Implementation of a Hybrid Voice Control System for a Colony of Robots

Mohamed Fezari and Mounir Bousbia-Salah
Department of Electronics, Badji Mokhtar University, Algeria

Abstract: *In this paper, a voice command system for autonomous robots is proposed as a project. The methodology adopted is based on hybrid techniques used in speech recognition which are zero crossing and extremes with dynamic time warping followed by a decision system based on independent methods test results. To implement the approach on a real time application, a personal computer interface was designed to control the movement of a set of robots by simple voice commands. The voice command system for four autonomous robots is designed. The main parts of the robots are a microcontroller from Microchip PIC16F84 and a radio frequency receiver module.*

Keywords: *Speech recognition, voice command, hybrid methods, DTW, robots, microcontroller*

Received January 27, 2007; accepted October 30, 2007

1. Introduction

Human-robot voice interface has a key role in many application fields [1, 4, 10]. Various studies made in the last few years have given good results in both research and commercial applications based on automatic speech recognition systems [6, 11, 17]. This paper proposes a new approach to the problem of the recognition of isolated words, using a set of traditional pattern recognition approaches and a decision system based on test results of classical methods [4, 11, 15] in order to increase the rate of recognition. The increase in complexity as compared to the use of only traditional approach is negligible, but the system achieves considerable improvement in the matching phase, thus facilitating the final decision and reducing the number of errors in decision taken by the voice command guided system.

Moreover, speech recognition constitutes the focus of a large research effort in Artificial Intelligence (AI), which has led to a large number of new theories and new techniques. However, it is only recently that the field of robot and Automatic Guided Vehicle (AGV) navigation have started to import some of the existing techniques developed in AI for dealing with uncertain information.

Hybrid method is a simple, robust technique developed to allow the grouping of some basic techniques advantages. It therefore increases the rate of recognition. The selected methods are: Zero Crossing and Extremes (CZEXM), linear Dynamic Time Warping (DTW), DTW with linear predictive coefficient parameters, Energy Segments (ES), and DTW with cepstral coefficients. This study is part of a specific application concerning robots control by simple voice commands. The application uses nine commands in Arabic words. It has to be implemented on a DSP [7] and has to be robust to any background noise confronted by the system.

The aim of this paper is therefore the recognition of isolated words from a limited vocabulary in the presence of background noise. The application is speaker-dependent. Therefore, it needs a training phase. It should, however, be pointed out that this limit does not depend on the overall approach but only on the method with which the reference patterns were chosen. So by leaving the approach unaltered and choosing the reference patterns appropriately, this application can be made speaker-independent.

As application, a vocal command for a set of robots is chosen. There have been many research projects dealing with robot control, among these projects, there are some projects that build intelligent systems [2, 3, 8]. Since we have seen human-like robots in science fiction movies such as in "I ROBOT" movie, making intelligent robots or intelligent systems became an obsession within the research group. Voice command needs the recognition of isolated words from a limited vocabulary used in AGV system [12, 16].

2. Description of Designed Application

The application is based on the voice command for a set of four robots. It therefore involves the recognition of isolated words from a limited vocabulary used to control the movement of a vehicle.

The vocabulary is limited to five commands, which are necessary to control the movement of an AGV, forward movement, backward movement, and stop, turn left and turn right. Four more command words are used as robot names (red, blue, green and black). The number of words in the vocabulary was kept to a minimum both to make the application simpler and easier for the user.

The user selects the robot by its name then gives the movement order on a microphone, connected to sound card of the PC. A speech recognition agent based on hybrid technique recognises the words and send to the parallel port of the PC an appropriate binary code. This code is then transmitted to the robots via a radio frequency emitter.

The application is first simulated on PC. It includes two phases: the training phase, where a reference pattern file is created, and the recognition phase where the decision to generate an accurate action is taken. The action is shown in real-time on parallel port interface card that includes a set of LED's to show what command is taken and a radio Frequency emitter.

3. The Speech Recognition Agent

The speech recognition agent is based on a traditional pattern recognition approach. The main elements are shown in the block diagram of Figure 1. The pre-processing block is used to adapt the characteristics of the input signal to the recognition system. It is essentially a set of filters, whose task is to enhance the characteristics of the speech signal and minimize the effects of the background noise produced by the external conditions and the motor.

