

# Implementation of a Biometric Interface in Voice Controlled Wheelchairs

Lamia Bouafif<sup>1</sup> and Nouredine Ellouze<sup>2,\*</sup>

<sup>1</sup>National Institute of Biomedical Studies of Tunis, 1092, Tunis, Tunisia

<sup>2</sup>Image and Signal Processing Laboratory, ENIT BP 37, University of Tunis El Manar, 1064, Tunisia

\*Corresponding Author: Nouredine Ellouze. Email: b2lamia@yahoo.fr

Received: 22 September 2019; Accepted: 26 December 2019

**Abstract:** In order to assist physically handicapped persons in their movements, we developed an embedded isolated word speech recognition system (ASR) applied to voice control of smart wheelchairs. However, in spite of the existence in the industrial market of several kinds of electric wheelchairs, the problem remains the need to manually control this device by hand *via* joystick; which limits their use especially by people with severe disabilities. Thus, a significant number of disabled people cannot use a standard electric wheelchair or drive it with difficulty. The proposed solution is to use the voice to control and drive the wheelchair instead of classical joysticks. The intelligent chair is equipped with an obstacle detection system consisting of ultrasonic sensors, a moving navigation algorithm and a speech acquisition and recognition module for voice control embedded in a DSP card. The ASR architecture consists of two main modules. The first one is the speech parameterization module (features extraction) and the second module is the classifier which identifies the speech and generates the control word to motors power unit. The training and recognition phases are based on Hidden Markov Models (HMM), K-means, Baum-Welch and Viterbi algorithms. The database consists of 39 isolated speaker words (13 words pronounced 3 times under different environments and conditions). The simulations are tested under Matlab environment and the real-time implementation is performed by C language with code composer studio embedded in a TMS 320 C6416 DSP kit. The results and experiments obtained gave promising recognition ratio and accuracy around 99% in clean environment. However, the system accuracy decreases considerably in noisy environments, especially for SNR values below 5 dB (in street: 78%, in factory: 52%).

**Keywords:** Handicapped persons; smart wheelchairs; voice control; speech recognition; training and classification

## 1 Introduction

To enable people with disabilities to move independently, researchers have been interested in improving the human machine interface based on biometric commands (voice, face) instead of manual control. The intelligence of a wheelchair lies in its ability to perceive the environment through different types of sensors and also thanks to its interaction with the user. Numerous prototypes of intelligent chairs have



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

been developed in research laboratories and industry using mobile robotics and embedded systems technologies. Unfortunately, these prototypes use joystick and manual controls. Currently, Automatic Speech Recognition systems (ASR) are widely used in Human-machine interfacing, mobile telephony and access control. Another aspect of speech recognition can be used by ASR, which is the assistance of people with moving disabilities [1]. In fact, with their voices, patients could control and move a wheelchair or turn on-off the light switch. Tab. 1 shows some examples of ASR applications demonstrating the benefits of automatic speech recognition. The first isolated word ASR system was developed since 1922 by Elmwood Button Company which produced a toy named “Radio Rex”. It was stuffed dog that jumps when someone says its name. Later in 1972, Glenn et al. proposed the first commercial device called “Voice Command system” [2]. This autonomous system requires a training phase and is capable of recognizing twenty-four isolated words.

**Table 1:** Examples of ASR applications

Domain	Applications
Communication	Automation of telephone transactions, E-banking.
Automobile	Free hands control equipment such as radio, navigation system, voice dialing.
Medical	Help for people with disabilities, rehabilitation, microscopic surgery.
Industry	Voice control machines.
Informatics	Internet browsing, voice navigator, intelligent keyboard, vocal dictation.
Security	Control access, bank transactions.

