



## Article

# Development of Smart Healthcare System Based on Speech Recognition Using Support Vector Machine and Dynamic Time Warping

Ahmed Ismail <sup>1,\*</sup> , Samir Abdlerazek <sup>1</sup>  and Ibrahim M. El-Henawy <sup>2</sup>

<sup>1</sup> Information System Department, Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt; samir.abdelrazek@mans.edu.eg

<sup>2</sup> Department of Computer Science, faculty of Computers and Information, El-Zagazig University, Zagazig, Sharqiyah 44519, Egypt; henawy2000@yahoo.com

\* Correspondence: A.ebada@students.mans.edu.eg or A.ebada@outlook.com

Received: 11 February 2020; Accepted: 17 March 2020; Published: 19 March 2020



**Abstract:** This paper presents an effective solution based on speech recognition to provide elderly people, patients and disabled people with an easy control system. The goal is to build a low-cost system based on speech recognition to easily access Internet of Things (IoT) devices installed in smart homes and hospitals without relying on a centralized supervisory system. The proposed system used a Raspberry Pi board to control home appliances through wireless with smartphones. The main purpose of this system is to facilitate interactions between the user and home appliances through IoT communications based on speech commands. The proposed framework contribution uses a hybrid Support Vector Machine (SVM) with a Dynamic Time Warping (DTW) algorithm to enhance the speech recognition process. The proposed solution is a machine learning-based system for controlling smart devices through speech commands with an accuracy of 97%. The results helped patients and elderly people to access and control IoT devices that are compatible with our system using speech recognition. The proposed speech recognition system is flexible with scalability and availability in adapting to existing smart IoT devices, and it provides privacy in managing patient devices. The research provides an effective method to integrate our systems among medical institutions to help elderly people and patients.

**Keywords:** speech recognition; healthcare; elderly people; patients; smartphones; dynamic time warping; IoT; smart home

## 1. Introduction

In recent years, there have been many new solutions designed for elderly people, blind people, handicapped people, and patients living in their own homes without help to give them a way to be independent. The most common problem for people who suffer from a disability is controlling home devices that require localization or help from others [1]. An Internet of Things (IoT) home is a home that has integrated sensory devices connected through a network to help the user obtain information and alerts or to control their home appliances. A smart home is a home that has IoT devices that inform the user about what is happening in the house and help the user control those devices.

Using speech recognition to control IoT-connected devices in a smart home is a great solution to facilitate the control of the devices. Smart homes are homes equipped with sensors and actuators, which can be controlled through wireless networks [2]. Many researchers used microphones as an input to complete interactions at smart homes and have added many available solutions for sending speech commands to control smart home devices [3,4]. There are some challenges in developing a speech recognition system that must be addressed involving distance, security, and noise [5].

This paper presents a proposed speech recognition system to control appliances at smart homes or smart hospitals. Many researchers have developed systems that depend on vocal commands, such as wheelchair commands for people with dysarthria [4] and vocal control of appliances and devices [6]. Voice-based systems can provide a smart home with a speech recognition system with features to help visually impaired and elderly people to control devices [7].

Smart homes usually provide different solutions for managing homes from inside or even from outside the home through the Internet. Adaptive smart homes primarily use machine learning models to recognize patterns from daily activities and automate the actions and rules that mimic these activities [8]. The main target of IoT smart homes is reducing physical movements and actions that need to be carried out by humans by catering to their needs using advanced sensors.

Dynamic Time Wrapping (DTW) is an algorithm for measuring the similarity between two-time series which may vary in timing. DTW works well with spoken words processing and automatic speech recognition. Many types of data which seem to not be time series can be transformed into time series, such as speech, DNA (Deoxyribonucleic Acid), historical handwriting, and shapes [9]. DTW is very efficient at finding the differences between two-time series by calculating the matrix distance between them.

This paper focuses on controlling smart home devices using a hybrid system for speech recognition and it is divided into four parts. Firstly, the introduction gives an overview of the problem and an overview of the solutions in general. In the second part, this paper presents an introduction to existing home automation technologies. In the third section, the proposed system and the system components are discussed. In the fourth section, the results are discussed. Finally, the paper presents future work prospects and concludes the proposed system.

## 2. Related Work

M. Rahman et al. [10] proposed an automatic speech recognition system based on Support Vector Machine (SVM) with the assistance of DTW for speakers of the isolated Bangla language. They collected data from 40 speakers for five different Bangla words with the highest acoustic and noise-proof environment. Mel Frequency Cepstrum Coefficients (MFCCs) were used as static features from the speech signal. They used DTW after determining the feature vectors for feature matching. They proposed a reliable model that was tested by 12 speakers, and the recognition rate that the system achieved is 86.08%. The limitation of their work was that it was only designed for isolated Bangla speech recognition and therefore cannot be generalized.

X. Kong et al. [11] presented an evaluation for sub-phonemics to study the effect of noise on speech recognition. They used some features such as manner, place, and voicing error patterns as grey-scale confusion matrices, and as distinctive-feature-distances in the comparison. They found that the place features are most susceptible to misperceptions in white noise, followed by manner features, and then voicing features. Interactions among the IoT devices in smart homes can be executed via user interaction techniques such as human speech [12,13].

I. Mohamad et al. [14] developed a system that used a voice-matching methodology by extracting images, data, and user voice inputs. They used Principal Component Analysis (PCA) for the extraction of images and MFCC for the extraction of voice inputs. The system achieved an accuracy of 87.14% with the DTW method and 92.85% with the Euclidean distance method.

G. Ruben et al. [15] proposed an Automated Speech Recognition (ASR) tool as an educational platform for visually impaired people. They achieved good accuracy levels for speech recognition, but the system did not reach a high enough level to help users use it independently.

Y. Mittal et al. [16] worked on a voice-controlled smart home based on an IoT connection that could be installed easily and at a low cost. It also required minimal training and maintenance overhead. They suggested that using wireless technology can further be used for better performance. However, their solution suffers from authentication lockage and distance problems; the closer the user is to the sound device, the more suitable the system is. Additionally, the developed mobile application

system can solve the distance problem, which requires that users be within a certain distance of the microphone.

E. Essa et al. [17] proposed a speech recognition system for isolated Arabic words by using a combined classifier based on backpropagation with different parameters and architecture. They used some recorded recitals from the Holy Quran by recruiting 10 famous reciters from different countries to test the system. The system could reach very high rates for speech recognition of up to 96% using MFCC. The system suffered from excessively large vocabulary sets and adaptation problems during the process of individual classifier combination.

