

# Recognition of numerical password by voice for opening of electronic lock

## Abstract

There are situations where the opening of doors, drawers and other compartments becomes unviable for reasons of hygiene, contamination or simply by not being available to use hands, as in the case of people with immobilized or disabled arms which prevent this type of movement. Systems that operate by voice commands can be very useful in these cases, and with the current technology available it is possible to develop a system with relatively low cost and that can be accommodated in small spaces. The present work presents the development of an embedded system capable of capturing a numeric password spoken by the user, analyze this password and compare it with a predefined password to decide whether or not to open a specially designed lock. The software of the embedded system is able to capture the ambient sound through a microphone, to process it in real time and, with the aid of an Artificial Neural Network (ANN) of the Multilayer Perceptron type, interpret the numerical sequence spoken by the user in order to verify if this sequence matches the previously programmed password. In this version, the system is able to interpret ten digits in Portuguese language ("zero" (zero), "um" (one), "dois" (two), "três" (three), "quatro" (four), "cinco" (five), "seis" (six), "sete" (seven), "oito" (eight), and "nove" (nine)), but can be expanded to interpret variations of the same, such as "meia" (half a dozen), or higher numbers such as "dez" (ten), "onze" (eleven), and so on. In order to train the ANN, a database was created containing the locutions of the selected numbers. For the construction of this base, the utterances of the numbers were recorded by a group of 50 volunteers, including men, women and children

**Keywords:** digital signal processing, artificial neural networks, speech recognition

**Abbreviations:** ANN, artificial neural network; MFCC, Mel frequency Cepstral coefficients

## Introduction

Accessibility is often restricted, or at least difficult, for those who cannot use their hands, whether due to hygiene, contamination or even natural causes, such as people with physical disabilities. This can be for room openings, safes, doors, drawers, among other compartments. Voice command systems can be very useful in these cases, and with current technology available it is possible to develop a low cost system that can be confined in small spaces. Intelligent voice based access control systems are developed since decades,<sup>1</sup> and they are widely studied for commercial purposes and as a tool for teaching and learning in electrical engineering schools. Wahyudi<sup>2</sup> used an intelligent system approach to develop authorized person models based on their voice using Adaptive-Network-based Fuzzy Inference Systems to open doors connected to a building security.

Monday et al.,<sup>3</sup> propose a design of an improved cost effective electronic locking system which restricts access to certain important data and valuables to certain individuals using low cost microcontrollers and a simple keypad solution. Other authors also showed different solutions for the same purpose,<sup>4-8</sup> this shows that the subject can be widely explored and with different proposals. The proposed work, developed by electrical engineering students, aims to develop a system capable of capturing a numeric password spoken by the user, analyzing this password and comparing it with the predefined password for opening a lock. This system features a microphone, a processor that runs an algorithm that analyzes audio signals, and

an electronically actuated lock. The digit password (from 0 to 9 in Portuguese) is spoken by the user and a microphone captures the audio. This audio is analyzed by an algorithm that compares it with the respective digits for the correct password. If the password spoken by the user is true, the lock opens.

## Methods

For the choice of the system HW board, it was established that it should fulfill some requirements, such as: low cost; good computational performance; readiness to run a Linux environment; compact size; own a microphone. After an extensive research involving the boards available in the Brazilian market, it was decided to choose the Orange Pi Lite board. As Operating System for Orange Pi Lite, Armbian V5.35 was chosen. All scripts used in this project were written using the Python Programming Language. The methodology used for project implementation is explained in the subsections below.

