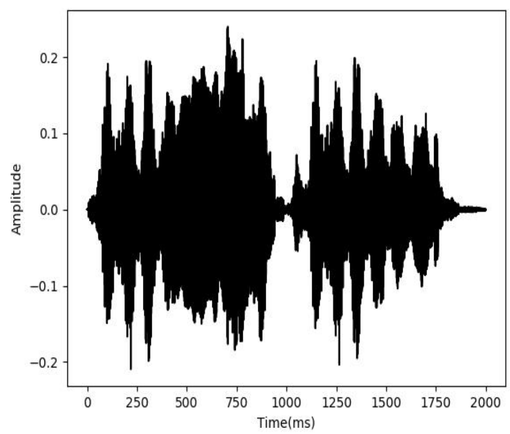

Fig 3 Time-Frequency Domain-based sound plot

2.4 FFT (Fast Fourier Transform) Process

FFT is an algorithm for calculating Discreate Fourier Transform (DFT) quickly and efficiently(Syaifuddin, 2014). FFT is a fast algorithm for Discrete Fourier Transform (DFT) implementations that are operated on discrete time-signals consisting of N samples(Sasilo et al., 2022). The FFT process is used to represent signals in discrete time domains and frequency domains. At this stage, the x-axis will represent frequency (kHz) and power in dB (Desibells). An example of one of the FFT results from the sound sample can be seen in Figure 4.

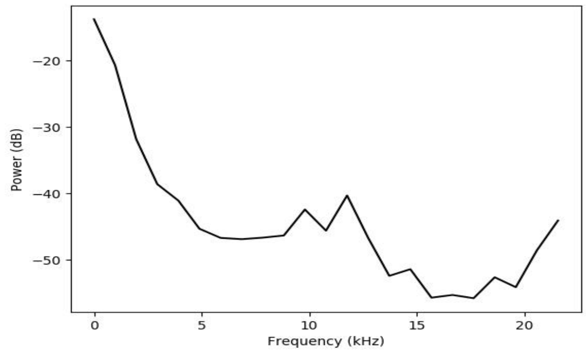

Fig 4 FFT sound sample results

2.5 K-NN Classification

The classification of K-Nearest Neighbor (KNN) is based on learning by comparing test data with training data following Equation (1). Train (n) attribute data, with each data describing the plus point of space with dimension n. When a data is unknown in the class, KNN will look for spatial patterns for the nearest training data, so that it can classify the data as needed(Anggoro et al., 2020). The classification stage uses the k-Nearest Neighbor algorithm with a value of k=3. At the classification stage, the testing data will be formed a frequency time matrix and then compared with all training data to check which data is closer to the data to be tested. K-Nearest Neighbor (KNN) is a method to classify objects based on learning data that are closest to the object. This method is widely used in the field of pattern recognition, the KNN classification is based on comparing a test data with a number of training data. The training data consists of n attributes, each data represents a point in an n-dimensional space, so all the training data is stored in the n-dimensional pattern space. When given a data of unknown class, the nearest neighbor will look for a spatial pattern for the nearest k training data (Wahyuni et al., 2022).

3. Result and Discussion

The principle of K-NN is to check whether the testing data is close to a number of the nearest neighbors. The number of classes that have more proximity to the data is considered to be the right class. Determination of proximity uses euclidean distance. The first step of this stage is to form a time-frequency matrix which is a 25x5 matrix into a 1-dimensional flat matrix measuring 1x125. The next step is to use the euclidean distance to determine the distance with each training data. The last stage is by sorting or sorting the distance data that has been calculated based on the shortest distance. Based on the sorted data, it can be seen the number of closest neighbors from the data to determine the class.

All data is audio data with the same specifications, namely with a bit rate of 16 bits and a sample frequency of 44100 Hz. Each data has been extracted from its features against the time frequency matrix. The results of the extraction of each data are as follows. Tests were carried out on 9 samples of voice data.

Table 1. Sample Voice data testing 1

Class Labels	Distance Values	3 Selected Distances
Voice A	134.937.620.179	
Voice A	94.890.362.477	v
Voice A	115.890.237.929	v
Voice B	820.924.339.216	
Voice B	91.583.846.454	v
Voice C	717.457.582.571	

Voice C	828.693.551.281	

Based on the table above, the sound tested for the first test is included in the A sound class because the 2 closest values include the A sound class and 1 value includes the B sound class.

Table 2. Sample Voice data testing 2

Class Labels	Distance Values	3 Selected Distances
Voice A	147.302.028.581	
Voice A	141.651.362.065	
Voice A	135.794.534.721	v
Voice B	171.530.006.046	
Voice B	185.817.328.139	
Voice C	128.181.163.864	v
Voice C	121.389.514.148	v

Based on the table above, the sound tested for the second test has a discrepancy because the sound tested is sound A but is classified in class C.

Table 3. Sample Voice data testing 3

Class Labels	Distance Values	3 Selected Distances
Voice A	119.133.824.754	v
Voice A	119.771.812.369	v
Voice A	131.927.470.215	v
Voice B	162.607.017.928	
Voice B	160.331.989.902	
Voice C	135.308.301.218	
Voice C	134.200.758.415	

Based on the table above, the sound tested for the second test is included in the A sound class because the three closest values belong to the A sound class.

Table 4. Sample Voice data testing 4

Class Labels	Distance Values	3 Selected Distances
Voice A	160.312.966.673	
Voice A	108.058.623.698	
Voice A	130.432.039.812	
Voice B	87.056.482.011	v
Voice B	72.770.592.845	v

Voice C	98.106.106.415	v
Voice C	102.157.758.525	

Based on the table above, the sound tested for the fourth test is correctly classified because the sound tested is a B sound classified in class B.