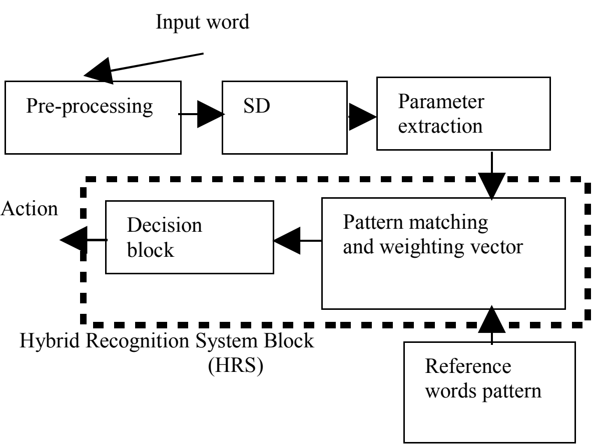

Figure 1. Block diagram.

The Speech Detector (SD) block detects the beginning and end of the word pronounced by the user, thus eliminating silence. It processes the samples of the filtered input waveform, comprising useful information (the word pronounced) and any noise surrounding the PC. Its output is a vector of samples of the word (i.e., those included between the endpoints detected).

The SD implemented is based on analysis of crossing zero points and energy of the signal, the linear prediction mean square error computation helps in limiting the beginning and the end of a word; this makes it computationally quite simple.

The parameter extraction block analyses the signal, extracting a set of parameters with which to perform the recognition process. First, the signal is analysed as a block, the signal is analysed over 20-mili seconds frames, at 256 samples per frame. Five types of parameters are extracted: Normalized Extremes Rate with Normalized Zero Crossing Rate (CZEXM), linear DTW with Euclidian distance (DTWE), LPC coefficients (Ai), Energy Segments (ES) and Cepstral parameters (Ci) [16].

These parameters were chosen for computational simplicity reasons (CZEXM, ES), robustness to background noise (12 cepstral parameters) and robustness to speaker rhythm variation (DTWE) [20].

The reference pattern block is created during the training phase of the application, where the user is asked to enter ten times each command word. For each word and based on the ten repetition, ten vectors of parameters are extracted from each segment and stored. Tests were made using each method separately. From the results obtained, a weighting vector is extracted based on the rate of recognition for each method. Figure 1 shows the elements making up the main blocks for the Hybrid Recognition System (HRS). The Weighting vector is affected some values, these values are used in order to make a decision based on test results of each method separately [5]. The matching block compares the reference patterns and those extracted from the input signal. The matching and decision integrate: a hybrid recognition block based on five methods, and a weighting vector.

A parallel port interface was designed to show the real-time commands. It is based on two TTL 74HCT573 Latches and 16 light emitting diodes (LED), 9 green LED to indicate each recognized word "Ameme", "wara", "Yamine", "Yassar", "Kif", "Ahmar", "azrak", "Akhdar", "Asuad" respectively, and a red LED to indicate wrong or no recognized word. The other LED's were added for future insertion of new command word in the vocabulary example the words: "Faster", "slower" and "light" as shown in Figure 2.

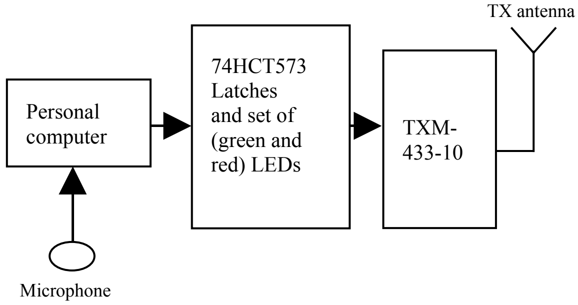

Figure 2. The voice command system.

A voice command system and four autonomous robots that receive voice commands are implemented. The Arabic command words for robots are selected from commands used to control a vehicle ("Ameme", "wara", "yassar", "yamine", and "kif"). The names to differentiate the robots are the colors ("Ahmar", "azrak", "Akhdar", and "Asuad") and their meanings are listed as in Table 1.

Table 1. The meaning of voice commands.

