High Level Tele-Operating Speech Communication System for Controlling a Colony of Robots

Al-Dahoud Ali[1], Mohamed Fezari[2] and Attoui Hamza[2]

[1]Faculty of IT, Al-Zaytoonah University, Amman, Jordan
[2]Faculty of Engineering, Department of Electronics, Badji Mokhtar University, Annaba
Laboratory of Automatic and Signals, Annaba, BP.12, Annaba, 23000, ALGERIA
aldahoud@alzaytoonah.edu.jo, mohamed.fezari@uwe.ac.uk

Abstract—The purpose of this article is to presents the implementation and design of a high level communication interface based on multi agents model for a set of autonomous robots. Classical techniques, used in speech recognition which are zero crossing and extremes, dynamic time warping and different features such as Mel frequency Cepstral coefficients, delta MFCC , Energy and LPC are merged in order to increase the rate of recognition, followed by a decision system based on independent methods test results, are used as a speech recognition agent. Two consecutive agents; namely syntactic and semantic agents, are added to improve the recognition rate and improve the human-machine communication language. To implement the approach for tele-operating a set of robots on a real time, a Personal Computer interface was designed based on Bluetooth wireless communication modules to control the movement of a set of robots using high level language. The voice command system for four autonomous robots is designed however the robot navigation techniques are not discussed in this work. The main parts of the robots are based on a microcontroller from Microchip PIC18F2450, a set of sensors modules and a Bluetooth module.

Keywords: Speech recognition, HLL (High Level Language, Hybrid methods, human -robot interaction, DTW, Robots, Microcontroller, wireless communication.

1. Introduction

Speech recognition is a process by which the elements of spoken language can be recognized and analyzed, and the linguistic message it contains transposed into a meaningful form so that a machine can respond correctly to spoken commands. The earliest attempts to devise systems for automatic speech recognition by machine were made in the 1950s. Today, speech recognition research is interdisciplinary, drawing upon work in fields as diverse as biology, computer science, electrical engineering, linguistics, mathematics, physics, and psychology. Within these disciplines, pertinent work is being done in the areas of acoustics, artificial intelligence, computer algorithms, information theory, linear algebra, linear system theory, pattern recognition, phonetics, physiology, probability theory, signal processing, and syntactic theory[1, 21,22,23,24].
Applications of speech recognition have been made in office or business systems, as well as in manufacturing, medicine, and telecommunications, and usually concern the recognition and retrieval of information, such as voice-activated data entry; the control, operation, and monitoring of various machines and devices; call-processing functions; and the automation of services normally requiring human beings. While speech recognition has many short-term applications, it also has the potential to change daily life profoundly as free communication between man and machine becomes a reality.

Human-robot voice communication interface has a key role in many application fields [1-3]. Moreover, robots are becoming increasingly complex. A human-oriented approach to control them is the key for better interaction between the user and the robot. The most natural way to facilitate the user task s to provide a spontaneous speech during the interaction process as it is the natural way for human to communicate. Various studies made in the last few years have focused on systems based on spontaneous and natural speech, and market opportunities for speech-based devices are growing [4-7]. High accuracy in speech recognition abilities of the system and human-machine interaction based on better evaluation-corrective abilities and high level language rules are therefore the key for the success of this paradigm [8]. This paper proposes a new approach to the problem of the recognition of spotted words, using natural language recognition system composed of multi agents. Automatique speech recognition of spotted word agent based on a set of traditional pattern recognition approaches and a decision system based on test results of classical methods [2][5] and [7] , syntactic language agent and a semantic robot command agent in order to increase the accuracy of recognition. The increase in complexity as compared to the use of only traditional approach is considerable, however 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 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 natural language in form of phrase containing spotted words (nine commands words used in Arabic language). It has to be robust to any background noise confronted by the system.

The best-known strategies for speech recognition are the statistical and the connectionist ones, but fuzzy sets can also play an important role. Based on HMM's the statistical strategies have many advantages, among them being recalled: rich mathematical framework, powerful learning and decoding methods, good sequences handling capabilities, flexible topology for statistical phonology and syntax. The disadvantages lie in the poor discrimination between the models and in the unrealistic assumptions that must be made to construct the HMM's theory, namely the independence of the successive feature frames (input vectors) and the first order Markov process.

Based on artificial neural networks (ANNs), the connectionist strategies for speech recognition have the advantages of the massive parallelism, good adaptation, efficient algorithms for solving classification problems and intrinsic discriminative properties. However, the neural nets have difficulties in handling the temporal dependencies inherent in speech data. The learning capabilities of the statistical and the neural models are very important, classifier built on such bases having the possibility to recognize new, unknown patterns with the experience obtained by training. The introduction of fuzzy sets allows on one hand the so-called fuzzy decisions, on other hand the "fuzzyfication" of input data, often more suitable for recognition of pattern produced by human beings, by speaking, for example. In a fuzzy decision, the recognizer realizes the classification based on the degree of membership to a given class for the pattern to be classified, a pattern belonging in a certain measure to each of the possible classes[25].