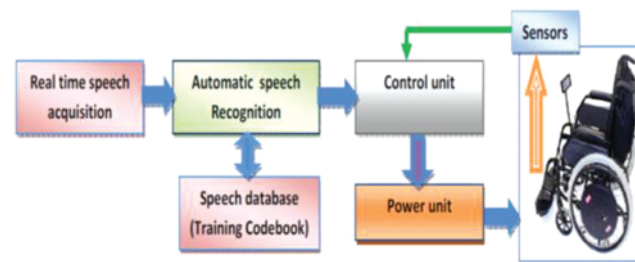
The same year, DARPA (Defense Advanced Research Projects Agency) funded the research program to develop a computer system capable to understanding the continuous speech. The most successful project was the HARPIE Carnegie Mellon which was able to recognize continuous speech with a vocabulary of 1011 words with 95% of accuracy [3]. Finally in 2010, Microsoft added a voice mail to text into his exchange mail server, so we can get the voice mail transcribed into Outlook. In 2011, Google added online its “personalized recognition” voice search engine on android phones and in chrome browser [4]. The main object of this work consists to develop a smart control aid system by replacing the wheelchair joystick control by a voice command. This application enables people with disabilities to facilitate travel and to ensure its autonomy. It provides all the necessary safety during the movement. It can be combined with mobile telephony to track the traceability and even measure the patient’s physiological parameters and transmit them to the medical staff (Tab. 2).

**Table 2:** ASR industrial products

System	Method	Recognition	Application	Vocabulary
DRAGON	HMM	Continuous mono speaker	Speech recoognition	5000 to 60000
SIEMENS	RN	Multi speaker	Telephony	10000
IBM	HMM	Multi speaker isolated words	Voice dictation	35000 to 65000

## 2 Methods and Materials

The aim of this project is to help people with disabilities to move easily by wheelchairs using their voice as a control tool. The interface of Fig. 1 is composed of an embedded system allowing a real time speech acquisition of the control word (on, off, stop, left, right and back). Once recognized, the control word is

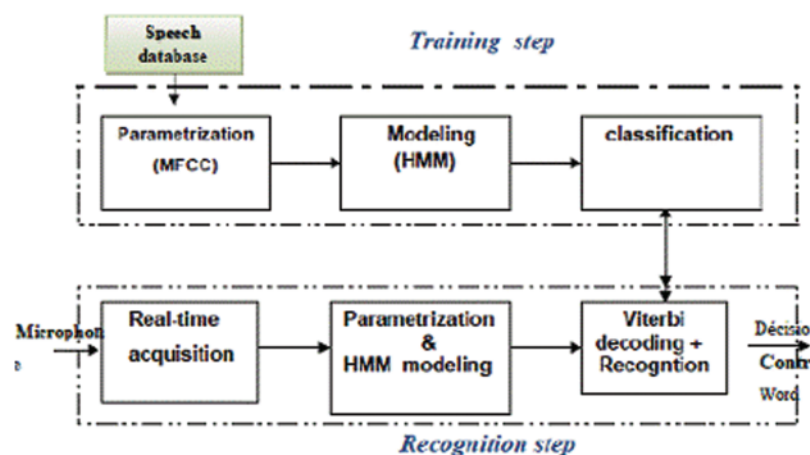


**Figure 1:** The architecture of the wheelchair control

used to operate the motors of the wheelchair thanks to an isolated word recognition system and a power unit (Fig. 1). In addition, the chair is equipped with sensors that collect data on its environment to detect obstacles and ensure the user safety.

### 2.1 Speech Acquisition, Training and Recognition

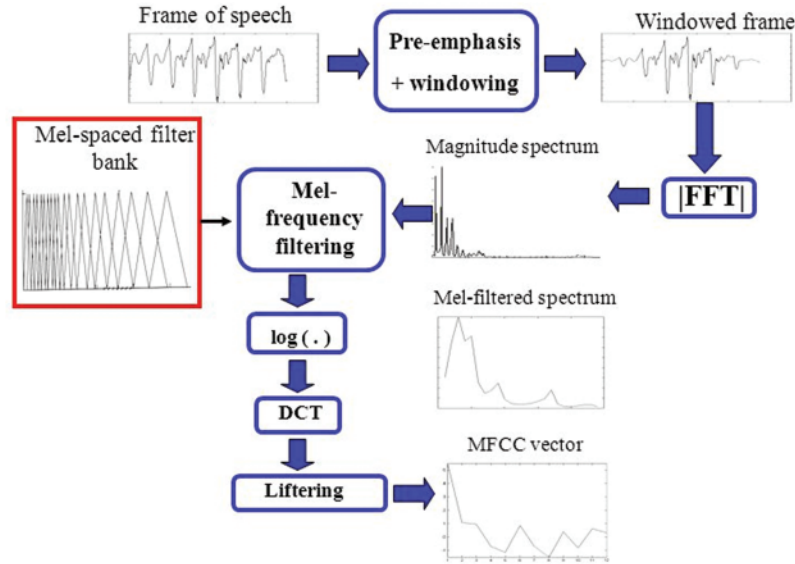
Most of the movement commands of the wheelchair are executed by the user through his voice. For safety reasons, an obstacle detection system based on ultrasonic sensors which are placed around the chair was inserted in order to have a satisfactory detection radius. These sensors are connected at the inputs of the DSP card as illustrated in the experimental section. To avoid collision, we have programmed short distances of detection causing the chair to stop or allowing it to slow when needed. When an obstacle is detected in the trajectory (front or rear) of the wheelchair at a distance less than 50 cm, the wheelchair must stop. Fig. 2 illustrates the training and recognition steps of our system.