In the study by M. De Brouwer et al. [18], a cascading reasoning framework was proposed to provide healthcare solutions with a responsive intelligent system that could convey data between sensors at smart homes and smart hospitals with pipelines. This framework was designed to remove latency from the process of transforming data and to improve the system's responsiveness.

There are a lot of underlying issues when providing smart solutions based on IoT devices. As a result, we studied the main issues that should be considered from the literature research in References [19,20]. The authors stated that security, and privacy are the main concerns for smart city applications. In the proposed system, the proposed solution worked on these issues to provide the users with a safe environment.

In another study by L. Sánchez et al. [21], they described a platform called SoundCity which is based on data streaming from smartphones to enable interoperability among heterogeneous IoT devices such as smart home devices and smartphones. They presented ways to integrate IoT devices and the other protocols from smartphones to control the devices. The big benefit of using smartphones to control IoT devices is that there is no coverage limit. The proposed system used an integration system to control smart home appliances with smartphones without restrictions on a user's location.

These existing solutions showed that they can be employed for speech recognition applications with only a little bit of degrading of the performances. However, all of those presented combination schemes of SVM and other methods are complicated and computationally expensive. Unlike an existing method, the proposed hybrid system of SVM and DTW does not require a lot of corpus training to train the speech recognition process and instead, only requires a predefined dictionary that contains the basic speech command and their templates. By using a supervised machine learning method that is based on a predefined dictionary with the controlling speech commands, the solution can automatically match the template commands with user commands. The proposed model was evaluated by using a smart home data set, using SVM alone, and using the hybrid system. The proposed SVM-DTW provides a direct and simple scheme to merge SVM and DTW and matches DTW templates for a speaker verification system in addition to using the hybrid model for the speech recognition process. The proposed SVM-DTW for smart home appliances controlling and accessing provides several advantages, as follows:

- (1) It provides user authentication to give grants to access the smart home devices.
- (2) It increases the probability of recognizing the speech commands with a more convenient method than using SVM alone.
- (3) It provides an efficient scheme to integrate SVM and DTW for speech recognition methods.

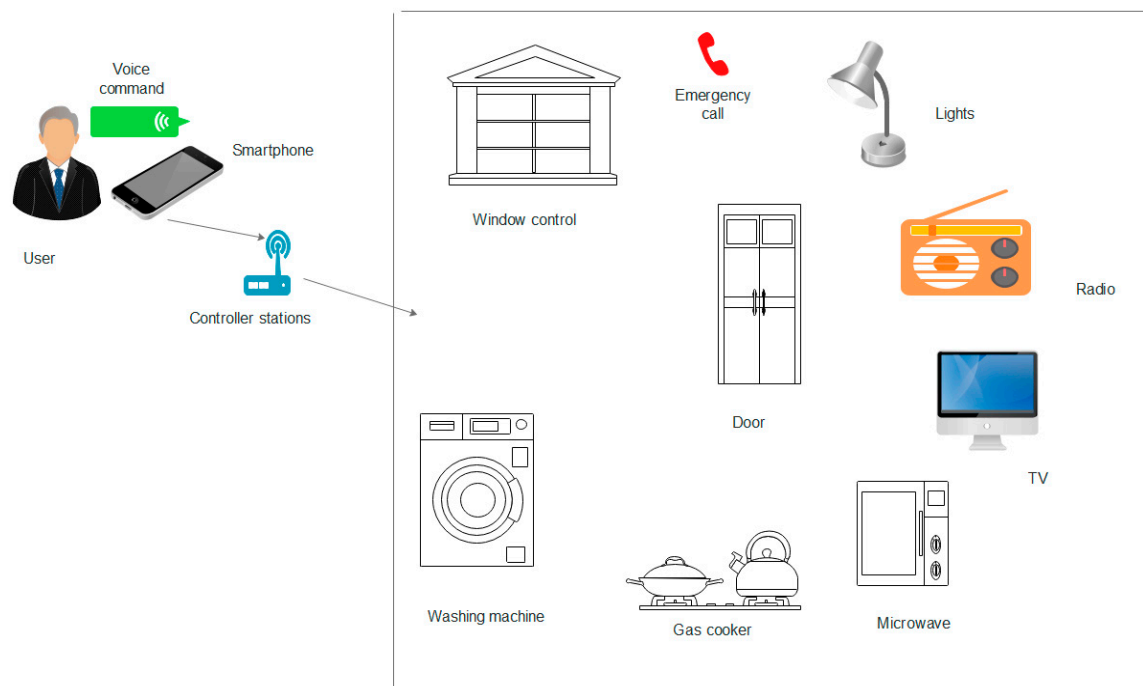
### 3. The Proposed Speech Recognition System

The proposed system is a smartphone-dependent system for speech recognition to execute one command based on the matching against the user command and the recorded speech templates. Each user should record these commands for different home appliances to train the system to recognize these commands using the microphone of the smartphone (to address distance issues) and voice command matching. Then, the commands are sent only through a smartphone. A machine learning model is used to match the voice commands from elderly people, patients, or disabled people based

on an expandable dictionary of predefined user commands to help the system recognize the user's speech commands.

### 3.1. Structure of the Proposed System

The proposed system is composed of three components: a smartphone, a controlling system, and the devices (Figure 1). The smartphone receives the voice command and sends it to the controller, which is a router that receives the command and IP address of the connected device. The control command is easily sent through the mobile connection from anywhere, as the command is sent by a smartphone, so it is very easy to use and solves the distance problem.