This relaxation in decision leads to significant enhancements in recognition performances, situation that can also be obtained by looking in a fuzzy way to the input data The learning capabilities offered by the statistical and the connectionist paradigms and also a "nuanced" inside in the reality of the input and output domains of the speech recognizers contribute to a kind of "human likely" behaviour of this automata. These three main strategies were applied in our speech recognition experiments however the the developed algorithms were not all incorporate in this paper. The statistical strategies played the most important role in the development of continuous

speech recognition based on HMMs. The neural strategies are applied in form of multilayer perceptrons (MLP) and Kohonen maps (KM) for tasks like vowel or digit recognition[26-27].

Algorithms for hybrid structures like fuzzy HMM, fuzzy MLP or MLP –HMM are also incorporated in our research platform.

The aim of this paper is therefore the recognition of spotted words from a limited vocabulary taking into account the syntactic conditions then the semantic rules 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. As example, 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 [10-12]. Voice command needs the recognition of words from a limited vocabulary used in Automatic Guided Vehicle (AGV) system [13] and [14].

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 spotted 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, 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 then a syntactic language agent will check the correctness of the order, i.e. "robot black move right" is not acceptable as a command because there is not black robot, then a semantic language agent will test the possibility to understand the command i.e. "you green robot pick the pen" is not acceptable because in this application the robots task is just execute movement commands. Once the system recognise the robot and the order affected to that element it then sends to the USB port of the PC an appropriate binary code. This code is then transmitted to the robots via a Bluetooth wireless transmission protocol.

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.

3. HIGH LEVEL COMMUNICATION SYSTEM

As mentioned in the introduction this system is based on three independent agents used in speech recognition, the speech uttered by the operator is a high level language sentence as used by natural speaker, the system will detect the spotted words within this uttered phrase, it checks for syntactic correctness then checks the meaning of the operator, finally it generates a tele-operated command to the set of robots. The man components are illustrated in figure 1. The system is composed of the following agents:

A. 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 2.a 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.

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) [14].

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].

A.1 Dynamic Time Warping Algorithm (DTW)

Dynamic Time Warping algorithm (DTW) [30] is an algorithm that calculates an optimal warping path between two time series. The algorithm calculates both warping path values between the two series and the distance between them.

Suppose we have two numerical sequences (a_1,a_2, ..., a_n) and (b_1, b_2, ..., b_m). As we can see, the length of the two sequences can be different. The algorithm starts with local distances calculation between the elements of the two sequences using different types of distances. The most frequent used method for distance calculation is the absolute distance between the values of the two elements (Euclidian distance). That results in a matrix of distances having n lines and m columns of general term:

$$d_{ij=} \left| a_i - b_j \right| \; i=1..n \; and \; j=1..m$$

Starting with local distances matrix, then the minimal distance matrix between sequences is determined using a dynamic programming algorithm and the following optimization criterion:

$$a_{ij}=d_{ij} + min(a_{i-1,j-1},a_{i-1,j},a_{i,j-1}),$$

where a_{ij} is the minimal distance between the subsequences (a_1,a_2, ..., a_i) and (b_1, b_2, ..., b_j).

A warping path is a path through minimal distance matrix from a_{11} element to a_{nm} element consisting of those a_{ij} elements that have formed the a_{nm} distance.

The global warp cost of the two sequences is defined as shown below:

$$GC = 1/P \sum_{i=1}^{P} Wi$$

where w_i are those elements that belong to warping path, and p is the number of them . There are three conditions imposed on DTW algorithm that ensure them a quick convergence:

- monotony – the path never returns, that means that both indices i and j used for crossing through sequences never decrease.
- continuity – the path advances gradually, step by step; indices i and j increase by maximum 1 unit on a step.
- boundary –the path starts in left-down corner and ends in right-up corner.

Because optimal principle in dynamic programming is applied using “backward” technique, identifying the warp path uses a certain type of dynamic structure called “stack”. Like any dynamic programming algorithm, the DTW one has a polynomial complexity. When sequences have a very large number of elements, at least two inconveniences show up:

- memorizing large matrices of numbers;
- performing large numbers of distances calculations.

There is an improvement in standard DTW algorithm that sorts out the two problems named above: FastDTW (Fast Dynamic Time Warping) [31].

Words identification can be performed by straight comparison of the numeric forms of the signals or by signals spectrogram comparison.