**Figure 2:** Block diagram of a recognition system

### 2.2 Parameterization

The speech signal is divided into frames of 30 ms. For each frame, a set of coefficients that represent the speech or the word is extracted. A segmental analysis is performed to provide observation vectors  $Y_t$ , which will be used for HMM training. The parameterization method is the Melcepstral Coefficients (MFCC) which is described in Fig. 3. The computing of the MFCC coefficients follows next steps:



**Figure 3:** Feature extraction with MFCC

-Pre-emphasis with a high pass filter:

$$H(z) = 1 - 0.95z^{-1} \quad (1)$$

-Segmentation with hamming windows of 10 ms with recovering ratio of 50%.

-Computing FFT for every window.

-Mel scale conversion by non-linear filter bank as [2]:

$$\text{Mel} = 2595 \log_{10}(1 + F/700) \quad (2)$$

-Computing the inverse DCT to retrieve MFCC as illustrated in expression (2):

$$\text{MFCC}(k) = \sum_{i=1}^N \log(E_i) \cos\left[\frac{\pi k}{N}(i - 0.5)\right] \quad (3)$$

with:

$E_i$ : energy coefficient;

$N$ : is the number of the filter bank.

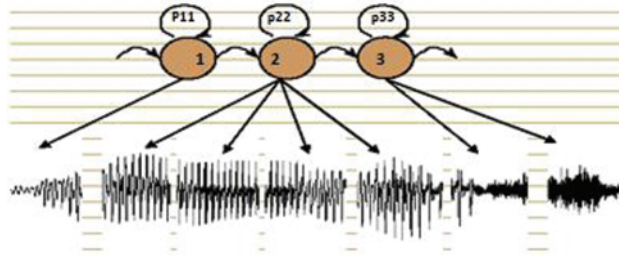
The vector of the acoustic parameters is constituted of 12 MFCC coefficients +  $\Delta$  +  $\Delta\Delta$ : energy derived [5].

### 2.3 HMM Training

Every word of the codebook is treated and expressed as a HMM model characterized by ( $N$ ) states and the transition probabilities  $p(i)$ . For example, Fig. 4 illustrates a HMM model of the word “RUN” with 3 HMM states.

The implementation of a recognition system based on hidden Markov models (HMM), has three phases:

1. Description of a network whose topology reflects the phrases, vocabulary words or basic units.
2. Make the training mode settings:  $\lambda = (\pi, A, B)$ .