No.	Command	Meaning
1	Amame	Go forward, action on M1
2	Wara	Change direction back, action on M1
3	Yamine	Turn right, action on M2
4	Yassar	Turn left, action on M2
5	Kif	Stop the movement, stops M1 and M2
6	Ahmar	Choose the first robot, red colour
7	Azrak	Choose the second robot, blue colour
8	Akhdar	Choose the third robot, green colour
9	Asuad	Choose the fourth robot, black colour

4. Autonomous Robots

As in Figure 3, the structures of the mechanical hardware and the computer board of autonomous robots in this paper are similar to those in [2] and [8]. However, since the autonomous robots in this paper need to perform simpler tasks than those in [2] and [8] do, these autonomous robots can more easily be constructed. The computer board of each autonomous robot consists of a PIC16F84, with 1K-instruction Electrically Programmable Read Only Memory (EEPROM) [10], two *H* bridges drivers using BD134 and BD133 transistors for DC motors, a RF receiver module from RADIOMETRIX which is the SILRX-433-10 (modulation frequency 433MHz and transmission rate is 10 Kbs) [13], and a four bit micro-switch to fix the address of each robot. Each autonomous robot performs the corresponding task to a received command as in Table 1. Commands and their corresponding tasks in autonomous robots may be changed in order to enhance or change the application.

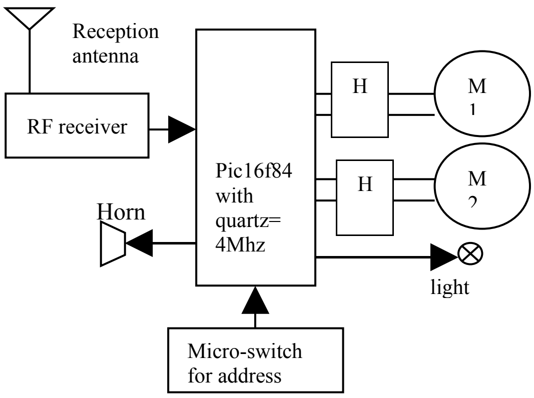

Figure 3. Autonomous robot block diagram.

In the recognition phase, the application gets the word to be processed, treats the word, then takes a decision by setting the corresponding bit on the parallel port data register and hence the corresponding LED is on. The code is also transmitted in serial mode to the TXM-433-10.

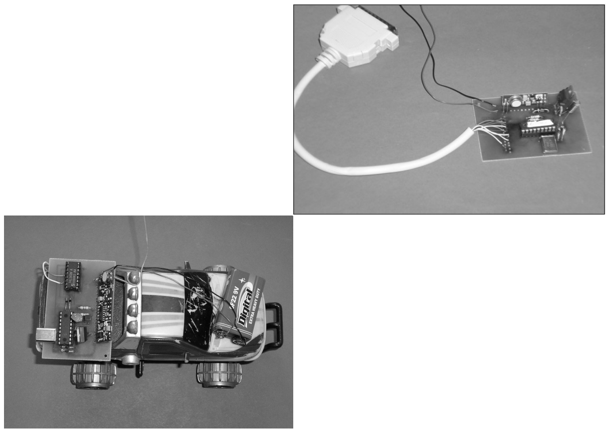

Figure 4. Overview of one robot and parallel interface.

5. Benefits of Designed System

In this work, a voice command system is designed and is constructed to make commands to four autonomous robots.

This voice command system can be applied to other systems as well as robot systems. The followings are the advantages of the proposed system:

- The proposed system, to command and control an autonomous robot by human voices, is totally based on microcontrollers as shown in Figure 5.
- Compared with previous projects that build intelligent systems as in [3] and [12], the cost of the proposed voice command system is much lower.
- An autonomous robot controlled by human voices is one of the projects that can be assigned to a heterogynous research group and therefore require the cooperation of each member. Depending on the research field of group members, this autonomous robot can be divided into several modules and each module can be assigned to one individual researcher. For example, one person designs a voice command system and the other person an autonomous robot, while a third person may work on behaviour of several robots.
- Several interesting competitions of voice-controlled autonomous robots will be possible. In the competition, autonomous robots will perform various tasks by human voice commands. Each team may use its own voice commands different from those of other teams. One example of the competitions is a robot soccer tournament by human voice commands.
- While previous intelligent systems as in [2] and [3] are under a full automatic control, voice-controlled autonomous robots are under a supervisory control. Therefore, it can be used to solve some problems in the supervisory control. One of problems in supervisory control is due to the time delay. The time delay mainly caused by the recognition time of

voices and the time of reception of an RF signal then reaction of the robot, the effect of time delays in controlling autonomous robots can be observed.
- Other systems besides autonomous robots can be combined with the proposed voice command system. For example, a voice-controlled remote controller of consumer electronic products can also be built by using this voice command system and an infrared transmitter/receiver pair if the codes of the remote controller in consumer electronic products are known, example the RC5 code for Philips products [9].

6. Tests on the Voice Command System

The developed system has been tested within the laboratory of LASA the tests were done only on the five command words. Three different conditions were tested:

- The rate of recognition using the classical methods with different parameters.
- The rate of the hybrid method.
- And the effect of noise on the system.