The comparison process in both cases must compensate for both the different length of the sequences and non-linear nature of the sound. The DTW Algorithm succeeds

in sorting out these problems by finding the warp path corresponding to the optimal distances between two series of different lengths.
Figure 1.a illustrates the DTW comparison between reference word and test word:

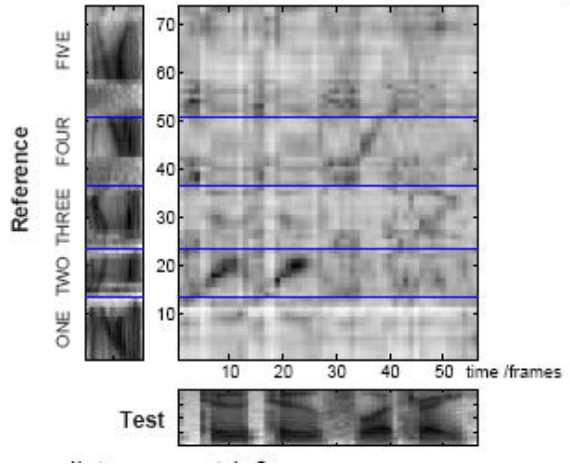
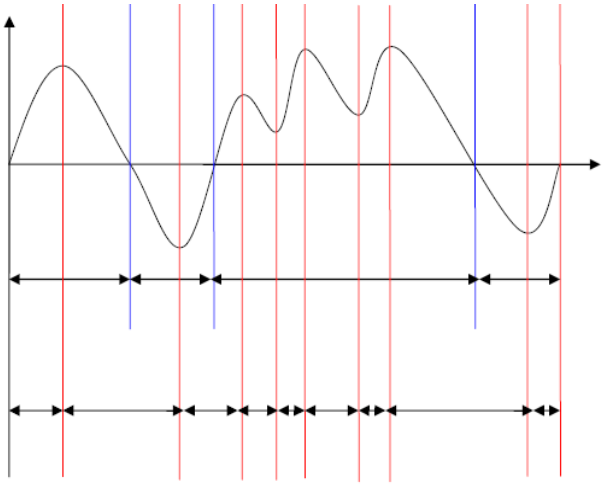


Figure 1.a Framewise comparison with stored templates

A.2 Energy Zero Crossing and Exremums features

Zero crossing and energy were used to detect the beginning and the end of a word however they were used also as parameters in our case because of fast computation of these parameters and then easy to implement them in microcontroller or DSP processor.



(y(k)*y(k+1) < 0) and Time between two CZ >200 ms
[y(k+1)-y(k)]*[y(k+1)-y(k+2)]) >0

A.3 LPC and MFCC features in sppech recognition

The combined use of LPC and MFCC cepstrals in speech recognition system is for calculating speech features.
Calculation of the speech features algorithm is defined in the following form.

- *Pre-processing.* The amplitude spectrum of a speech signal is dominant at "low frequencies" (up to approximately 4 *kHz*). The speech signals is passed through a first-order FIR high pass filter:

$$s_p(n) = s_{in}(n) - \alpha \cdot s_{in}(n-1)$$

where α – is the filter coefficient $(\alpha \in (0{,}95;1))$, $s_{in}(n)$ (is the input signal.

- *2. Voice activation detection (VAD).*

The problem of locating the endpoints of an utterance in a speech signal is a major problem for the speech recognizer. An inaccurate endpoint detection will decrease the performance of the speech recognizer. Some commonly used measurements for finding speech are short-term energy estimate E_s , or shortterm power estimate P_s , and short term zero crossing rate Z_s .

For the speech signals $s(n)_p$ these measures are calculated as follows:

$$E_s(m) = \sum_{n=m-L+1}^{m} s_p^2(n), \qquad P_s(m) = \frac{1}{L}\sum_{n=m-L+1}^{m} s_p^2(n),$$

$$Z_s(m) = \frac{1}{L}\sum_{n=m-L+1}^{m} \frac{\left|\operatorname{sgn}(s_p(n)) - \operatorname{sgn}(s_p(n-1))\right|}{2}$$

where

$$\operatorname{sgn}(s_p(n)) = \begin{cases} 1, & s_p(n) \geq 0, \\ -1, & s_p(n) < 0. \end{cases}$$