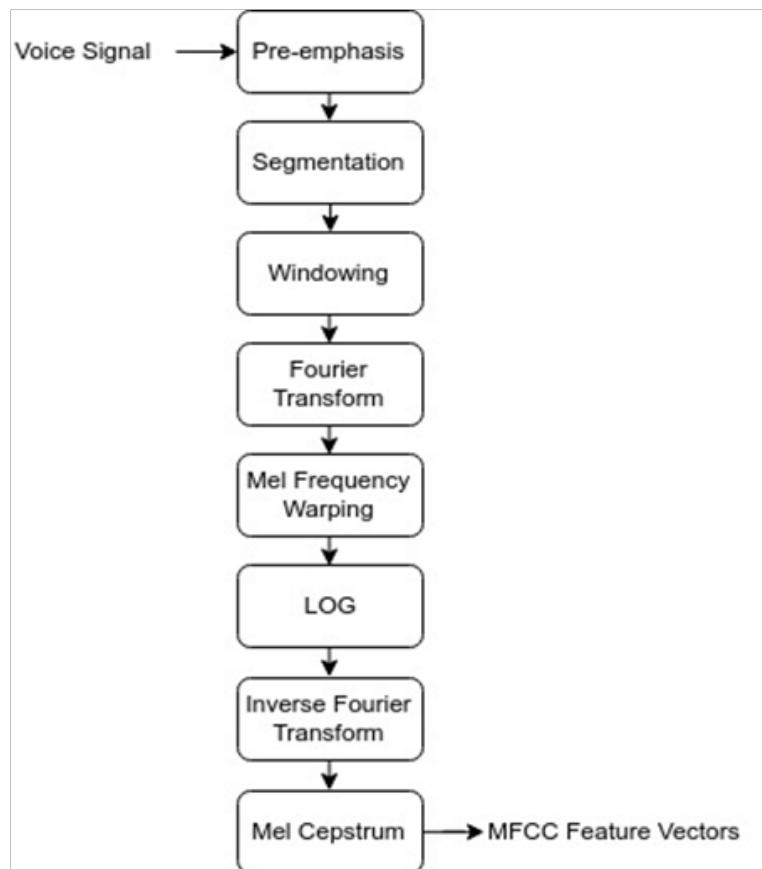
## Database construction

The database was recorded using Orange Pi Lite itself. A script was used to record ".wav" audio files with duration of 2 seconds and a sampling frequency of 16 kHz. The samples were recorded with the board exposed and the recordings took place in various environments and with the presence of environmental noises such as people talking around the board, TV sets and radios on, etc. The participants were 50 volunteers, distributed between 0 and 89 years old, men and women. We sought to create a diverse database to make the classifier robust enough to recognize different types of voices. The volunteers were asked to pronounce each digit three times from zero to nine. At the end of the collection, the database had 150 samples of each digit.

## Feature extraction

For the extraction of characteristics of audio signals, the algorithm MFCC (Mel Frequency Cepstral Coefficients) was used, because it presents a high capacity to discriminate the characteristics of voice signals.<sup>9</sup> For the extraction of MFCCs an open source API called “python-speech-features” was used.<sup>10</sup> Figure 1 shows the steps to which an audio signal is subjected to acquire the MFCC feature vector.<sup>11</sup> Before starting the feature extraction process, the audio sample goes through the normalization step to ensure that all samples that make up the file have values between 1 and -1. Next, the normalized file goes through a pre emphasis filter to amplify the higher frequencies that

have been attenuated due to the physical characteristics of the human speech system. Since the audio acquisition system records 2-second samples, an audio segmentation process based on signal energy variation that identifies the attack and the end of the pronounced word sustain has been developed. This procedure is necessary because only the part of the audio containing the word is sent for MFCCs extraction. Each segmented audio section is divided into 10 windows of varying length, and from each window the 13 most relevant coefficients for the speech recognition process are extracted. Therefore, each digit can be described as a feature vector of 130 elements. The use of windows of variable size is justified by the fact that they adapt to the length of each command, which varies from digit to digit.



**Figure 1** Steps for MFCC feature vector acquisition.

## ANN training and validation

For the Artificial Neural Network<sup>12</sup> (ANN) training process, an open source API called “Pybrain”<sup>13</sup> was used. Initially, the 50 volunteers were separated into 44 for ANN training and 6 [3 women and 3 men] for trained ANN validation. A thorough analysis of the 132 samples of each digit from the 44 volunteers was performed and the corrupted samples were discarded with 88 samples remaining for each of the ten digits. Tests were performed to find the best setting for ANN. The tests consisted of varying the number of hidden layer neurons, the hidden layer activation function, the output layer activation function, the learning rate (which has a direct impact on ANN learning speed) and also the momentum rate (a constant that determines the effect of past weight changes on the current direction of motion in the weight space). The parameters with which ANN presented the best