Table 5. Sample Voice data testing 5

Class Labels	Distance Values	3 Selected Distances
Voice A	171.131.162.539	
Voice A	109.944.476.515	
Voice A	1.328.726.325	v
Voice B	76.095.273.381	v
Voice B	85.723.797.714	v
Voice C	972.592.422.508	
Voice C	975.291.823.188	

Based on the table above, the sound tested for the fifth test is correctly classified because the sound tested is a B sound classified in class B.

Table 6. Sample Voice data testing 6

Class Labels	Distance Values	3 Selected Distances
Voice A	140.348.340.796	
Voice A	110.139.564.214	
Voice A	107.508.441.062	
Voice B	18.849.327.679	v
Voice B	11.238.002.456	v
Voice C	82.308.299.991	v
Voice C	623.508.324.137	

Based on the table above, the sound tested for the sixth test has a discrepancy because the sound tested is a C sound but is classified in the B sound class.

Table 7. Sample Voice data testing 7

Class Labels	Distance Values	3 Selected Distances
Voice A	13.657.533.156	v
Voice A	962.950.842.622	
Voice A	102.484.847.033	
Voice B	108.708.902.009	

Voice B	109.233.889.045	
Voice C	69.988.737.837	v
Voice C	82.798.252.072	v

Based on the table above, the sound tested is correctly classified because the sound tested is a C sound classified in class C.

Table 8. Sample Voice data testing 8

Class Labels	Distance Values	3 Selected Distances
Voice A	160.312.966.673	
Voice A	108.058.623.698	
Voice A	130.432.039.812	
Voice B	87.056.482.011	v
Voice B	72.770.592.845	v
Voice C	98.106.106.415	v
Voice C	102.157.758.525	

Based on the table above, the sound tested is correctly classified because the sound tested is a B sound classified in class B.

Table 9. Sample Voice data testing 9

Class Labels	Distance Values	3 Selected Distances
Voice A	150.232.799.928	
Voice A	122.992.088.724	v
Voice A	124.250.687.688	v
Voice B	873.455.939.002	
Voice B	101.561.639.154	v
Voice C	869.374.327.201	
Voice C	772.241.391.444	

Based on the table above, the sound tested is correctly classified because the sound tested is a A sound classified in class A. The data that was not successfully classified were 2 data outside the classroom so that it could not be known which class he was in.

4. Conclusions

After the sound data was preprocessed, feature extraction using FFT, classification using KNN and finding the nearest distance using the Euclidean distance method, an accuracy result of 79% of the tested data was obtained. This research can still be developed with different voice datasets, and the classification method uses deep learning to get even better accuracy results.

5. References

Anggoro, A., Herdjunanto, S., & Hidayat, R. (2020). MFCC dan KNN untuk Pengenalan Suara Artikulasi P. *Avitec*, *2*(1), 13–19. https://doi.org/10.28989/avitec.v2i1.605

Apsari, M. S. A., & Widiartha, I. M. (2021). Classification of Women's Voices Using Fast Fourier Transform (FFT) Method. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, *10*(1), 59. https://doi.org/10.24843/jlk.2021.v10.i01.p08

Heriyanto. (2013). *Instrumentasi dan Pengukuran*. *Mkb 7056*, 1–101. https://slideplayer.info/slide/13942689/

Manikandan, P., Kiruthika, M., Shrimathi, K., & Mubeena, A. (2022). Speech Recognition using Fast Fourier Transform Algorithm. *International Journal of Engineering Research & Technology*, *10*(08), 29–33. www.ijert.org

Riyani, A., Nurrochman, A., Sanjaya, E., Rizqiyah, P., & Junaidi, A. (2019). Mengidentifikasi Sinyal Suara Manusia Menggunakan Metode Fast Fourier Transform (Fft) Berbasis Matlab. *Journal of Informatics, Information System, Software Engineering and Applications Mengidentifikasi*, *1*(2), 42–50.

Sasilo, A. A., Saputra, R. A., & Ningrum, I. P. (2022). Sistem Pengenalan Suara Dengan Metode Mel Frequency Cepstral Coefficients Dan Gaussian Mixture Model. *Komputika : Jurnal Sistem Komputer*, *11*(2), 203–210. https://doi.org/10.34010/komputika.v11i2.6655

Syaifuddin, A. (2014). Fast Fourier Transform (Fft) Untuk Analisis Sinyal Suara Doppler Ultrasonik. *Youngster Physics Journal*, *3*(3), 181–188.

Tobing, F., & Agung, H. (2020). Penerapan Algoritma Fast Fourier Transform (Fft) Pada Prototype Door Lock System Berbasis Voice Recognition. *Jurnal Algoritma, Logika Dan Komputasi*, *3*(1), 248–254. https://doi.org/10.30813/j-alu.v3i1.2090

Wahyuni, N. M. P., Putri, L. A. A. R., Putra, I. G. N. A. C., Darmawan, D. M. B. A., Raharja, M. A., & Muliantara, A. (2022). Implementasi Metode K-Nearest Neighbor Dalam Mengklasifikasikan Jenis Suara Berdasarkan Jangkauan Vokal. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, *11*(1), 187. https://doi.org/10.24843/jlk.2022.v11.i01.p20

Yohanes, S. R., Lumenta, A. S., & RUA Sompie, S. (2014). Simulasi Sistem Pengacak Sinyal Dengan Metode FFT (Fast Fourier Transform). *E-Journal Teknik Elektro Dan Komputer (2014), ISSN 2301-8402*, 1–9.