B. *Learning strategies*

In the same way into which a human learns to perceive speech from examples by listening, the automatic speech recognizer does apply a learning strategy in order to decode the pronounced word sequence from a sequence of elementary speech units like phonemes, with or without context. By learning are created models for each elementary speech unit, serving in the comparisons to make a decision about the uttered speech unit. We will describe further the learning strategies implemented in our research platform: hidden Markov models (HMM), and hybrid methode based on grouping more parameters or Kohonen maps (KM). Some hybrid learning strategies are also implemented in form of a HMM-ANN combination and a fuzzy perceptron or fuzzy HMM. **HMMs** are finite automata, with a given number of states; passing from one state to another
is made instantaneously at equally spaced time moments. At every pass from one state to another, the system generates observations, two processes taking place: the transparent one represented by the observations string (features sequence), and the hidden one, which cannot be observed, represented by the state string [28].
In speech recognition, the left - right model (or the Bakis model) is considered the best choice. For each symbol, such a model is constructed; a word string is obtained by connecting corresponding HMMs together in sequence (Huang et al., 2001). For limited vocabulary, word models are widely used, since they are accurate and trainable. In the situation of a specific and limited task they become

valid if enough training data are available, but they are typically not generalizable. Usually for not very limited tasks are preferred phonetic models based on monophones (which are phonemes without context), because the phonemes are easy generalizable and of course also trainable. Monophones constitute the foundation of any training method and we also started with them (for any language). But in real speech the words are not simple strings of independent phonemes; these phonemes are affected for the immediately neighboring phonemes by coarticulation.

This monophone models are changed now with triphone models (which are phonemes with context) that became actually the state of the art in automatic speech recognition for the large vocabularies (Young, 1992). A triphone model is a model that takes into consideration the left and the right context of the phonemes. Based on the SAMPA (Speech Assessment Methods Phonetic Alphabet) in Romanian language there are 34 phonemes; the number of necessary models (triphones) to be trained is about 40000, situation which is unacceptable.

In the continuous speech recognition task we modelled only internal – word triphones and we adopted the state tying procedure, conducting to a controllable situation. If triphones are used in place of monophones, the number of needed models increases and it may occur the problem of insufficient training data. To solve this problem, tying of acoustically similar states of the models built for triphones corresponding to each context is an efficient solution [28-29]

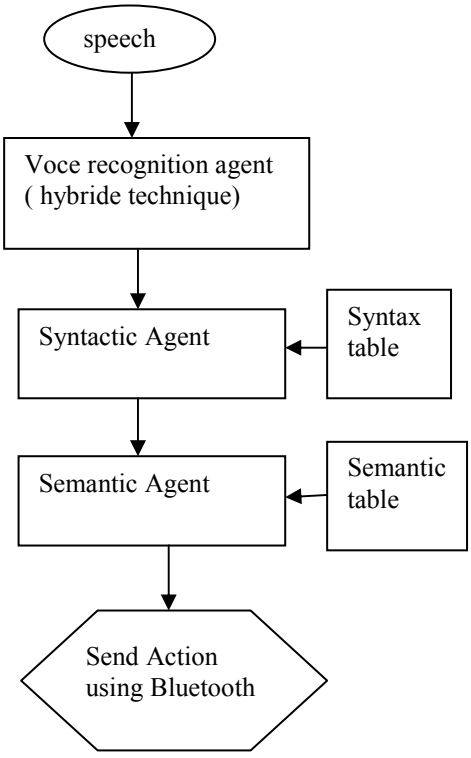


Fig.1.b High level communication system components

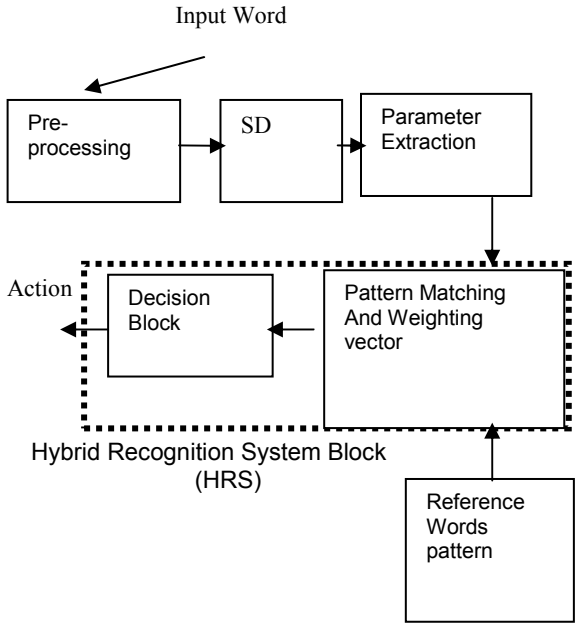


Fig. 2: Block Diagram of Voice rec. Agent

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 2 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 [9]

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.

C. Syntactic agent

Based on some syntactic language rules created and saved in a dictionary the agent checks for the order of spotted words within the uttered phrase. The syntax for each sentence should follow the natural language rule i.e. "key-word + Robot-name + action". The key-word is necessary as a security so that the system would not replay to any generated phrase that might contain the spotted word and spoken by another person than the operator. If there is any syntactic error or non defined syntax in dictionary then the agent will not provide any action to the following agent. As example, if the operator says "you red go forward", in this case the key-word "robot" has not been detected n the sentence therefore the command is refused.

The rules are:

Phrase= key-word + nominal-group + verb + complement.