performance were: 130 neurons in the input layer; 36 neurons in the hidden layer; 10 neurons in the output layer; hyperbolic tangent as the activation function of the hidden layer; soft max as activation function of the output layer, 0.01 of learning rate; and 0.06 of moment. After trained ANN, it was validated using 18 samples of each digit from the 6 volunteers who were not used in training and who were recorded in a low noise environment. The average ANN hit was 88.3%, having hit 159 commands in 180 attempts. The details of the test can be seen in Figure 2 which presents a confusion matrix that shows, in the first column, the digits pronounced by the volunteers. Lines 2 through 11 show how each of the 10 digits was sorted by ANN. For example, in line 11, it can be seen that out of 18 times the number “9” was spoken, 14 times the ANN classified it correctly but on 4 occasions ANN classified it as the “4” digit.

Classified as =>	0	1	2	3	4	5	6	7	8	9
0	18	0	0	0	0	0	0	0	0	0
1	2	13	1	0	0	0	0	0	2	0
2	0	0	15	1	2	0	0	0	0	0
3	0	0	1	16	0	1	0	0	0	0
4	0	0	0	0	18	0	0	0	0	0
5	0	0	0	0	0	18	0	0	0	0
6	0	0	0	1	0	1	16	0	0	0
7	1	0	0	3	0	0	1	13	0	0
8	0	0	0	0	0	0	0	0	18	0
9	0	0	0	0	4	0	0	0	0	14

**Figure 2** ANN hit ratio in the validation base.

### Tests during development

After discovering the best parameters for training the Neural Network, we sought to validate the effectiveness of an ANN trained under these conditions when subjected to audio samples from different environments. In this stage, 12 volunteers participated, being women and men of different ages and who did not have their voices used for ANN training. Volunteers were asked to speak three times

each digit, from zero to nine. The 36 samples of each digit obtained were submitted directly to ANN without any treatment or removal of the corrupted samples. ANN's average hit was 78%, having hit 281 commands in 360 attempts. Test details can be seen in Figure 3. With these preliminary results a hit rate was obtained that justified the maintenance of the current ANN parameters and the continuity of the project.

Classified as =>	0	1	2	3	4	5	6	7	8	9
0	32	0	0	0	0	0	1	3	0	0
1	0	26	1	0	1	1	0	2	5	0
2	2	1	28	0	0	0	1	1	1	2
3	0	2	2	27	0	1	4	0	0	0
4	0	0	0	0	31	0	0	0	3	2
5	2	4	2	0	0	25	2	1	0	0
6	1	0	2	2	0	1	29	1	0	0
7	2	1	0	0	0	3	1	28	0	1
8	1	3	0	1	0	3	0	3	24	1
9	2	0	0	0	3	0	0	0	0	31

**Figure 3** ANN hit ratio in variable noise environments.

### Final prototype

To perform the activation of the device that opens the door, an electronic board was developed. The board has as main components a 12V powered relay, an opt coupler and three RGB LEDs to facilitate user interaction. The Orange Pi Lite + electronics board set was then housed inside the box that was built to simulate a door. This box is illustrated in Figure 4. The prototype has, as a form of interaction with the user, 3 RGB LEDs that can show the colors red, green, blue and white. When the system is idle, the 3 LEDs turn blue. When the user pronounces a command to wake the system, which can be any word, all the LEDs go out and immediately only LED 1 lights up in white indicating the moment for the user to speak the first digit of the password. Two seconds after the first LED lights up, the second LED also lights white showing the user the moment to speak the second digit of the password. And finally, 2 seconds after LED 2 lights up, LED 3 also lights white indicating the time for the user to speak the third digit of the password. Immediately after the three digits of the password are pronounced, the system processes the information received and, if the

password is wrong, illuminates the three red LEDs. If not, all three LEDs light up green and the door opening mechanism is activated.

**Figure 4** Ready prototype.

## Results and discussion

### System accuracy

The system as a whole performed satisfactorily, with an average hit rate of 88.3% in low noise environments, with the digits “um” (one) and “sete” (seven) having the worst hit rate, 72.2% each. However, it was observed that when pronounced by female voices, the digits “zero” and “sete” (seven) were sometimes confused, respectively, with “sete” (seven) and “zero”. Another point observed concerns the number “três” (three). When the user pronounces it as “três” (three) the system in some cases classifies it as “seis” (six).

