eISSN: 2477-7986

Journal of Measurements, Electronics, Communications, and Systems *(Vol.06, Pg.28-33, 30 June 2020)*

Design and Implementation of Learning Tools to Read the Braille Letters Based on Speech Processing Using Arduino

R. Patmasari, Sofia Saidah , A.F.Akbar, R. Magdalena
School of Electrical Engineering, Telkom University, Bandung, 40287, Indonesia
raditiana@telkomuniversity.ac.id
Manuscript received Marc 01, 2020; revised June 10, 2020, Accepted June 25, 2020

Abstract

Ability to read Braille is critical skill for blind students. Without the skill, blind students would encounter difficulties in their learning activities because most learning materials are written using the Braille system. The currently applied Braille learning system uses printed paper that is time consuming and pricey. This research attempts to develop a tool for helping the blinds to learn how to read braille letters. The tool processes inputs in the form of speech signal into a text by applying Mel Frequency Cepstral Coefficient (MFCC) as a feature extraction method and K- Nearest Neighbor (KNN) as a classifier method. The text will subsequently be transformed into Braille pattern by using Arduino UNO. The test results discover the combination of Mel Frequency Cepstral Coefficient and K-Nearest Neighbor method are able to recognize the speech signal of different alphabets with 87,3% accuracy. Furthermore, the computing time for alphabet recognitions decreases 85 % when the device is applied This finding helps the blind students to recognize the alphabets easily and faster.

***Keywords*:** Arduino; Blinds; Braille; KNN; MFCC; Speech Processing

doi: 10.25124/jmecs.v6i1.2019

1. Introduction

Reading activity builds a strong foundation to learn and understand any knowledge. It is very important for everyone, especially for blinds. Without basic skill of read braille letter, blinds will have problem to learn, because almost all learning activity are conducted through read and write activities [1].

Learning braille letters with conventional method (using printed paper) is commonly used in special school called Sekolah Luar Biasa (SLB). This method is pricey and time consuming [2]. As an alternative, a previous research entitled "Kannada Speech Recognition Using MFCC and KNN Classifier for Banking Applications" had been conducted by applying Mel Frequency Cepstral Coefficient and K-Nearest Neighbor method with accuracy 91,5% [3] Besides, another research by students joining the MOCO Warrior Team from the Sepuluh Nopember Institute of Technology (*Institut Teknologi Sepuluh Nopember* – ITS), Indonesia has proposed a tool called Edu Braille for supporting the read activities of the blind based on Arduino [4]. The first research, however, focused on the processing of speech-to-text for banking applications implementations, while the latter research covered a limited set of input data from only a prepared database and no additional speech-related feature. Therefore, our research aims to not only develop a similar tool to edu-Braille but also utilizes speech signals as its input data to generate the text data. This research built a tool for helping blind students in SLB to learn how to read Braille letters easily without the need to print any learning materials. Hence, it is expected to reduce cost, saving time, and promote an autonomous learning for the students. In particular, this research focuses on the design and analysis of a Braille learning tool based on speech processing.

2. Design and Implementation

In general, human use the same physical anatomy to speak. However, human abilities to control speech producing system of organs, including the mouth, nasal cavity and human respiratory, differ between individuals, hence resulting in different characteristics of human voices [5].

Theoretically, Braille has been recognized as a tactile writing system for the blinds created by Louise Braille (1809 - 1852). Braille system applies a 6-dot scheme so-called the Braille pattern. The pattern is divided into 2 columns, in which dots in the left column is numbered 1, 2 and 3 from top to down, while dots in the right column is numbered 4, 5, and 6 in the same direction. Figure 1 shows the numbering scheme of Braille pattern [6].

Fig. 1. Braille Cell

The design for identifying speech signal by applying MFCCs and k-NN method is displayed in Figure 2 [7].