Nominal-group= determinant + name [+ preposition+ nominal-group].

D. Semantic agent

The uttered phrase should have a meaning, because we can construct correct phrases using the spotted words however they have no meaning and therefore the system can detect errors and hence it will not generate commands (i.e. "robot eats red carrots" or "robot go forward next to red table", or "robot blue go to the table in its right"). This luck of precision leads the agent to not classify the sentence as a correct one. This type error is due in general to some phonemes of different successive words in the phrase. The semantic agent has a rule semantic table, in which accepted set of words are stored in the table called semantic table.

4. HARDWARE PARTS IN THE APPLICATION

A. Transmission module

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". On USB port a Bluetooth wireless transmission module is added in order to have a Tele-Operated system as shown in Figure 2.b.

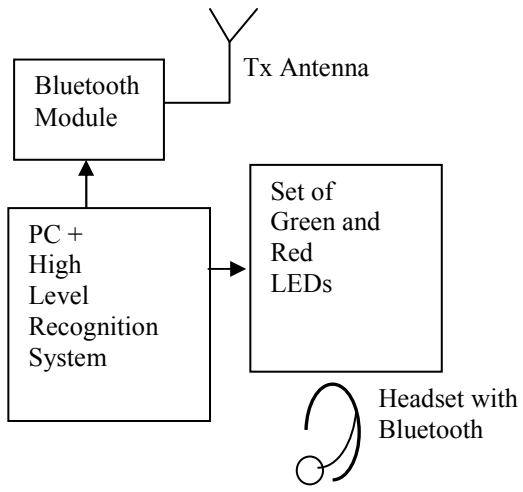


Figure 2.b : The Voice Command system

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 colours ("Ahmar", "azrak", "Akhdar", and "Asuad") and their meanings are listed as in Table 1.

TABLE 1
The meaning of voice commands

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

B. Autonomous robots and reception module

As in Figure 3 and 4, the structures of the mechanical hardware and the computer board of autonomous robots in this paper are similar to those in [11-12]. However, since the autonomous robots in this paper need to perform simpler tasks than those in [11-12] do, these autonomous robots can more easily be constructed. The computer board of each autonomous robot consists of a PIC18F2450, with 16 to 32K-instruction EEPROM (Electrically Programmable Read Only Memory), 256 byte of Ram, ADC converters and integrated full speed USB module [15], two H bridges drivers using BD134 and BD133 transistors for DC motors, a Bluetooth radio transmitter, 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.

The Bluetooth specification defines three power classes for radio transmitters with an output power of 1 mW, 2.5 mW and 100 mW. The output power defines the range that the device is able to cover and thus the functionality of your product must be considered when deciding which power class to use. The user would not want to have to get up from his desk to connect to the LAN and therefore requires a higher power radio. Conversely, a cellular phone headset is likely to be kept close to the phone, making a lower range acceptable, which allows smaller batteries and a more compact design. Table 2 details the respective maximum output power versus range. It is important to realize that the range figures are for typical use. In the middle of a stadium, where the land is flat and there is not much interference, a Class 1 device has been

successfully tested at over 250 meters. But in a crowded office with many metal desks and a lot of people, the Bluetooth signal will be blocked and absorbed, so propagation conditions are far worse and ranges will be reduced.

B.1 SFR08 Module Sensors

In order to avoid and maintain a safe distance from obstacles, two modules (type SFR08 provided by Lextronic) of ultrasonic sensors are installed in front of the robot.

The reasons, ultrasonic sensors were chosen are: these sensors are readily commercially available as a module, to be integrated in the robot base, and have a reasonable cost, and they complied with the requirement specification for Robot movements. With the aid of its ultrasonic sensors, each robot is able to keep track of the distance between itself and other robots and to avoid collision with obstacles such as walls.

Most distance-measuring ultrasonic systems are based on the time-of-flight method. This method comprises:

1. Transmitting an ultrasonic pulse.
2. Radiating ultrasonic pulses over a certain range.
3. A receiver receiving the ultrasonic pulses.
4. Calculating the time between the transmission and the reception of the ultrasonic pulse, where the distance (d) to the object having reflected the ultrasonic pulse can be calculated as: $d=v*t/2$.

The time measured can easily be transformed into distance. The ultrasonic signal processing module used in the design is the 'SFR08' Provided by LEXTRONIC [33].