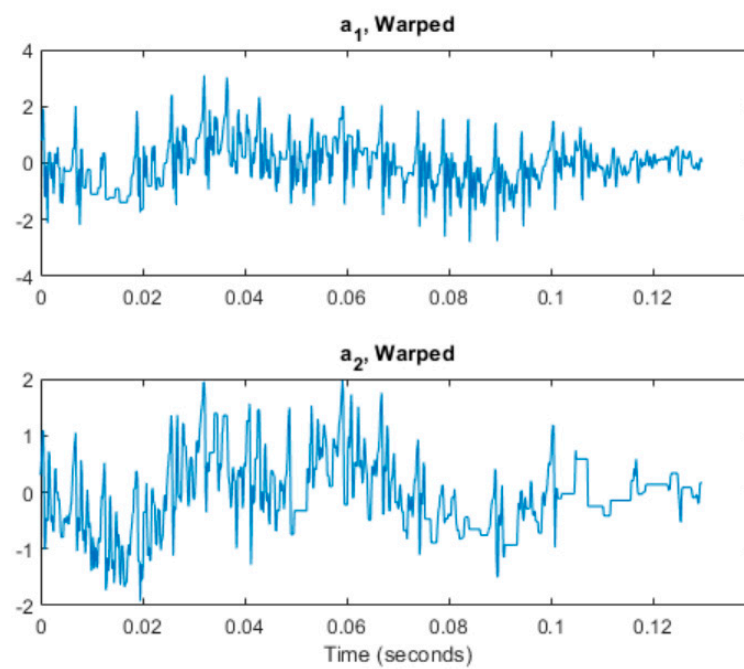
**Figure 1.** The proposed system structure of smart devices control.

The proposed system used SVM with DTW to detect the user's commands quickly and efficiently. The DTW algorithm measures the similarity between two time series, which may differ in speed or time. The SVM method is trained on biological similarity detection [22,23].

#### 3.1.1. DTW in the Proposed Speech Recognition System

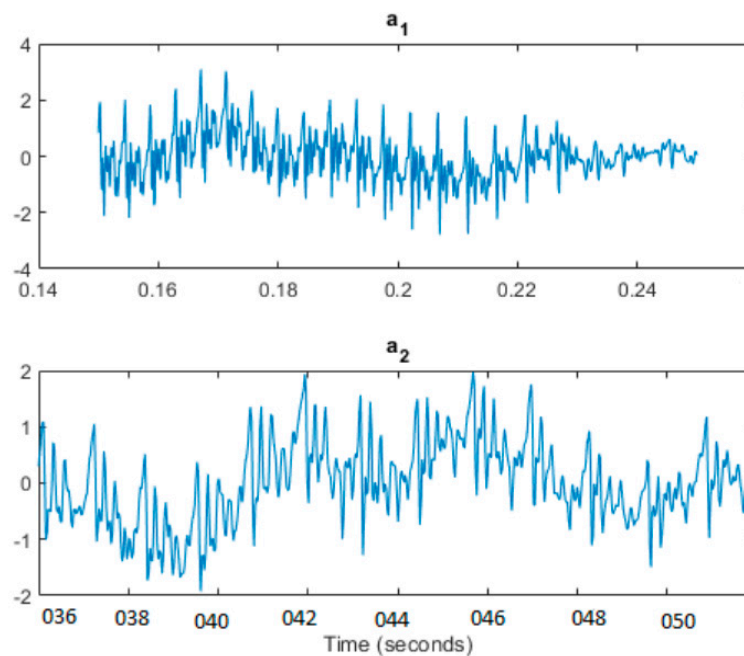
The proposed system used a database of training speech templates. Each training template has an utterance for a given word class from men and women i.e., "door", "close", "one". Each template has more than one record. DTW can find the nearest record by choosing the class of template with the minimum DTW distance by making an alignment between the two series.

The proposed system investigates the efficiency of time-series classification with similarities based on DTW. The recognition of a specific time is faster than the recognition of a sample with a smaller time-series length. For example, the length of the time series varies from 20 to 80, instead of providing recognition across the length of the overall series. Two signals with equivalent features arranged in the same order can appear to be very different because of the differences in durations in their sections. This is shown in Figure 2, where the spoken words used were "The main door open". The same words are used in both examples, but the order turned out in very different ways.



**Figure 2.** Two waves for the same word before DTW.

DTW distorts these durations so that the corresponding features appear at the same location on a common time axis, thus highlighting the similarities between the signals, as shown in Figures 3 and 4. Two main operations are performed to find the similarities between two-time series: the analysis of a single time series and analysis of two-time series.



**Figure 3.** Two waves for the same word after wrapping.

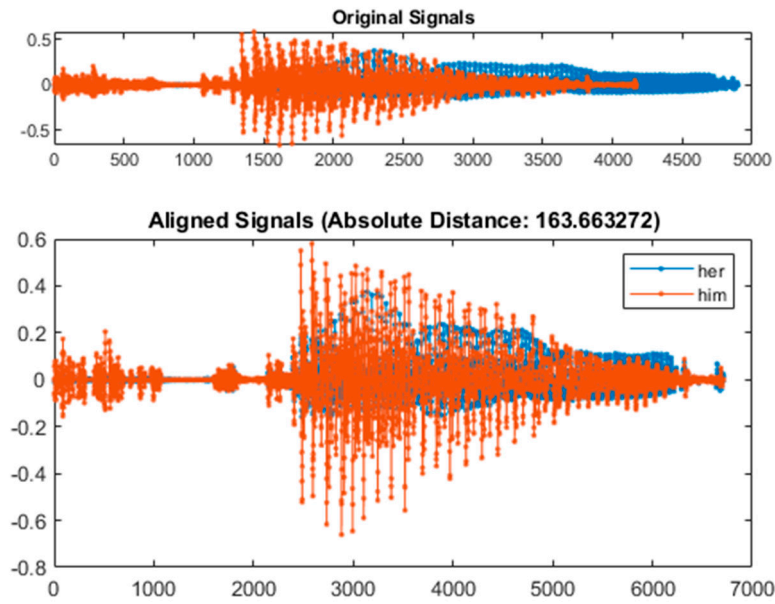


Figure 4. The two waves after DTW minimization.

DTW minimizes the Euclidean distance between the signals. It also computes the shared duration of the warped signals. The time axes are warped so that the absolute distance between the signals is minimized. The original and transformed signals are plotted. The proposed system uses DTW to find the optimal alignment between two given time-dependent sequences. It implements the method on each user input, and it generates a list of recorded matches from specific dictionary words. The target is to choose the closest match based on the time sequence of the user's commands.