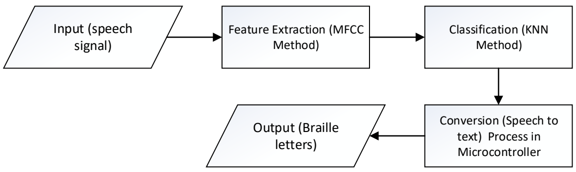


Fig. 2. Block Diagram of Speech Signal Identification

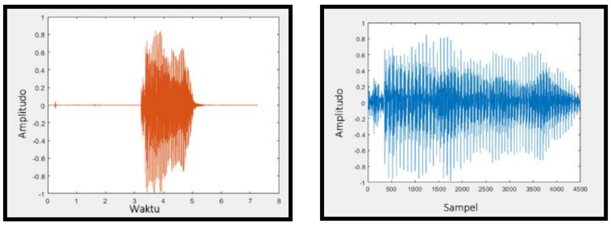
There are two stages namely training and testing. Both of training and testing phase start with the preprocessing and feature extraction stages (using the Mel Frequency Cepstral Coefficient). However, the different between training and testing phase, the output data of the feature extraction stage in training phase will be stored into the database, while in the testing stage, a speech recognition process is carried out to determine a reference model that is similar to the speech input signal using the K- Nearest Neighbors method.

In this research, 1300 training data and 260 test data were collected. The training data speech collection consists of 26 letters with each letter being repeated 50 times in a soundproof condition. The testing speech data colletion consists of 26 letters with each letter being repeated 10 times in a soundproof condition. The data were taken using the Now Smart Sound Recoder application and the data saved in * m4a format.

2.1 Pre- Processing

The first step in this research is pre-processing. Steps in the pre-processing stage include convert stereo to mono, normalization and signal cropping.

Every input signal that has been read will be carried out stereo (two channels) to mono (one channel). This is done because the extraction process will be much faster using a signal with one channel. Normalization is a process to modify the signal to ensure comparable volume levels of all incoming signals [8]. The output of this process is speech signal that has an amplitude range from -1 to +1. Finally, signal cropping is the last process in pre-processing to remove the signals that are not required during feature extraction process [9].



3(a) 3(b)

Fig. 3(a). Normalization signal, 3(b). Normalization Signal after Cropping

Cropping is done manually by removing the start and end of signals that do not contain spoken letter information.

2.2 Feature Extraction using Mel Frequency Cepstral Coefficient Method

When introducing a sound signal, an extraction feature is required to identify the signal. The feature extraction is conducted by transforming voices into standardized data in the form of wave spectrum [10]. Steps in MFCC method are explained as follows.

2.2.1 Frame Blocking

In this frame blocking step, incoming signal is split into several segments called frames. Every frame consists of N sound samples, in which the distance between two frames is separated by M samples. The length of a frame divides every sample into a few frames based on time period [10].

2.2.2 Windowing

The purpose of a windowing is to give weight for each frame. An example of windowing is the use of Hamming window. The formula of Hamming is expressed as follows [10]

$$w(n) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) & 0 < n < N-1 \\ 0 & \text{others} \end{cases} \quad (1)$$

where n is indicates the sequence of signals while N is filter order.

2.2.3 Fast Fourier Transform

In FFT, the signal from N frame of time domain will be transformed into the signal of frequency domain. The FFT-transformed speech signal is useful for convolution vibration process of incoming sound and responding to sound waves within time domain [10].

2.2.4 Mel Wrapping

Mel scale is applied for adjusting the resolution of each frequency into a human-audible property. Mel scale is arranged into several critical banks by using a filter bank. Range of the frequency in a spectrum is very large and does not correspond to a linear scale. After the spectrum is prosessed, the mapped in Mel scale using a triangular scaling [11]. The formula of Mel wrapping for transforming frequency into Mel scale is expressed in Eq. (2) [11].

$$\mathrm{Mel}\ (f) = 2595 \times \mathrm{Log}_{10}(1 + f/100) \qquad (2)$$

where f is frequency used in a spectrum.

2.2.5 Cepstrum

Mel spectrum coefficient takes the form of a real number, hence it is convertible into time domain using DCT [11]. Performing the DCT with a decorelated Mel spectrum will produce a good representation of local spectral property [12]. The formula of DCT is expressed in Eq. (3) [12].

$$DCT(i) = \frac{2}{N}\, C(i) \sum_{x=0}^{N-1} pixel(x) cos\left[\frac{(2x+1)i\pi}{2N}\right] \qquad (3)$$

where *DCT(i)* is the dct index value to i; N is the matrix size; *pixel(x)* is pixel value at x^{th} index; *C(i)* =1 if x>0 and *C(i)* = $\frac{1}{\sqrt{2}}$ if x=0