C. Bluetooth protocol

Each Bluetooth device has its own unique 48 bit IEEE MAC Bluetooth address (BD_ADDR), which identifies it to other devices; if the device is a master; the connection timing and the hopping sequence are also derived from this address. Addresses are obtainable from the SIG in blocks and need to be programmed into every Bluetooth product at manufacture—all silicon is shipped with the same default address that must be changed. A "friendly name" may also be programmed into your product either by the user or at manufacture to enable the MMI to connect to "CSR development module," "Daisy's phone," "Lara's headset," or "Amy's little black book," concealing the actual address. The address is concealed from the user because it is a string of numbers (typically expressed in hexadecimal) which is not a very user-friendly format. An example of a Bluetooth device address is 0x0002 5bff 1234. The Bluetooth system operates in the 2.4GHz band. This band is known as the Industrial Scientific and Medical (ISM) band. In the majority of countries around the world, this band is available from 2.40–2.4835GHz and thus allows the Bluetooth system to be global, however many other technologies also reside in the band:

- 802.11b
- Home RF
- Some Digital Enhanced Cordless Communications (DECT) variants
- Some handheld short-range two-way radio sets (walkie-talkies).

More details in [17][18] and [19].

TABLE 2
Bluetooth Radio Power Classes

Power Class	*Max Output*	*Power Range*
Class 1	100 mW	Up to 200 meters
Class 2	2.5 mW	20 meters
Class 3	1 mW	1 meter

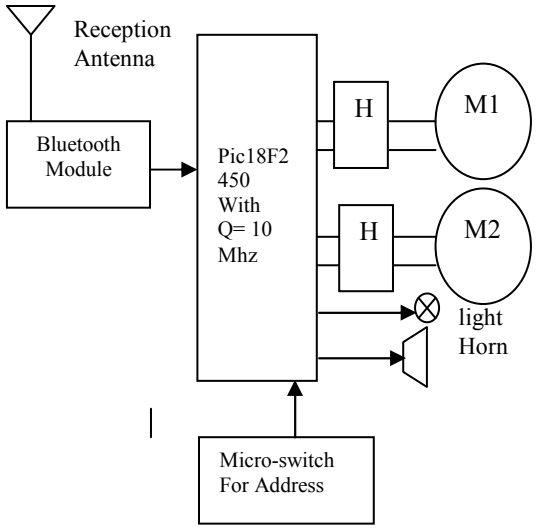


Fig.3. Autonomous robot block diagram

Fig. 4. Overview of one Robot and Bluetooth modules

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 to the Bluetooth module.

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:

1) The proposed system, to command and control an autonomous robot by human voices, is based on hardware components such as microcontrollers and parallel PCs interface and Bluetooth modules.

2) Compared with previous projects that build intelligent systems as in [10-13], the cost of the proposed voice command system is lower and modules combination design.

3) An autonomous robot controlled by high level communication speech 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 automatic speech recognition system, other member will work natural language processing and the other person an autonomous robot, while a fourth person may work on behaviour of several robots.

4) Several interesting competitions of voice-controlled autonomous robots will be possible. One example of the competitions is a robot soccer tournament by human voice commands.

5) While previous intelligent systems as in [10-12] 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 transmission signal by wireless communication devices then reaction of the robot, the effect of time delays in controlling autonomous robots can be observed.

6. TESTS ON THE VOICE COMMAND SYSTEM

The developed system has been tested within the laboratory of L.A.S.A. 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 syntactic and semantic agents on the accuracy of recognition commands.

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

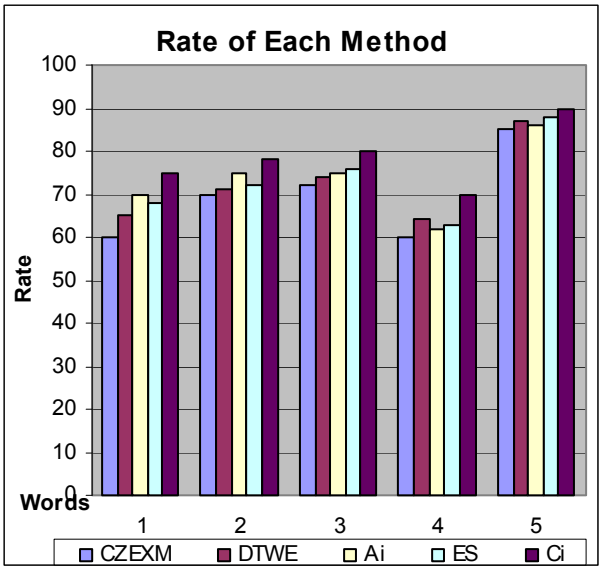
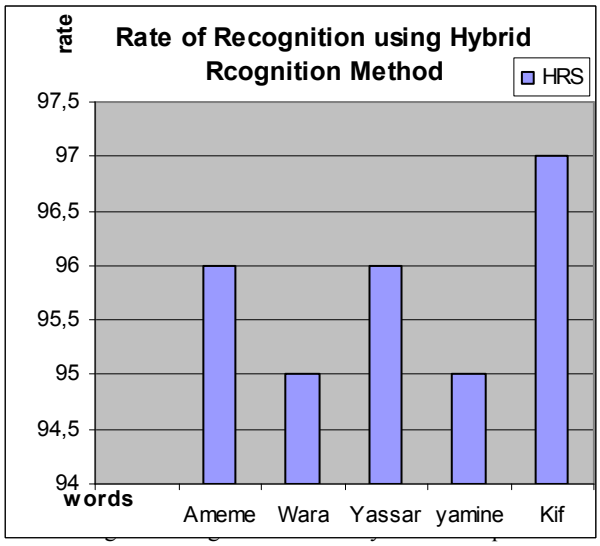


Fig. 5.a Recognition Rate of each method.