When the user records a command on the smartphone, the system uses SVM to compare the command against specific words in a specific time series. The proposed speech recognition system aims to help elderly people, disabled people, and patients to pronounce one of the pre-defined words to control a specific device. Therefore, we used DTW to build a match between each command and its defined category. The SVM is used to measure the local distance between vectors as shown in the Equation (1) below. To measure the distance between two series, for instance,  $a = \{a_1, a_2, \dots, a_M\}$  and  $b = \{b_1, b_2, \dots, b_m\}$ . Let  $M(a, b)$  be  $m \times m$  pointwise distance matrix between  $a$  and  $b$ , where  $M_{i,j} = (a_i - b_j)^2$ . The DTW distance between series is the path through  $M$  that minimizes the total distance subject to constraints on the amount of warping allowed, as in Equation (1).

In the proposed system, DTW plays two roles: it is an authentication tool to grant the user access to control the devices and also increases the effectivity of SVM of the speech recognition process. As shown in Figure 5, for the process of matching, a lower distortion between the testing data and the recorded template means there is a higher matching degree. The identification process for the user is composed of  $T$  frames where  $t$  is the feature vector; the recorded speech template involves  $R$  frames where the arbitrary frame in the recorded template is indicated as  $r$ ; the similarity between  $T$  and  $R$  frames can be represented by  $d[T(t), R(r)]$ ; the starting point of the overall comparison path is  $T(1), R(1) = (1,1)$ , and the endpoint of both of the frames is  $T(M), R(N) = (T, R)$ . The path optimal distance for the similarity between  $R$  and  $T$  is based on Equation (2).

$$DTW(a + b) = \sum_{k=0}^n (a_i - b_i)^2 \quad (1)$$

$$Optimal a = \sum_{m=1}^M a(T(m), R(m)) \quad (2)$$



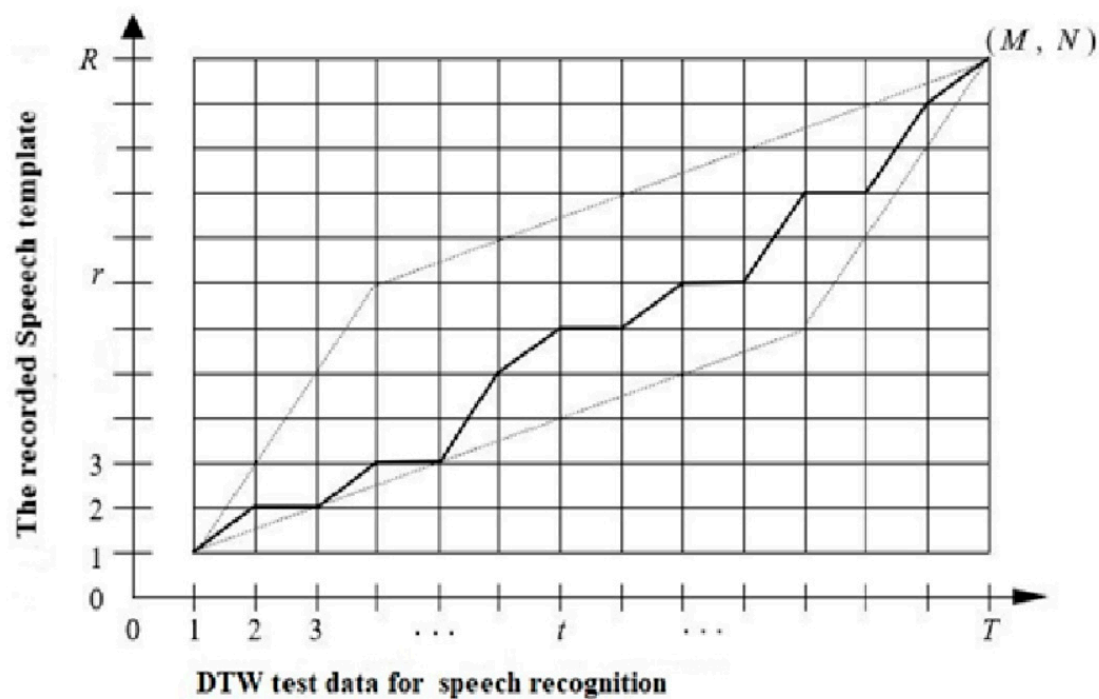


Figure 5. DTW method in speech recognition.

### 3.1.2. SVM in the Proposed Speech Recognition System

The SVM is used to measure the local distance between vectors. The speech commands are recognized using smartphones as a tool to recognize the user's voice and to send commands to the devices. SVM does not perform well for large datasets because it requires a high training time and it also takes more time in training compared to the Naïve Bayes algorithm. The proposed system works only for a limited pre-defined command, therefore the SVM can work with small datasets very well. The SVM is one of the most successful pattern classifications algorithms. The system detects voices which are analyzed by using the smartphone microphone; afterwards, the Open Ears [24] smartphone application analyzes the voice based on the two possible groups (i.e., either on or off for the lights, grades, and levels for television sound or words for security issues), as shown in Figure 6. The user starts the command with the device name and then the needed action based on the device group. For instance, in Table 1, for the air conditioner, the user needs to add a command with the device name with the action and the desired degree for the room. To control the TV, the user needs to start the command with "TV" then the desired action "on" or "off" then the channel number "2".