2.3 Classification using KNN Method

Data classification using KNN method has been recognized based on a set of training data closest to the set of tested data [13]. Basically, KNN is applied to find the nearest distance between training data and testing data. The formula of Euclidean distance is expressed in Eq. (4) [13].

$$d_i = \sqrt{\sum_{i=1}^{p} (x_{2i} - x_{1i})^2} \qquad (4)$$

Meanwhile, the formula of city block distance to find the nearest distance is written in Eq. (5) [14].

$$d_{ij} = \sum_{k=1}^{n} |x_{ik} - x_{jk}| \qquad (5)$$

Finally, formula of the Cosine distance to find the nearest distance is presented in Eq. (6) [15].

$$\cos a = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \qquad (6)$$

2.4 Device Design

Figure 4 presents the system that is connected to hardware . The hardware used in this system are Arduino Uno which is a microcontroller board based on Atmega328, 5V power supply , 6 servo as a drive, and plastic tube (embosser) to form a braille letter pattern.

The feature extraction process using MFCC method and classification using KNN method are carried out on a personal computer. The output of this process then becomes the input for Arduino, then Arduino drives the servo to move plastic tube to form braille letter pattern.

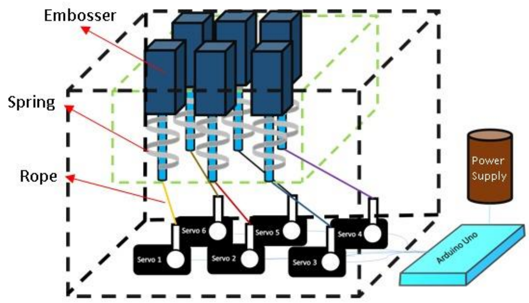


Fig. 4. Device Design

2.5 System Performance

The accuracy and average time of the system are used to show the index of performance. The formulas are expressed in Eq. (7) and (8) [15].

$$\mathrm{Accuracy} = \frac{\text{number of true data}}{\text{number of testing data}} \times 100\% \qquad (7)$$

$$\text{time average} = \frac{\sum_{1}^{n} Time}{\text{the number of testing data}} \qquad (8)$$

3. TESTING AND ANALYSIS

The following tests and analysis are conducted on the designed tool to measure the performance and to analyze the effect of the number of MFCC coefficients, the K parameter, and the type of distance in the KNN classification

3.1 Analysis of the effects of a MFCC coefficient

In this test, several different numbers of MFCC coefficients are used to see the effect of the number of MFCC coefficients on the system accuracy value. The number of coefficients used is 10, 20, 30, 40, 43, and 45.

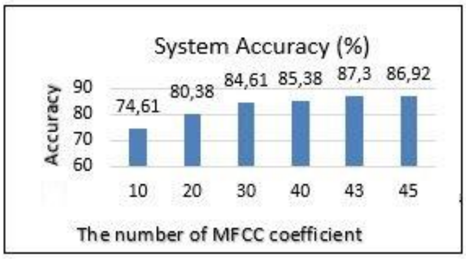


Fig. 5. Results of MFCC Coefficient Testing

Based on Figure 5 the best accuracy is at the number of MFCC coefficient is 43, with the accuracy is 87,3% and the worst accuracy is at the number of MFCC is 10, with the accuracy is 74,61%.
Based on Figure 4 the best accuracy is obtained when using 43 MFCC coefficient. With a few MFCC coefficient can't representing data sounds for identify, but too much the coefficient MFCC make the characteristic unclear and can't recognize.

3.2 Analysis of the effects of Parameter K

In this test, several different K parameters are used to see the effect of the K parameter on the system accuracy value. The number of K parameters used is 1, 3, 5, 7, and 9.

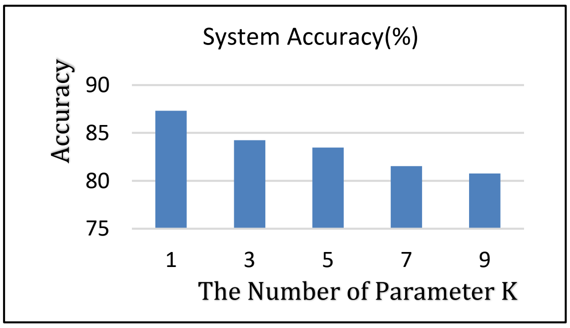


Fig. 6. Results of the number of Parameter K testing

Based on Figure 6, the best accuracy 87,3% is obtained when using K=1 parameter. The bigger K value will increase the neighbors to compare with. So it makes the system probably recognize a wrong speech signal.