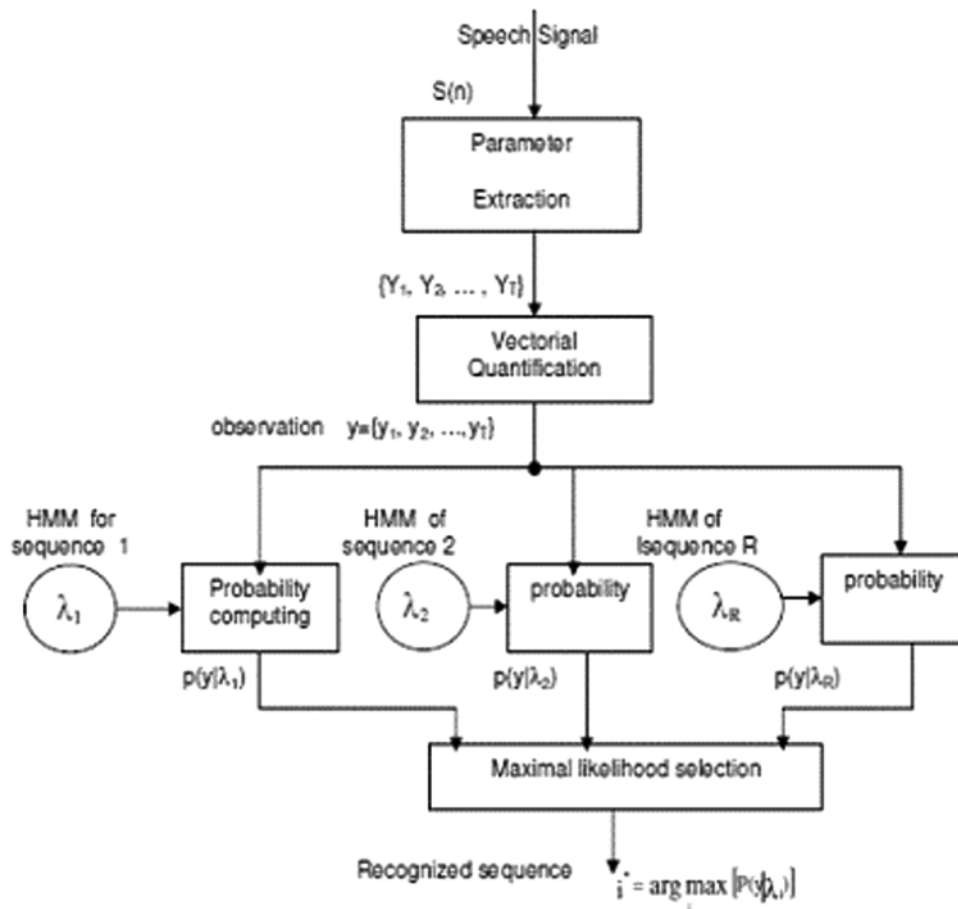
**Figure 4:** HMM model with 3 states

3. Carry out the actual recognition occurrence by calculating the maximum likelihood.

At this phase, each vocabulary word is associated with a Hidden Markov model (HMM) represented by Fig. 5 [7]:

$$\lambda = (A, B, \pi) \quad (4)$$

Since  $Q = \{q_1, q_2, \dots, q_T\}$  is the optimal state sequence that represents the sequence of observations  $\{y = y_1, y_2, \dots, y_T\}$  then the model is defined in function of:



**Figure 5:** Algorithm of an isolated word recognition system

1. The number of states  $S$ .
2. The number of possible observations of  $K$  codebook.
3. The matrix of transitions probabilities:  $A$ .
4. The matrix of probabilities associated with observation  $B$ .
5. The Initial probability  $\pi_i$  of each of the  $S$  states.

As each word is represented by a vector  $y = \{y = y_1, y_2, \dots, y_T\}$  among the  $K$  codebook vectors, the training problem will be to re-estimate the parameters  $(A, B, \pi)$  of  $l$  model to maximize the probability  $p(y|\lambda)$ .

## 2.4 Vector Quantization VQ

VQ is an operation that represents a vector ( $y$ ) with  $N$  components deduced from  $K$  vectors of a codebook. Vector quantization is organized to minimize the average distortion and uses the famous K-means algorithm [6].

## 2.5 Recognition

As each unknown word which will be recognized, the treatment described in Fig. 1 is to be performed by the observation sequence  $\{y = y_1, y_2, \dots, y_T\}$ , and a vector quantization, followed by a calculation of the likelihood probabilities for all possible models  $p(y|\lambda_i)$ . The selection of the word whose likelihood probability is the greatest will optimize the identification the desired word [5].