For the two first tests, each command word is uttered 25 times. The recognition rate for each word is presented in Figures 5 and 6.

The effect of noise is tested in the following conditions: (1) outside the Laboratory (LASA) with NSN, (2) outside the LASA with Stationary Noise (SN), (3) inside the LASA with NSN, and (4) inside the LASA with SN. The results are shown in Figure 7 where the numbers in abscess axe corresponds to the order of voice command word as they appear in Table 1.

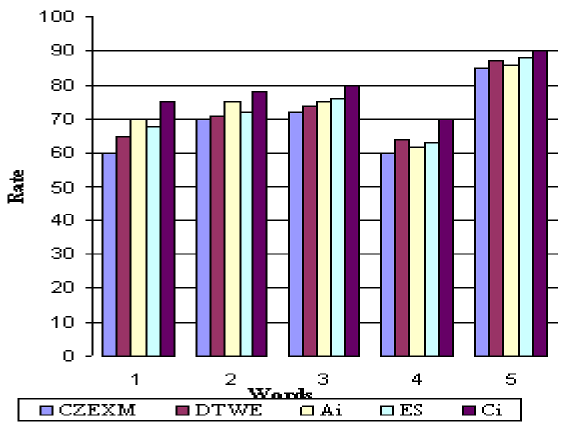

Figure 5. Recognition rate of each method.

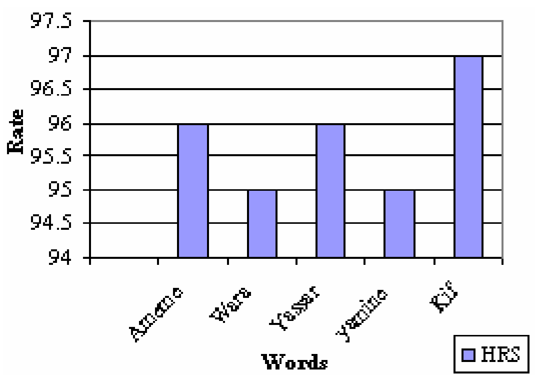

Figure 6. Recognition rate of hybrid thechnique.

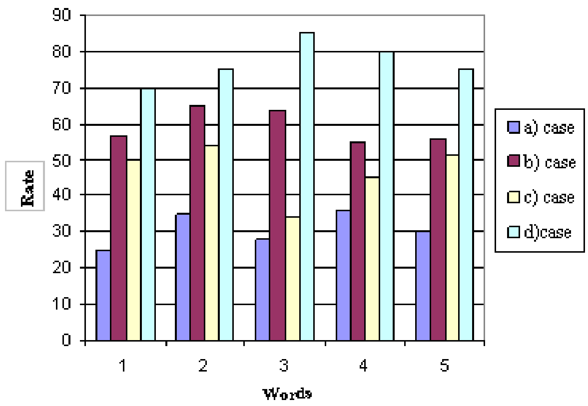

Figure 7. The effect of SN or NSN in and out the laboratory.

7. Conclusions and Future Work

A voice command system for autonomous robots is proposed and is designed based on an HRS for isolated words. The results of the tests show that a better recognition rate can be achieved inside the laboratory and especially if the phonemes of the selected word are quite different. However, a good position of the microphone and additional filtering may enhance the recognition rate.

The use of hybrid technique based on classical recognition methods makes it easier to separate the class represented by the various words, thus simplifying the task of the final decision block. Tests carried out have shown an improvement in performance, in terms of misclassification of the words pronounced by the user. The increase in computational complexit1, however, negligible. Segmentation of the word in three principal frames for the CZEXM method gives better results in recognition rate.

Beside the designed application, a hybrid approach to the implementation of an isolated word recognition agent HSR was used. This approach can be implemented easily within a DSP or a CMOS RISC microcontroller. Since the designed robots consists of a microcontroller, and other low-cost components namely RF transmitters, the hardware design can easily be carried out.

The idea can be implemented easily within a hybrid design using a DSP with a microcontroller since it does not need too much memory capacity. It is possible to increase the number of robots or the number of commands to be executed by a robot. Several interesting applications of the proposed system different from previous ones are possible as mentioned in section 5. We notice that by simply changing the set of command words, we can use this system to control other objects by voice command such as an electric wheelchair or robot-arm movements.