### Processing time

The average time between when the third digit of the password is pronounced until the system signals the door to open or not is 3.8 seconds. This delay is due to the script block responsible for segmenting the 2-second sample and sending only the part containing the pronounced word for feature extraction. Some strategies to reduce system processing time were tested. One was the use of the fixed window segmentation technique where the script identifies the beginning of the word and segments the audio after a fixed time thus making all samples the same length. However, these strategies eventually interfered with the ANN hit rate which decreased to 82% in controlled environments.

### Prototype

Due to the fact that all samples obtained for the creation of the database were recorded with the exposed Orange Pi Lite board, the question arose as to how the set would behave when operating within a confined environment. Tests were performed on random volunteers and it was found that the hit rate remained similar to the hit rate obtained with the exposed board, however, the hit began to decrease as the volunteers pronounced the digits more than a meter apart. From a hardware standpoint, the device proved reliable. During part of the project development, the Orange Pi Lite board remained on for two uninterrupted weeks and it did not lock or shut down due to overheating. It is noteworthy that we chose not to use heat sink or fan because of the limited space of the box and also the noise in the case of the fan. For the built-in electronic board test, a “stress test” was developed, which consisted of activating and deactivating all the board outputs to study their behavior against overheating. The set did not malfunction even when the test was performed with the entire system in the box.

### Conclusion

Opening doors, drawers or other compartments is often restricted, or at least difficult, for those who cannot use their hands, whether due to hygiene, contamination or even physical limitations, such as for persons with physical disabilities. The present project proposed to be an alternative to solve this type of problem by presenting a compact and low cost device capable of recognizing a numeric voice password for the activation of electronic locks. The goal has been achieved; the device efficiently performs the purpose for which it was designed. The development was also a test of concepts for undergraduates who performed the project. The students developed a sense of research and application of concepts in the development, implementation and testing of software and hardware. However, on the road to excellence,

there is always room for change and improvement. In the case of this work, for future improvements, the following may be cited as suggestions: investigate and test other audio signal feature extractors, investigate and test other audio segmentation methods, investigate and test other pattern classification methods, test the use of directional and external microphones on the board, and moreover, more voice samples could be added to the database used to train the classifier, either by recording the voices of new volunteers or through Data Augmentation techniques. Such changes could result in increased command hit rate, increased rating speed, and increased set robustness when operating in high noise environments.

### Funding details

None.

### Acknowledgements

None.

### Conflict of interests

Authors declare that there is no conflict of interest.

### References

1. Collier, DA. The service sector revolution: The automation of services. *Long Range Planning* 1983;16(6):10–20.
2. Wahyudi WA, Syazilawati M. Intelligent voice-based door access control system using adaptive-network-based fuzzy inference systems (ANFIS) for building security. *Journal of Computer Science*. 2007;3(5):274–280.
3. Monday HN, Li JP, Nneji GU, et al., Design of an Improved Cost Effective Electronic Locking System. *IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. 2018
4. Dahe, SA. Study conceptual design of biometrics technology in door lock security system. *Journal Of Kerbala University*. 2015. 13(3):312–328.
5. Rahman MM, Ali MS, Akther MS. Password Protected Electronic Lock System for Smart Home Security. *International Journal Of Engineering Research & Technology*. 2018;7(4):541–544.
6. Madhusudhan M, Shankaraiah. Implementation of automated door unlocking and security system. *International Journal of Computer Applications*. 2015;5:8.
7. Jagdale R, Koli S, Kadam S, et al. Review on intelligent locker system based on cryptography, wireless & embedded technology. *International Journal of Technical Research and Applications*. 2016:75–77.
8. Johnson J, Dow C. Intelligent door lock system with encryption. *US Patent Application Publication*. 2016:1–92.
9. Gordillo, CDA. *Realce e Reconhecimento de Voz Contínua em Ambientes Adversos*. PUC-Rio de Janeiro; 2018.
10. Lyons, J. *Python\_speech\_features's documentation*. 2013.
11. Rao KS, Manjunath KE. *Speech recognition using articulatory and excitation source features*. Springer; 2017.
12. Haykin S. *Neural Networks: A Comprehensive Foundation*. 2nd edn. 2001.
13. Schaul, T et al. PyBrain. *Journal of Machine Learning Research*. 2010;11:743–746.