## 3 Simulation Results

### 3.1 Speech Database

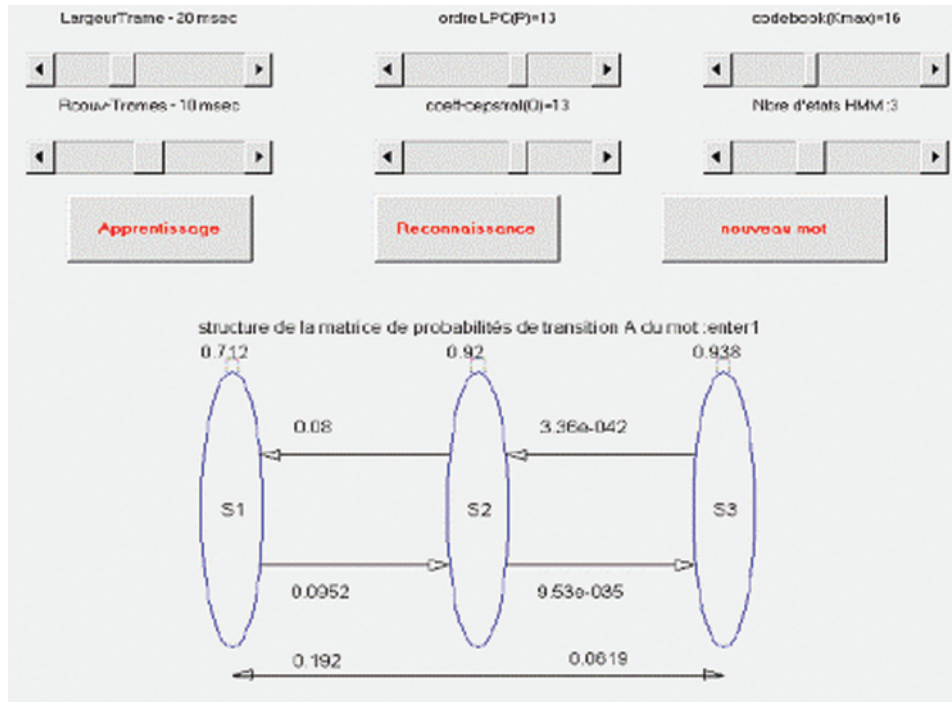
We used a codebook of isolated words for PC communication and control (as a motor control and wheelchair voice control). The codebook has the following words of Tab. 3.

**Table 3:** Example of the control words

Control word	Significance	Recognition ratio
Khalf	Back	78%
Amam	Forward	97%
Asraa	Accelerate	88%
Sir	Go	99%
Istader	Return	74%
Takadam	Start	83%
Tarajaa	Decelerate	88%
kif	Stop	99%
Yamin	Left	94%
Yassar	Right	91%
laa	No	98%
Naam	Yes	95%
Ayid	Repeat	91%

### 3.2 Results

We developed a Matlab GUI interface for processing, training and recognition which allows us to choose several settings, such as the window type and size, recovery time, LPC order, number of MFCC coefficients, the codebook size and the number of HMM states. The algorithm computes the likelihood ratio and deduces the recognized word [8]. For example, we report on Figs. 6 and 7 the recognition results for the word “Khalf” which signifies go-back. Each figure contains three red boxes (training, recognition and classification). With a sliding cursor, we can choose the processing parameters of every step. At the bottom of the figures, we observe the word HMM model represented with three states and with computed probabilities.



**Figure 6:** HMM modeling and training

To evaluate the sensitivity of our application to the language, we added another database using another language (French). The results gave similar and sometimes better recognition rates reaching 99% such as illustrated in Tab. 4. We can conclude that our system is valuable to several languages and can be used by every user after training.