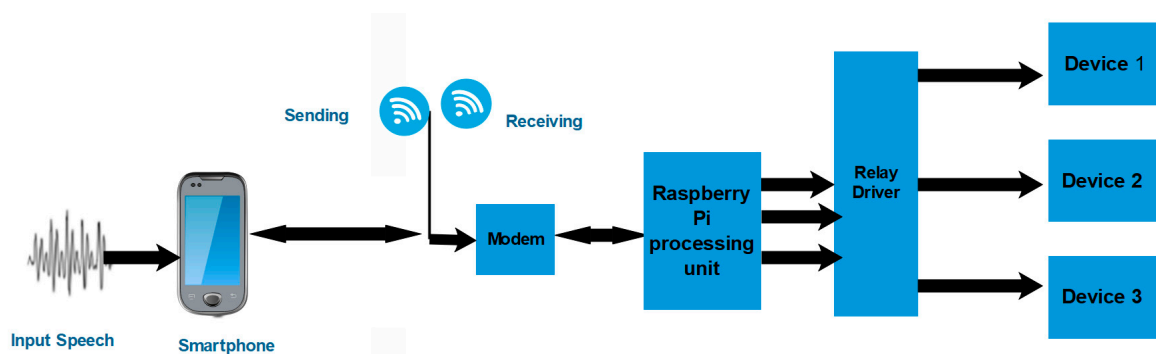


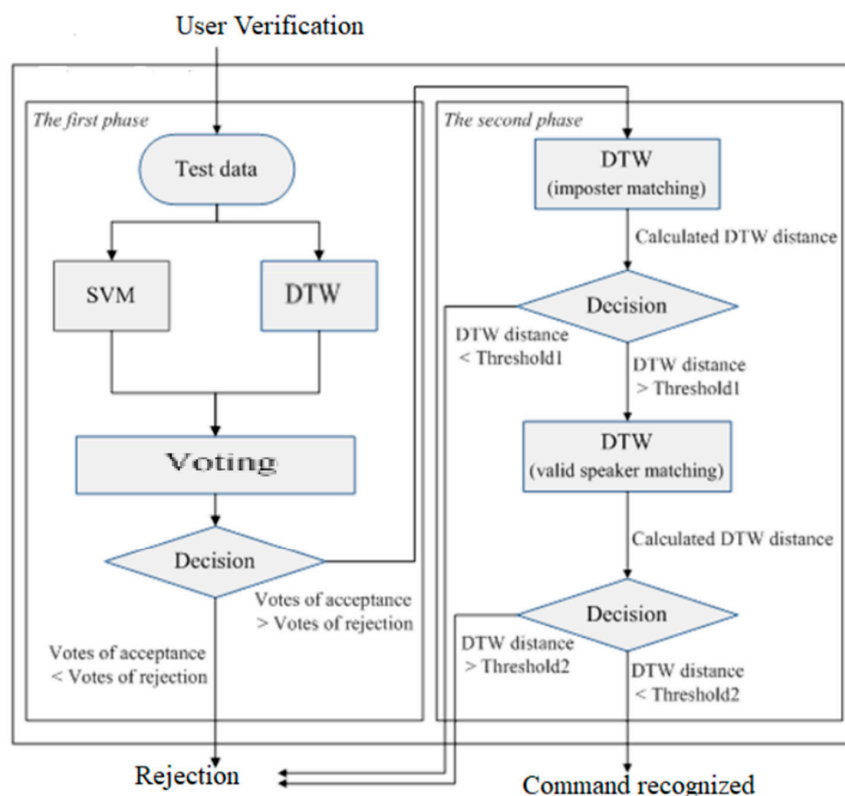
Figure 6. The proposed system with Raspberry Pi architecture.

**Table 1.** A user command example.

Number	Device	Action
1	TV	ON-2
2	Air conditioning	ON-23

The proposed system ensures the privacy for each user with a smartphone to control devices. The proposed system is working with nonnative English using an ASR platform and proposes a system for mobile phones that do not require a special voice system. The system depends on the mobile phone microphone, which uses the user's speech on a mobile application to detect the speech sounds. Finally, the ASR result is sent to the Raspberry Pi board a command, which processes the command and sends it to the targeted device through the relay driver as the controller. The proposed system as shown in Figure 6 uses the wireless connection between the mobile phone and Raspberry Pi board to send the commands. It directly connects the Raspberry Pi board to a display monitor to display the current command and the action being carried out. It also connects the Raspberry Pi board to smart devices such as a smart TV, smart door and other devices through a relay driver responsible for controlling the devices.

The system works when the user starts the mobile application, which listens to the user through the phone's microphone. Then, the user gives a command to control smart home devices, as shown in Figure 7. There are two steps for recognition: the first step is using DTW, where the user speech is compared to the template to ensure if the user is authenticated. Then the user says the device name to control it. For that purpose, there is a vote from DTW and SVM to check if the device name is recognized from among the saved words that the user can use to state a command to the device. The system uses DTW. Thus, instead of comparing the stored samples, and the user inputs over time, it compares only a specific time for every command to increase the probability of recognition.

**Figure 7.** The proposed system block diagram.



Applying a voting method based on machine learning method with labels provided a powerful classifiers using defined speech commands. The proposed system used a voting algorithm that exploits the individual predictions of SVM and DTW based on the most efficient self-training labeled algorithm. Using the major voting to find the matching between the templates and speech commands could improve the classification efficiency of voting by using DTW matching as a base for recognizing speech. The novelty of the presented system increased recognition accuracy of voting using a labeled machine learning algorithm.

The speech command for the devices depends on the device type and the controlling type. Controlled devices in smart homes belong to one of two types. The first group is on/off devices, such as doors or windows. The IoT devices are composed of groups according to the commands. The devices are controlled by on or off commands or by grades and levels. For example, to switch off the lights in the hall, the voice command is 'Hall' and then 'H-Off'. The smart home's utility group comprises some basic functions such as curtains controls. To open the curtains, the voice used command is 'Curtain'. 'Help' is a command parameter that sounds an alarm in the smart home to alert the residents and neighbors. 'Silence' is command is used to silence the alarm. The 'Light' command is followed by the 'Room' command and either the 'Off' command or the 'On' command to turn the room light off or on.

### 3.2. Speech Recognition Process

The proposed system used the out-of-speaker (OOS) detection algorithm enhanced by defining specific words that were used for controlling the smart home [24]. The proposed OOS detection algorithm non-linearly reflects the feature vectors from a low-dimensional space into a high-dimensional space, which is used to enlarge the differences between different classes, further classifying the different data to provide an effective way to describe the speech feature distribution. The pseudo-code for the Algorithm 1 is:

---

#### Algorithm 1 Speech Recognition

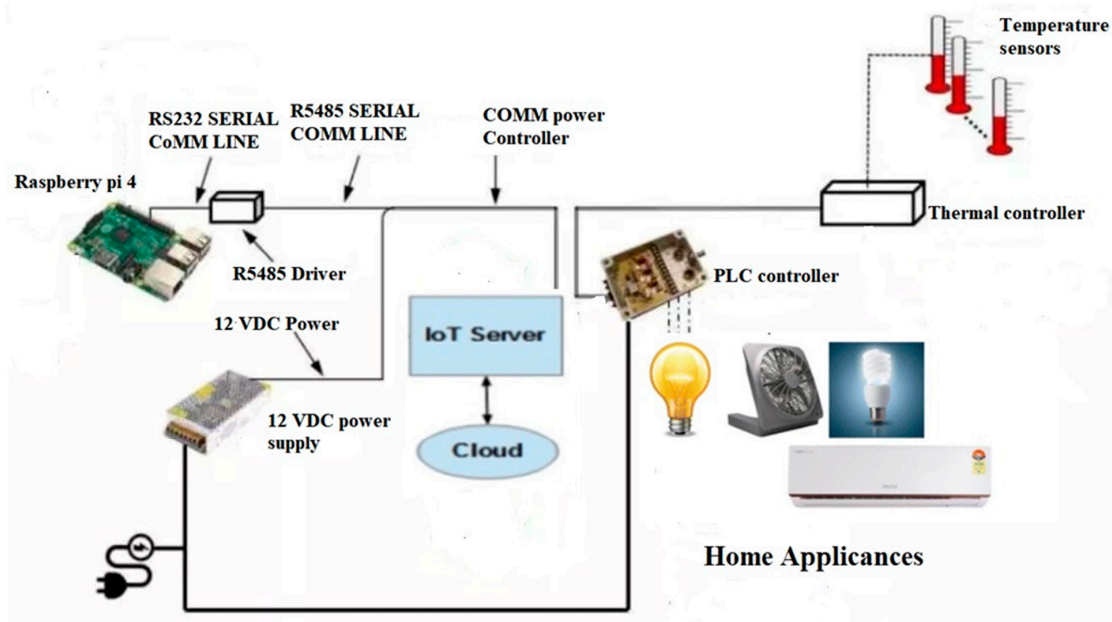
---

1. Iterate until convergence
  2. **for each** sequence  $S_i$  in  $D$
  3.  $A_i = \text{GetAlignment}(\text{DTW}(S_i, \text{av}))$
  4. **for each** observation  $j$  in  $\text{av}$
  5.  $\text{av}[j] = \text{mean}([A_1[j] \ A_2[j] \ \dots \ A_n[j]])$
- $S_i$  the sequence  
 $D$  represents the whole time series  
 $\text{av}$  represent the average
- 

The proposed speech recognition system does not depend on the order of the points in the recorded voice samples. The system can detect an unenrolled speaker who is excluded from the trained speaker model. The proposed system depends on using the defined word for controlling devices in smart homes to make the general process of detection faster and decrease the probability of false alarms. The proposed system also worked on the problem of how elderly and disabled people and patients pronounce words.

### 3.3. Healthcare Speech Recognition System Components

The proposed system provides a healthcare system for smart homes and clinics to control devices based on speech recognition. The proposed speech recognition system is based on Raspberry. The Raspberry Pi board is connected to the home appliances through a relay driver to control the devices. The system used serial cables between the controller and the Raspberry Pi board, as shown in Figure 8.



**Figure 8.** The proposed voice recognition system components for smart homes.

The Raspberry Pi board and the entire system is connected to a central power system as a power source. The system is connected to the IoT server, which allows the system to be controlled from outside the home or clinic through the cloud. The temperature sensors, thermal controller, lighting led, fans, lighting systems, and all appliances are connected directly to the PLC (Programmable Logic Controllers) controller to execute the commands which come from the Raspberry Pi board. Raspberry Pi provides multiple choices and an easy way to connect to smart homes, sensors, and connected devices efficiently [25].

## 4. Evaluation

### 4.1. Dataset and Settings

We tested the proposed system using MATLAB installed on a laptop PC (2.2 GHz with 8 GB RAM) with an SVM library to train the system using a dataset of commands that needed to be recognized. We tested the proposed system using the dataset called Home Automation Speech subset [26], which contains words for the kitchen, living room, office, and washing machine. The experiments were tested by different people's (male and female) voices. The proposed system is a voice recognition system that may help the user with specific tasks indoors or outdoors.

### 4.2. Experimental Results

The proposed system achieved high success rates in detecting sound features of the users to recognize their commands from any location using SVM and DTW. The proposed system solved the authentication problem of controlling smart homes that use speech recognition. The system will be very effective in hospitals and smart homes to help patients, the elderly, and disabled people be independent. The system training is performed by using only a pre-defined word that comes with a specific device to speed up the training and of recognition processes. The words used are on, off, open, close, silent, degree/any number, lower, higher, access, above, down, channel number, and exit. The system achieved high success rates by using specific supported words to increase the probability

of detection and by using DTW with SVM, as shown in Table 2. The System efficiency (SE) is calculated based on Equation (3).

$$\text{System efficiency (SE)} = \frac{\text{Recognized samples} * 100}{\text{total sample number}} \quad (3)$$

**Table 2.** The comparison between the Proposed system with SVM.

Dataset	SVM	SVM with DTW
D1	88%	98%
D2	97%	99%
D3	61%	100%
D4	91%	99%
D5	90%	94%
D6	92%	100%
D7	84%	99%
D8	66%	94%
D9	94%	97%
D10	99%	98%
Accuracy	79%	97%

A total of 20 male and female speakers were asked to record 10 datasets to compare the speech recognition performance and test the system. The records are only words used for smart home appliances controlling such “kitchen, door, fridge, open, close air condition”. The difference between the datasets was that they were recorded by different people with ages that varied between 18 and 28 years old. A total of 200 records were tested. Table 2 shows the recognition performance of the SVM only and the SVM with DTW. The hybrid SVM with DTW achieved a very high success rate of 97% compared with only the SVM.

#### 4.3. Discussion

The proposed speech recognition system operates on finite-length segments of speech commands. the command can be up to one minute long. The system evaluation is based on the right recognition true positive (TP) and true negative (TN) and the irregular speech recognition represented as false positive (FP) and false-negative (FN). The overall performance of these algorithms is presented in Table 3. The evaluation module gives accuracy, precision, recall, and F1 metrics that collectively show how good the model is based on the test to the evaluate module. The precision, recall, and F1 score are calculated as in Equations (4)–(6).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