Acknowledgements

I would like to thank Dr. Mokhtar Nibouche for providing support and material during my trainings at University of the West of England, 2005 and 2006. Also, I am Grateful to my colleagues Dr. Mounir Bousbia-Salah and Dr. Mohamed Boughazi for their valuable feedback during the hardware design period. Thanks are also extended to Prof. Mouldi Bedda, Prof. H.A Abbassi, members of the Laboratory of Automatics and Signals at Annaba University for their advices in speech recognition model design.

References

[1] Beritelli F., Casale S., and Cavallaro A., "A Robust Voice Activity Detector for Wireless Communications Using Soft Computing," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 16, no. 9, pp. 1337-1339, 1998.

[2] Chen N., "A Vision-Guided Autonomous Vehicle: An Alternative Micromouse Competition," *IEEE Transactions*, vol. 40, no. 4, pp. 253-258, 1997.

[3] Chen N., Chung H., and Kwon Y., "Integration of Micromouse Project with Undergraduate Curriculum: A Large-Scale Student Participation Approach," *IEEE Transactions on Education,* vol. 38, no. 2, pp. 136-144, 1995.

[4] Clark L., "MIT Team Guides Airplane Remotely Using Spoken English," *Massachusetts Institute of Technology*, November, 2004.

[5] Fezari M., Bousbia-Salah M., and Bedda M., "Hybrid Technique to Enhance Voice Command System for a Wheelchair," *in Proceedings of the international Arab Conference on Information Technology (ACIT'05)*, pp. 139-144, 2005.

[6] Gu L. and Rose K., "Perceptual Harmonic Cepstral Coefficients for Speech Recognition in Noisy Environment," *International Conference Acoustics, Speech, and Signal Processing (ICASSP),* pp. 452-463, 2001.

[7] Hongyu L., Zhao Y., and Wang Z., "A Secure Voice Communication System Based on DSP," *in the IEEE 8th International Conference*, China, pp. 132-137, 2004.

[8] Kim J., Hyun S., Heung S., Myung J., and Vadakkepat P., "Cooperative Multi-Agent Robotic Systems: From the Robot-Soccer Perspective," *in Proceedings of Micro-Robot World Cup Soccer Tournament*, Koria, pp. 3-14, 1997.

[9] Kim W., Fezari M., and Bousbia M., "Development of a Voice Remote Control System," *in Proceedings of the 1998 Korea Automatic Control Conference*, Korea, pp. 1401-1404, 1998.

[10] Microship PIC16F876 from Microchip, http://www.microchip.com, 2001.

[11] Neiderjohn B., "An Experimental Investigation of the Perceptual Effects of Altering the Zero Crossing of Speech Signal", *IEEE Transaction on magnetic,* vol. 35, no. 12, pp. 618-625, 1987.

[12] Nishimoto T., Nobutoshi S., Tetsunori K., and Katsuhiko S., "Improving Human Interface in Drawing Tool Using Speech, Mouse and Key-Board," *in Proceedings of the 4th IEEE International Workshop on Robot and Human Communication*, Japan, pp. 107-112, 1993.

[13] Radiometrix Components, http://www. radiometrix.com, 2004.

[14] Rao R., Rose K. and Gersho A., "Deterministically Annealed Design of Speech Recognizers and its Performance on Isolated Letters", *IEEE International Conference Acoustics, Speech, and Signal Processing (ICASSP'98*), pp. 461-464, 1998.

[15] Rogers F., Van Aken P., and Cuperman V., "Time-Frequency Vector Quantization with Application to Isolated Word Recognition," *in Proceeding of IEEE International Symposium on Time-Frequency and Time-Scale Analysis*, pp. 173-176, 1996

[16] RSC-364 Manual, Sensory Company, http://www.voiceactivation.com/html/support/docs/80-0165-O.pdf, 2003.

[17] Wang T. and Cuperman V., "Robust Voicing Estimation with Dynamic Time Warping", *IEEE International Conference Acoustics, Speech, and Signal Processing (ICASSP'98)*, pp. 533-536, 1998.

Mohamed Fezari is a Lecturer in Electronics and computer architecture at the University of Badji Mokhtar, Annaba, Algeria. He got a Bachelor degree in electrical engineering from University of Oran, 1983. An MSc, degree in computer science from University of California, Riverside, 1987. He holds PhD degree in electronics from the University of Badji Mokhtar, Annaba, 2006.

Mounir Bousbia-Salah is a lecturer in electronics and computer architecture at the University of Badji Mokhtar Annaba, Algeria. He received an MSc degree in electrical engineering from Cardiff University, UK, 1987. He holds PhD, degree in electronics from the University of Badji Mokhtar, Annaba, 2004.