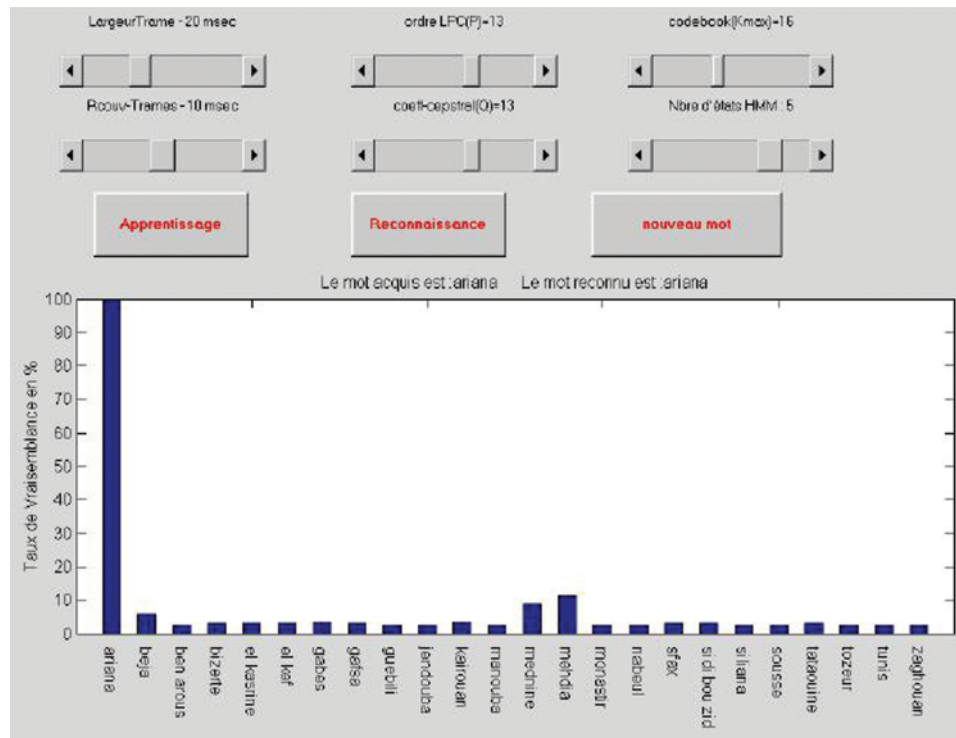
Figure 8 represents the obtained recognition ratio for the others words of the codebook. We can observe that these values vary between 70% (for the word help) to 99% (for the words enter and yes).

The recognition tests results of Tab. 5, demonstrate that our ASR control system can reach a recognition rate of 100% in safety conditions with a reduced time of processing (about 0.24 s). The simulation results of Fig. 6 show that the maximal recognition ratio for this word is about 100% with a good discrimination with the others control words.

### 3.3 Tests under Noisy Environments

We note that the recognition ratio decreases with noisy environments [9]. According to Fig. 9 and Tab. 5, the recognition ratio decreases from 98% to 75% with the presence of noise (SNR < 5 dB). These results are





**Figure 7:** Recognition ratio: word “khalf”: go-back

**Table 4:** Example of the control words

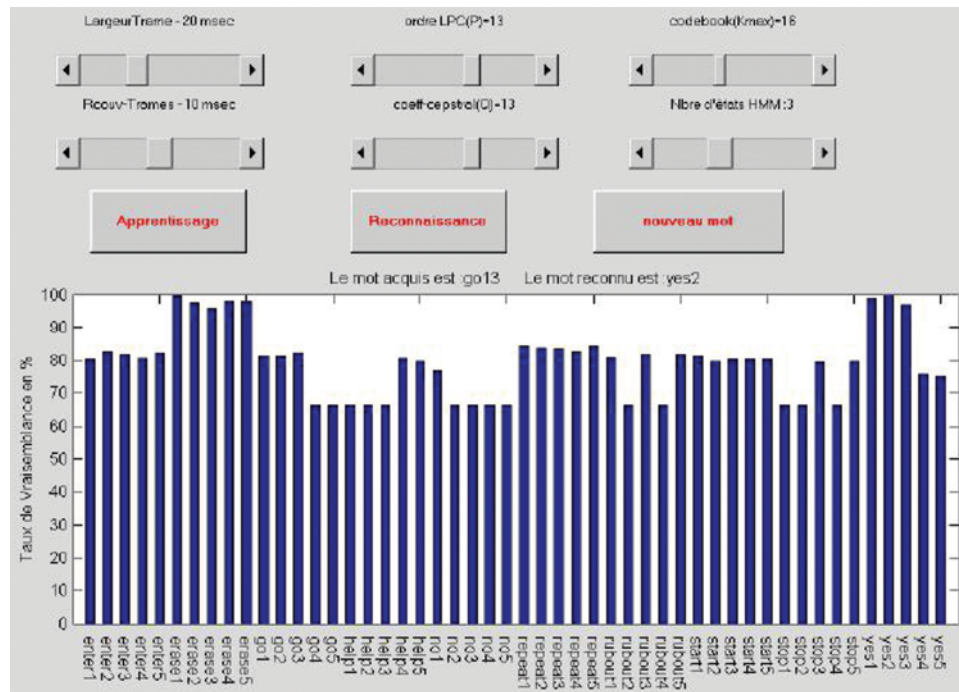
Control word	Significance	Recognition ratio
Reculer	Back	89%
Avancer	Forward	95%
Démarrer	Start	86%
Arrêter	Stop	95%
Retourne	Return	86%
Droite	Right	99%
Gauche	left	96%