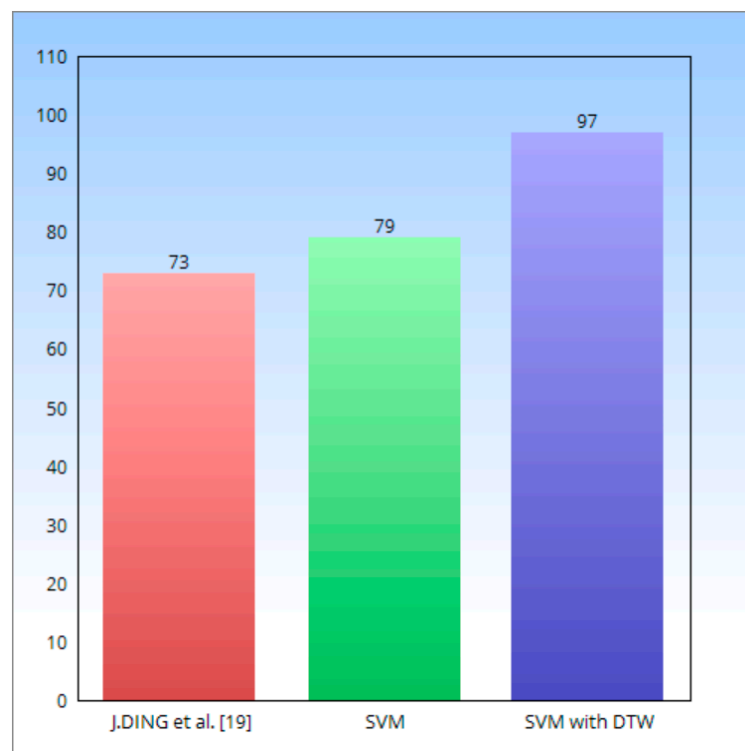
$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$F1 = \frac{2(\text{Precision} - \text{Recall})}{\text{Recall} + \text{Precision}} \quad (6)$$

**Table 3.** The proposed speech recognition system evaluation.

Accuracy	0.976
Precision	0.972
Recall	0.958
F1	0.965

The most difficult problem that we solved was enhancing recognition (i.e., microphone problems when there are different distances). We did this by developing a system dependent on the user's smartphone. We used a few words for the recognition process to decrease the probability of false detections. The proposed systems used words such as on, off, open, close, increase, decrease, microwave, door, and window. We compared the proposed speech recognition system performance to the results from J. Ding et al. [23] as shown in Figure 9. They used speech recognition for user verification applications and used the DTW method just for verification, instead of for the speech recognition process as well. They mainly applied SVM with a Gaussian mixture model (GMM) and achieved an accuracy of 73%. The proposed system proved better accuracy in matching the speech commands with 97%, which was better than for applying SVM only. The proposed system can help patients and elderly people at smart homes to control home appliances.



**Figure 9.** The proposed system performance's comparison with existing work.

The proposed system was compared to another work for smart home applications to present the impact of speech recognition applications for smart homes to help patients and the elderly. In Reference [27], the authors proposed a solution like our proposed speech recognition system to use daily activities using speech commands to control smart home devices. They used the Markov Logic Network (MLN) algorithm and provided a high performance at 85%, with a recall rate of 92.7%. Our proposed speech recognition provided a better performance in comparison with the previous solutions as shown in Figure 10, but this does not mean that the system can be helpful for all situations because other relevant factors should be considered such as distance, noise, and stress of speech. The experiment results can be affected by the location of the human, the speech command recognition, and distress detection in real-time.

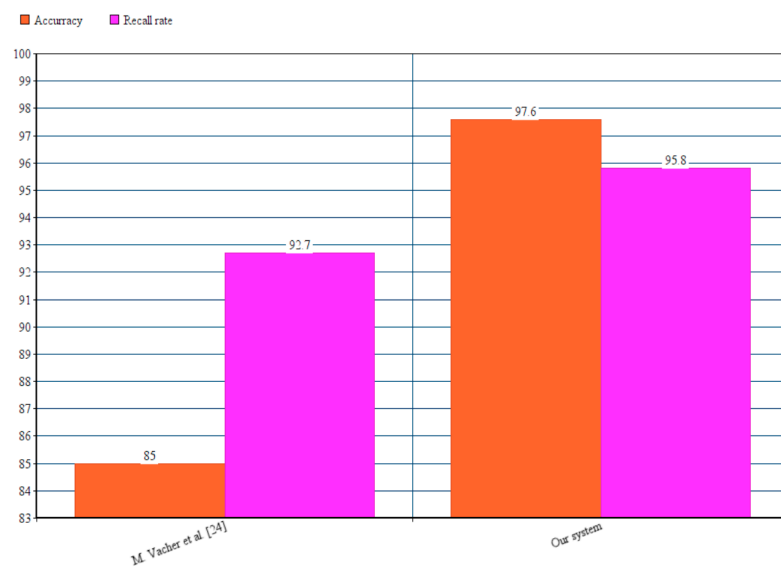


Figure 10. Comparison graph analysis.

## 5. Conclusions

The proposed system is an effective system that can be used for generalized in all hospitals and smart homes for patients and elderly people. The system is quick in employing voice recognition, as the system uses only defined words for controlling the smart home or hospital. Thus, the process of recognition depends on only using specified commands to be compared to the command word. The proposed system solved many problems from previous systems, such as the speech distance issues, the system privacy, the accuracy of the speech recognition, and the command speed or how fast the user says the voice command. The proposed system is a low-cost system because it is based on smartphones and low-cost boards (i.e., Raspberry Pi boards). The system is reliable, efficient, and secure for controlling devices in smart homes and clinics.

The proposed hybrid system achieved a high accuracy of 97%, which is higher than the accuracy of using an SVM only (79%). The system used DTW with SVM to overcome the differences among sound segmentations.

The system's limitation is the difficulty of speech recognition if the user's voice is affected by illness or is not clear enough to be detected. Useful future work would involve research on integrating multiple recognition systems such as voice recognition and video recognition to help patients who suffer from speech problems.

**Author Contributions:** Conceptualization, A.I.; Formal analysis, S.A.; Project administration, I.M.E.; Software, A.I. All authors have read and agreed to the published version of the manuscript.