To test the effect of semantic and syntactic agents, about 100 of phrases as utterances were tested taking into account the rules and conditions to recognize a valid command for the set of robots, the results shows that these two agents can improve de rate of recognition and hence reduce the errors.

Syntactic agent improved the recognition by 10% and semantic agent reduced the errors due to non meaning phrases about 15%.

7. CONCLUSION AND FUTURE WORK

A Tele-Operated voice command system for autonomous robots is proposed and is designed based on an ASR for spotted 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 words are quite different in phonemes.

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. Inserting new agents that take care of syntactic and semantic errors has upgraded the recognition rate. Tests carried out have shown an improvement in performance, in terms of misclassification of the words pronounced by the user and incorrect phrases. The increase in computational complexity as compared with a traditional approach is, however, negligible. Segmentation of the word in three principal frames for the Zero Crossing and Extremes method gives better results in recognition rate.

Since the designed robots consists of a microcontroller, and other low-cost components namely Bluetooth as 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. 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 [9] and [20].

REFERENCES

[1] S. Furui, "Recent advances in spontaneous speech recognition and understanding", in Proc. IEEE-SCA Workshop on Spontaneous Speech Processing and Recognition (SSPR) pages; 1-6, 2003.

[2] L.J. Clark, "MIT team guides airplane remotely using spoken English", News Office journal, Massachusetts Institute of technology, November, 2nd 2004.

[3] Rao R.S., Rose K. and Gersho A., "Deterministically Annealed Design of Speech Recognizers and Its Performance on Isolated Letters," Proceedings IEEE ICASSP'98, pp. 461-464, May 1998.

[4] Wang T. and Cuperman V., "Robust Voicing Estimation with Dynamic Time Warping," Proceedings IEEE ICASSP'98, pp. 533-536, May 1998.

[5] B.M. Neiderjohn, "An Experimental Investigation of the Perceptual effects of Altering the Zero Crossing of Speech Signal", IEEE transaction, Acoustic, Speech and signal Processing, vol. ASSP-35, pp. 618-625, 1987.

[6] Hazem M. El-Bakry, N. Mastorakis "Fast Word Detection in a Speech Using New High Speed Time Delay NeuralNetworks" WSEAS WSEAS TRANSACTIONS on SIGNAL PROCESSING, Issue 7, Volume 5,pp. 261-271, July 2009

[7] Furui S., "Overview of the 21st Century COE program framework for systematique and application of large-scale knowledge resources", in Proc. Int. Symp. On Large-Scale Knowledge Resources,pp. 1-8, 2004.

[8] Quartieri J., Troisi A., Guarnaccia C., Lenza TLL, D'Agostino P., D'Ambrosio S., Iannone G. *Analysis of Noise Emissions by Trains in Proximity of a Railway Station*, Submitted to 10th WSEAS Int. Conf. on "Acoustics and Music: Theory & Applications", Prague, Czech Republic.

[9] Fezari M., M. Bousbia-Salah and M. Bedda, "Hybrid technique to enhance voice command system for a wheelchair", ACIT'05, Al_Isra University, Jordan, 2005.

[10] K. Wada, T. Shibata, T. Saito, K. Sakamoto, and K. Tanie. Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged. In 2005 IEEE Int. Conf. on Robotics and Automation, pages 2796–2801, 2005.

[11] Kim J. H., et al, "Cooperative Multi-Agent robotic systems: From the Robot-Soccer perspective", 1997 Micro-Robot World Cup Soccer Tournament Proceedings, Taejon, Korea, pp. 3-14, 1997.

[12] S. Caselli, E. Fantini, F. Monica, P. Occhi, and M. Reggiani. Toward a mobile manipulator service robot for human assistance. In 1st Robocare Workshop, 2003.

[13] Ussama Javed Rai, Abbas Dehghni, *Design & Development of an Automated (Robotic) Snapping, Banding & Sorting System*, Proceedings of 8th WSEAS on Signal Processing, Robotics and Automation, pag. 309-314, ISSN 1790-5117, Cambridge, 2009

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

[15] Data sheet PIC18F2450 from Microchip inc. User's Manual, 2005, http://www.microchip.com.