confirmed by others researches which have studied the effect of noisy level on the recognition ratio and speech intelligibility such as in automobile and telecommunications [10].

With HMM system, the RR reaches 100% by using MFCC features and SNR levels between 15-30 dB. Yet, the RR decreases to 25% for high noisy environments with SNR between 0 dB and −5 dB. In this case, the use of a noisy database (Noisex DB) becomes necessary for training. We can also propose the insertion of a speech enhancement and de-noising step which can be added after speaker acquisition and integrated in the sound device acquisition of the wheelchair [11].

Figure 10 gives an illustration of the confusion values of the Recognition Ratio (RR) in the case of a noisy environment of speech acquisition and control (example in the street). We can observe in Fig. 10 that the RR varies between 10% and 99%.

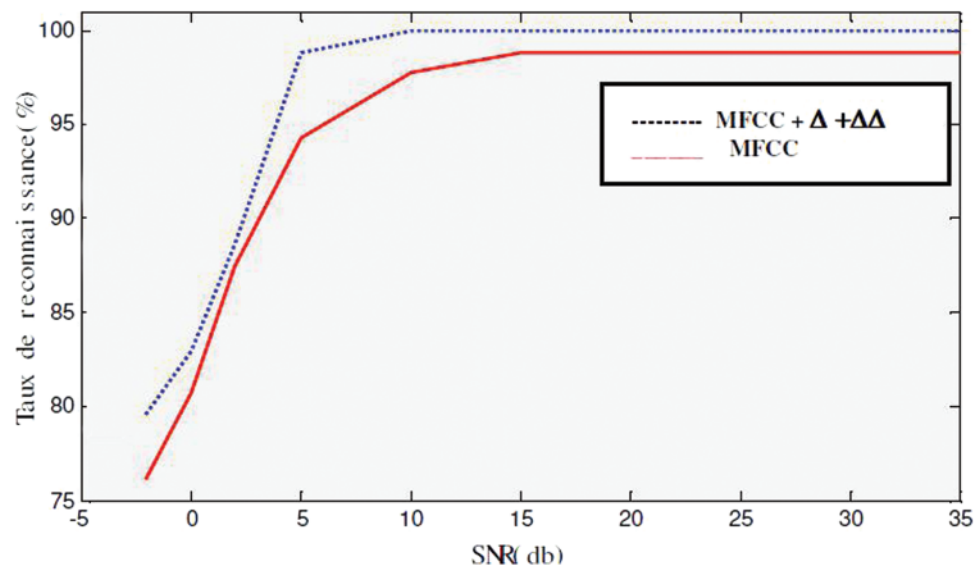




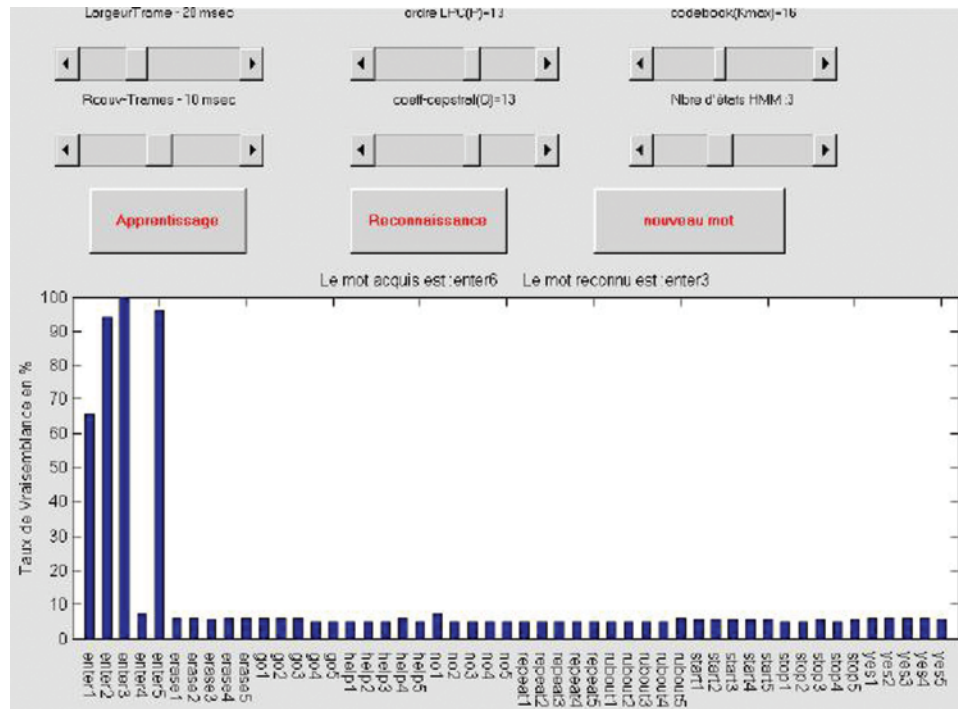
**Figure 8:** Recognition Ratio (%) evaluation of several words of the database

**Table 5:** ASR performances

Recognition method	Training time	Recognition Time	Recognition ratio RR
MFCC+ $\Delta$ + $\Delta\Delta$ + MMC	1,56 mn	0.24 s	99.8%



**Figure 9:** ASR recognition ratio vs. Noise level (SNR)



**Figure 10:** Recognition ratio: for the word “enter” in noisy environment

For example: For the word “Enter 4” with SNR = −5 dB, we obtained a RR about 10%, this word corresponds to the control word “enter” pronounced in a very noisy environments (car, street,...), for the word “Enter 1” with SNR = +5 dB, we obtained a RR about 65%, for the word “Enter 3” with SNR = 20 dB, we obtained a RR about 99%.

