# Human Machine Communication Interface System Based on Merging Best Features and Semantic Models

**Mohamed Fezari**[1], **Ahmad Al-Dahoud**[2]
1-Laboratory of Automatic and Signals Annaba,Faculty of Engineering, Department of Electronics,
Badji Mokhtar Annaba University, Po Box:12 Annaba, 23000 Algeria
2-Faculty of IT, JUST University, Jordan
Mohamed.fezari@uwe.ac.uk, black4online@yahoo.com

## Abstract

This work is part of a research project to develop a human machine communication interface based on multi agents model for a set of autonomous robots. Classical techniques, used in Automatic Speech Recognition (ASR) are: pre-processing, features extraction, models creation and then classifiers. As features in ASR, we mention: zero crossing and extremes, 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, dynamic time warping 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 main parts of the robots are based on a microcontroller from Microchip PIC18F2450 and a Bluetooth module.

***Keywords:*** Speech recognition; Hybrid methods; human -robot interaction; DTW; Semantic rules; wireless communication.

## 1. INTRODUCTION

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 natural speech and gesture, and market opportunities for speech-based devices are growing [4-7]. This paper proposes a new approach to the problem of the recognition of spotted words, using natural language recognition system composed of multi agents. Automatic 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 as done in [14].

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.

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 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.

As application, a voice 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] .

## 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:

### *3.1 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 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) .

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].

## 3.2 Dynamic Time Warping Algorithm(DTW)

Dynamic Time Warping algorithm (DTW) [6] 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.

In case 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:

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

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

$$a_{ij}=d_{ij} + min(a_{i-1,j-1},a_{i-1,j},a_{i,j-1}), \qquad (2)$$

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:

$$GC = 1/P\sum_{i=1}^{P} Wi \qquad (3)$$

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 and performing large numbers of distances calculations.

Words identification can be performed by straight comparison of the numeric forms of the wav 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:

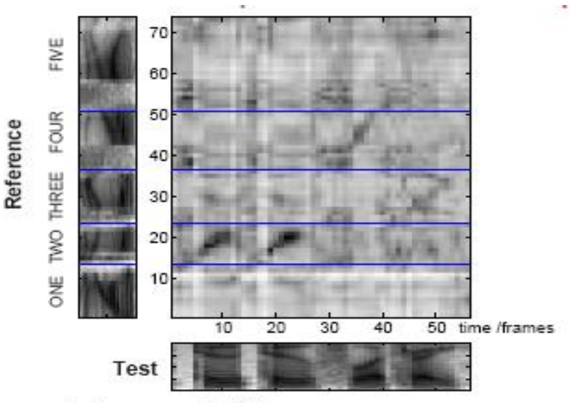

Figure 1.a Framewise comparison with stored templates

### 3.3 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. 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.

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.

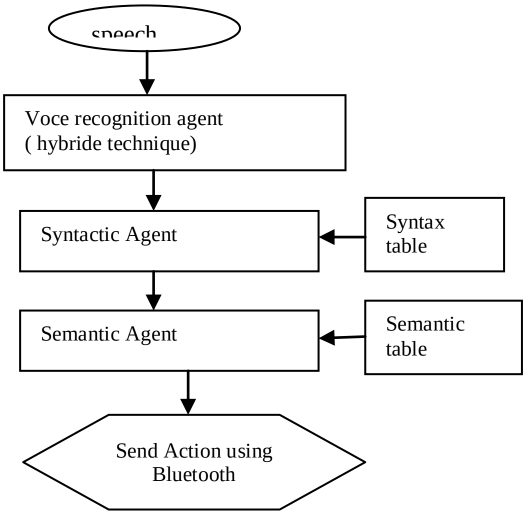

Fig.1.b High level communication system components

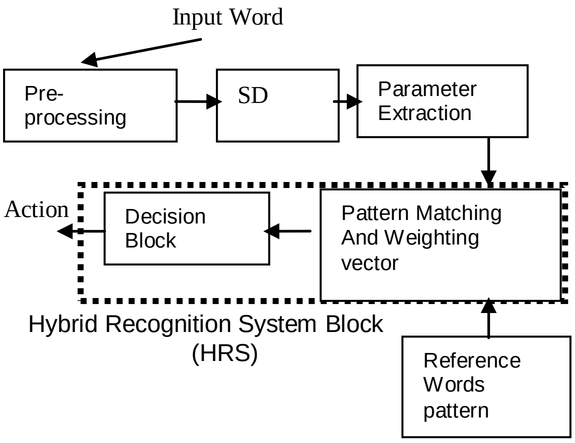

Fig. 2: Block Diagram of Voice rec. Agent

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.

### 3.4. 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].

### 3.5. 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.

Table1: Semantic Table in our Case.

| Keyword | Noun | Adjectif | Action | Execute |
|---|---|---|---|---|
| Robot<br>"LASA" | Ahmar<br>Akhdar<br>Azrak<br>Asuad | -<br>Fast<br>slow | Amame<br>Wara<br>Yamine<br>Yasar<br>KIF | Tabek |

## 4 Tests and Results

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.3.a and Fig.3.b.

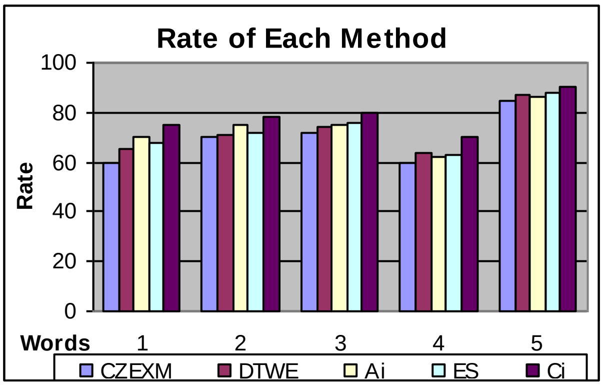

Fig. 3.a Recognition Rate of each method.

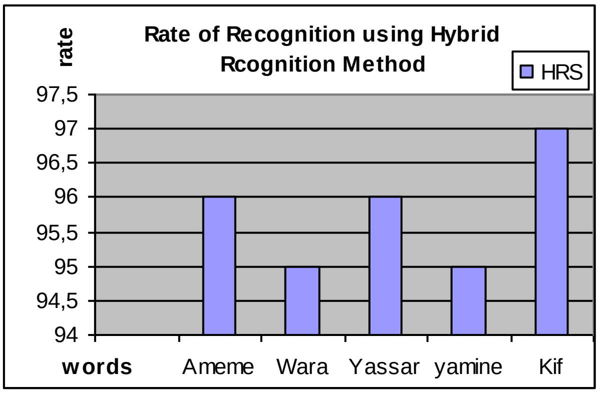

Fig 3.b Recognition Rate of hybrid technique.

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%.

## 6. 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.

## 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] R.S Rao., 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] T. Wang and 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] M. Hazem, N. 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] S Furui., “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] J. Quartieri , A. Troisi A., C. Guarnaccia , 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] M. Fezari , 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] J. Kim 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] U. 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] 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