**Conflicts of Interest:** The authors declare that there is no conflict of interest.

## References

1. Peetoom, K.; Witte, L.; Joore, M.; Dirksen, M.L.C. Literature review on monitoring technologies and their outcomes in independently living elderly people. *Disabil. Rehabil. Assist. Technol.* **2015**, *10*, 271–294. [[CrossRef](#)] [[PubMed](#)]
2. Ni, Q.; Hernando, A.G.; Cruz, I. The elderly's independent living in smart homes: A characterization of activities and sensing infrastructure survey to facilitate services development. *Sensors* **2015**, *15*, 312–362. [[CrossRef](#)] [[PubMed](#)]
3. Choi, D.; Choi, H.; Shon, D. Future changes to smart home based on AAL healthcare service. *J. Asian Archit. Build. Eng.* **2019**, *18*, 190–199. [[CrossRef](#)]
4. Helal, S.; Bull, C. From Smart Homes to Smart-Ready Homes and Communities. *Dement. Geriatr. Cogn. Disord.* **2019**, *47*, 157–163. [[CrossRef](#)] [[PubMed](#)]

5. Gomez, C.; Chessa, S.; Fleury, A.; Roussos, G.; Preuveneers, D. Internet of Things for enabling smart environments: A technology-centric perspective. *J. Ambient Intell. Smart Environ.* **2019**, *11*, 23–43. [[CrossRef](#)]
6. Park, G.; Kim, H. Low-Cost Implementation of a Named Entity Recognition System for Voice-Activated Human-Appliance Interfaces in a Smart Home. *Sustainability* **2018**, *10*, 488. [[CrossRef](#)]
7. Zhu, M.; Wu, X.; Lu, Z.; Wang, T.; Zhu, X. Long-term speech information-based threshold for voice activity detection in massive microphone network. *Digit. Signal Process.* **2019**, *94*, 156–164. [[CrossRef](#)]
8. Rashid, P.; Cook, D. Keeping the Resident in the Loop: Adapting the Smart Home to the User. *IEEE Trans. Syst. Man Cybern.* **2009**, *39*, 949–959. [[CrossRef](#)]
9. Tian, Y.; Wang, Z.; Lu, C. Self-adaptive bearing fault diagnosis based on permutation entropy and manifold-based dynamic time warping. *Mech. Syst. Signal Process.* **2019**, *114*, 658–673. [[CrossRef](#)]
10. Rahman, M.; Dipta, D.; Hasan, M. Dynamic Time Warping Assisted SVM Classifier for Bangla Speech Recognition. In Proceedings of the 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 8–9 February 2018; pp. 1–6.
11. Kong, X.; Choi, J.Y.; Shattuck-Hufnagel, S. Testing automatic speech recognition systems in comparison with human perception results using distinctive feature measures. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 5810–5814.
12. Lecouteux, B.; Vacher, M.; Portet, F. Distant speech processing for the smart home: Comparison of ASR approaches in scattered microphone network for voice command. *Int. J. Speech Technol.* **2018**, *21*, 601–618. [[CrossRef](#)]
13. Koskela, T.; Mattila, K. Evolution towards smart home environments: An empirical evaluation of three user interfaces. *Pers. Ubiquitous Comput.* **2004**, *8*, 234–240. [[CrossRef](#)]
14. Mohamad, I.; Zainal, I.; Guntur, R. Implementation of Dynamic Time Warping algorithm on an Android-based application to write and pronounce Hijaiyah letters. In Proceedings of the 2016 4th International Conference on Cyber and IT Service Management, Bandung, Indonesia, 26–27 April 2016.
15. Ruben, G.; Muñoz, J.; Salazar, J.; Duque, N. Voice Recognition System to Support Learning Platforms Oriented to People with Visual Disabilities. In Proceedings of the International Conference on Universal Access in Human-Computer Interaction, Toronto, ON, Canada, 17–22 July 2016; Springer: Cham, Switzerland, 2016; pp. 65–72.
16. Mittal, Y.; Sharma, S.; Toshniwal, P.; Singhal, D.; Gupta, R.; Mittal, V. A voice-controlled multi-functional smart home automation system. In Proceedings of the Annual IEEE India Conference (INDICON), New Delhi, India, 17–20 December 2015.
17. Essa, E.; Tolba, A.; Elmougy, S. Combined classifier based Arabic speech recognition. In Proceedings of the INFOS 2008, Cairo, Egypt, 27–28 March 2008; pp. 27–29.
18. de Brouwer, M.; Ongena, F.; Bonte, P.; de Turck, F. Towards a Cascading Reasoning Framework to Support Responsive Ambient-Intelligent Healthcare Interventions. *Sensors* **2018**, *18*, 3514. [[CrossRef](#)] [[PubMed](#)]
19. Lytras, M.; Visvizi, A. Who uses smart city services and what to make of it: Toward interdisciplinary smart cities research. *Sustainability* **2018**, *10*, 1998. [[CrossRef](#)]
20. Visvizi, A.; Lytras, M. It's not a fad: Smart cities and smart villages research in European and global contexts. *Sustainability* **2018**, *10*, 2727. [[CrossRef](#)]
21. Sánchez, L.; Lanza, J.; Santana, J.; Agarwal, R.; Raverdy, P.; Elsaleh, T.; Fathy, Y.; Jeong, S.; Dadoukis, A.; Korakis, T.; et al. Federation of Internet of Things testbeds for the realization of a semantically-enabled multi-domain data marketplace. *Sensors* **2018**, *18*, 3375. [[CrossRef](#)] [[PubMed](#)]
22. Liao, L.; Noble, W. Combining pairwise sequence similarity and support vector machines for detecting remote protein evolutionary and structural relationships. *J. Biol.* **2003**, *10*, 857–868. [[CrossRef](#)] [[PubMed](#)]
23. Ding, J.; Yen, C.; Cheng, O. A method to integrate GMM, SVM and DTW for speaker recognition. *Int. J. Eng. Technol. Innov.* **2014**, *4*, 1–38.
24. Avila, M.; Ruggiero, R.; Leite, J.; Bueno-Junior, L.; Del-Ben, C.M. Better ears with eyes open: Effects of multisensory stimulation with nonconscious visual stimuli on auditory learning. *bioRxiv* **2019**. [[CrossRef](#)]
25. Singh, P.; Nayak, P.; Datta, A.; Sani, D.; Raghav, G.; Tejpal, R. Voice Control Device using Raspberry Pi. In Proceedings of the 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, UAE, 4–6 February 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 723–728.



26. The Testing Dataset. Available online: [https://de.mathworks.com/matlabcentral/fileexchange/45054-gui-environmental-sound-recognition?s\\_tid=mwa\\_osa\\_a](https://de.mathworks.com/matlabcentral/fileexchange/45054-gui-environmental-sound-recognition?s_tid=mwa_osa_a) (accessed on 21 January 2020).
27. Vacher, M.; Lecouteux, B.; Chahuara, P.; Portet, F.; Meillon, B.; Bonnefond, N. The Sweet-Home speech and multimodal corpus for home automation interaction. In Proceedings of the 9th Edition of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 26–31 May 2014.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).