3.3 Analysis type of K value in various type of distances

In this test, several different types of distance are used to see the effect of distance on the system's accuracy value. The types of distance used are euclidean distance, cityblock distance and cosine distance.

The number of K and type of distance calculation have an effect on the accuracy system, because in principle the KNN works based on the closest distance between the test data and the training data for the K's closest neighbours.

Table 1. The Effect of Distance Type

System Accuracy (%)			
The Number of K	Euclidean Distance	Cityblock Distance	Cosine Distance
1	87,3	85,76	87,3
3	84,23	84,61	82,6
5	83,46	83,84	82,3
7	81,53	81,92	81,5
9	80,76	82,69	80

From table 1 the best accuracy 87,3% is obtained when using Euclidean distance and cosine distance, and the number of K = 1.

3.4 System Accuracy using device

The system that has been designed is connected to the device and then tested for its accuracy and the average time to produce the output.

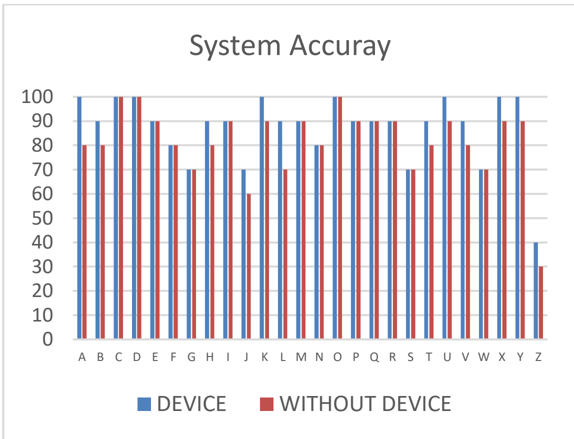


Fig. 7. Result of system accuracy using device

From Figure 7, We can see that the accuracy of system with using device have accuracy mean 81.54%. The accuracy with device is more less than accuracy mean system without using device is 87.3%. The average accuracy using device is smaller than the average accuracy of the system without using the device, this is because the servo used cannot always move according to the command.

Table 2. Computing Time

Letter	Time (Second)	
	Using Device	Without Device
A	4,015616	14,847847
B	1,719449	15,083895
C	3,526800	13,217571
D	1,578302	12,742087
E	3,440878	20,981758
F	1,425718	14,569611
G	1,717372	14,854592
H	3,657082	12,803158
I	1,501457	13,032386
J	1,432018	12,437143
K	1,363207	12,694556
L	1,916947	12,960620
M	1,349017	12,639308
N	1,257130	12,500697
O	1,375051	12,217072
P	1,431131	12,542711
Q	1,460380	12,493731
R	1,245931	14.680354
S	1,269456	12,719555
T	1,391584	13,128457
U	1,257626	12,468708
V	1,636006	12,807492
W	3,668810	14,759756
X	1,598503	12,792915
Y	1,316286	14,467909
Z	1,265674	14,212520
Average Time	1,902070	13,71632

From the table 2 we can see that average computation times are 1.902070 s and 13.71632 s with and without device application, respectively.

4. Conclusion

We can conclude that the application of MFCC and K-NN method to speech recognized system is able to identify a speech with a best accuracy of 87,3% when 43 MFCC coefficient , K value equal to 1, and using Euclidean distance are used. Besides, the average computation times are decreased from 13.71632 s to 1.902070 s when device is applied. This study helps the blind students to recognize speech faster and easily.

References

[1] D. A. L. Fitria, "Permulaan Braille bagi Siswa Tunanetra Diajukan kepada Universitas Negeri Surabaya untuk memenuhi Persyaratan Penyelesaian Program Sarjana Pendidikan Luar Biasa Oleh : Universitas Negeri Surabaya," 2017.

[2] J. T. Komputer and U. K. Indonesia, "Braille code trainer," Seminar Nasional Aplikasi Teknologi Informasi. 2009.

[3] S. B Harisha, S. Amarappa, "Kannada Speech Recognition Using MFCC and KNN Classifier for Banking Applications" vol.5, issue 1, DOI: 10.15680/IJIRCCE.2017.