[16] Kim W. J., et al., "Development of A voice remote control system." Proceedings of the 1998 Korea Automatic Control Conference, Pusan, Korea, pp. 1401-1404, 1998.

[17] B. A. Miller and Ch. Bisdikian, *Bluetooth Revealed: The Insider's Guide to an Open Specification for Global Wireless Communications*, Prentice Hall PTR, 2002.

[18] D. Kammer, G. McNutt and B. Senese, "Bluetooth Application Developer's Guide", book printed by Syngress Publishing, Inc. 2002, ISBN: 1-928994-42-3

[19] Müller, T., ed., "Bluetooth Security Architecture," White Paper Revision 1.0, Bluetooth Special Interest Group, July 1999.

[20] M. Fezari, Attoui Hamza, Mouldi BEDDA "Arabic Spotted Words Recognition System Based on HMM Approach to control a didactic Manipulator Arm", In Proc. MS'08, nt. Conf. On Modelling and Simulation, PETRA/ Jordan, Vol. 2008

[21] A. Kaddouci, H. Zgaya, S. Hammadi, F. Bretaudeau "Multi-Agents Based Protocols for Negotiation in a Crisis Management Supply Chain", Proceedings of the 8th WSEAS International Conference on COMPUTATIONAL INTELLIGENCE, MAN-MACHINE SYSTEMS and CYBERNETICS (CIMMACS '09), Puerto De La Cruz, Tenerife, Canary Islands, Spain, December 14-16, 2009, p: 143.

[22] Roxana Grejdanescu, Loredana Paun, Valeriu Avramescu, Eugen Strajescu, "Aspects Regarding the Motion Possibilities of a CNC Multifunctional Machine-Tool ", Proceedings of the 8th WSEAS International Conference on COMPUTATIONAL INTELLIGENCE, MAN-MACHINE SYSTEMS and CYBERNETICS (CIMMACS '09), Puerto De La Cruz, Tenerife, Canary Islands, Spain, December 14-16, 2009, P: 155

[23] P. Papantoni-Kazakos, A. T. Burrell, " Stable Protocols for the Medium Access Control in Wireless Networks", Proceedings of the 8th WSEAS International Conference on DATA NETWORKS, COMMUNICATIONS, COMPUTERS (DNCOCO '09), Morgan State University, Baltimore, USA, November 7-9, 2009, P: 19.

[24] Ching-Chiang Chen, Dong-Her Shih, Cheng-Jung Lee, "Web 2.0 Trends based on E-Learning for Troops Training Process Improvement (TTPi)", Proceedings of the 8th WSEAS International Conference on DATA NETWORKS, COMMUNICATIONS, COMPUTERS (DNCOCO '09), Morgan State University, Baltimore, USA, November 7-9, 2009, P: 240.

[25] C. O. Dumitru and I. Gavat "Progress in Speech Recognition for Romanian Language", Source: Advances in Robotics, Automation and Control, Book edited by: Jesús Arámburo and Antonio Ramírez Treviño, ISBN 78-953-7619-16-9, pp. 472, October 2008, I-Tech, Vienna, Austria

[26] Draganescu, M., (2003). "Spoken language Technology", *Proceedings of Speech Technology and Human-Computer-Dialog (SPED2003)*, pp. 11-12, Bucharest, Romania.

[27] Dumitru, C.O., Gavat, I. (2006). "A Comparative Study of Features Extraction Methods Applied for Continuous Speech Recognition in Romanian Language", Proceedings the 48th International Symposium ELMAR 2006, pp. 115-118, Zadar, Croatia.

[28] Gavat, I., Dumitru, C.O., Costache, G., Militaru, D. (2003). Continuous Speech Recognition Based on Statistical Methods, Proceedings of Speech Technology and Human-Computer-Dialog (SPED2003), pp. 115-126, Bucharest.

[29] Furtună, F., Dârdală, M., Using Discriminant Analisys in Speech Recognition, The Proceedings Of The Fourth National Conference Humman Computer Interaction Rochi 2007, Universitatea Ovidius Constanţa, 2007, MatrixRom, Bucharest, 2007

[30] Sakoe, H. & S. Chiba. (1978) Dynamic programming algorithm optimization for spoken word recognition. IEEE, Trans. Acoustics, Speech, and Signal Proc., Vol. ASSP-26.

[31] Stan Salvador, Chan, FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space, IEEE Transactions on Biomedical. Engineering, vol. 43, no. 4

[32] K.R.Ayda-zade, S.S.Rustamov. Research of Cepstral Coefficients for Azerbaijan speech recognition system. Transactions of Azerbaijan National Academy of sciences."Informatics and control problems". Volume XXV, №3. Baku, 2005, p.89-94.

[33] Lextronic, datasheet, SFR05/SFR08, Ultrasonic sensors module, 2005. http://www.lextronic.com