[4] Dadang ITS (s), "Edu Braile Buatan ITS, Mudahkan Tunanetra Belajar", ITS, 21 Oktober 2016, [Online]. Tersedia: https://www.its.ac.id/news/2016/10/21/edu-braile-buatan-its-mudahkan-tunanetra-belajar/ [Diakses: 24 Oktober 2018].

[5] S. B. Bhaskoro, I. Ariani, and A. Anandha, "Transformasi Pitch Suara Manusia Menggunakan Metode PSOLA," ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika,vol. 2, no. 2, pp. 129–151, 2014.

[6] E.D.Susanti, B. Hidayat, and N. Andini, "Steganografi Teks pada Citra Digital menggunakan Metode Braille dan Metode Pemilihan Piksel LSB," e-Proceeding of Engineering, vol. 2, no. 2, pp. 2951–2958, 2015.

[7] S. K. Gaikwad, B. W. Gawali, and P. Yannawar, "A Review on Speech Recognition Technique," Int. J. Comput. Appl., vol. 10, no. 3, pp. 16–24, 2010.

[8] B. M. Nema, and A. A. A. Kareem, "Preprocessing Signal for Speech Emotion Recognition," Al-Mustansiriyah Journal of Science, vol 28, Issue 3, 2017.

[9] A. Suroso, Y. Fitri, S. F. Retnowaty, and Nurkhamdi, "Aplikasi Pengenalan Ucapan dengan Ekstraksi Ciri Mel Frequency Cepstral Coefficient (MFCC) dan Jaringan Syaraf Tiruan (JST) Propagasi Balik untuk Buka dan Tutup Pintu," Jurnal Komputer Terapan Vol. 1, No. 2, November 2015.

[10] T. Chamidy, "Metode Mel Frequency Cepstral Coeffisients (MFCC) Pada

klasifikasi Hidden Markov Model (HMM) Untuk Kata Arabic pada Penutur Indonesia," Matics, vol. 8, no. 1, pp. 36–39, 2016.
[11] S. Dhingra, G. Nijhawan, and P. Pandit, "Isolated speech recognition using MFCC and DTW,"International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, vol. 2, no. 8, pp. 4085–4092, 2013.
[12] Y. Miftahuddin, M. B. Murini, and M. R. Hakim, "Implementasi Mel Frequency Cepstral Coefficient dan Dynamic Time Wrapping untuk Pengenalan Nada Pada Alat Musik Bellyra," Seminar Nasional Teknologi Informasi 2017.
[13] R. M. Ikhsanudin, "Identifikasi Citra Pada Plat Nomor Kendaraan Mobil Pribadi Menggunakan Metode K-Nearest Neighbour," Skripsi. Fakultas Ilmu Komputer. Universitas Dian Nuswantoro. 2014.
[14] D. P. Pamungkas, and F.R. Hariri, "Implementasi Metode PCA dan City Block Distance untuk Presensi Mahasiswa Berbasis Wajah," Seminar Nasional Teknologi, Informasi, Komunikasi dan Aplikasinya, vol. 4, pp. 190-194. des 2017.
[15] R. T. Wahyuni, D. Prastiyanto, and E. Supraptono, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi," Skripsi. Fakultas Teknik Elektro. Universitas Negeri Semarang. 2018.

Author Information

Raditiana Patmasari is currently a Lecturer in School of Electrical Engineering Telkom University, Bandung, Indonesia. She received the bachelor and master degree from Telecommunication Engineering, Telkom Institute of Technology, Bandung. Her research interest is information signal processing, especially in biomedical engineering, embedded system/ hardware system development, and information/ coding theory.

Sofia Saidah received the B.S. and M.S. degree from Telecommunication Engineering, Telkom Institute of Technology, Bandung, Indonesia in 2012 and 2014 respectively. She is currently a Lecturer in School of Electrical Engineering Telkom University. Her research interest includes, image processing, audio processing, biomedical engineering, steganography and watermarking.

Azri Fathul Akbar received a bachelor's degree in telecommunication engineering from Telkom University. He is passionate in signal processing and embedded system/ hardware system development.

Rita Magdalena is currently is currently a Lecturer in School of Electrical Engineering, Telkom University. Her research interest is information signal processing especially in image processing, audio processing and biomedical engineering