### 3.4 Comparison with Previous Studies

In order to validate our results and to compare it with others references, we presented in [Tab. 6](#) an overview of the main industrial and research performances of ASR systems in the last 10 years [12, 13]. We can observe that the system accuracy varies between 56% and 98% (MFCC + DWT).

The feature extraction with hybridizations between Mel cespstral coefficients and discrete wavelet coefficients gives the better values and performances according to [Tab. 6](#). If we compare our ASR

**Table 6:** ASR accuracy

SNR in dB	Recognition ratio
30	100%
20	99.80%
10	95%
5	78%
0	25%
−5	12%

system, which has a RR from 78% (respectively from 25 dB to 5 dB) with the others works of Tab. 6; we can conclude that we obtain satisfactory and good performances especially in the case of clean environment.

## 4 Experimental Implementation

### 4.1 Electronic Conception

Figures 11 and 12 illustrate the prototype of the wheelchair on DSP TMS320-C6416 de Texas instruments [14]. After the control word is recognized by the ASR engine, the DSP generates codes in its output to activate and control the dc motors corresponding to the voice control (forward, backward, left, right and stop,...). The card delivers six digital outputs corresponding to the recognized words. These values are observed also by 6 diodes (Fig. 13). The programming of the speech processing is made by visual DSP++ language, the Matlab Simulink and C++ runtime server.

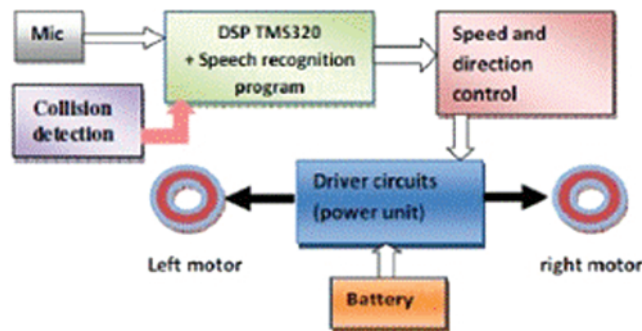


Figure 11: DSP implementation of the wheelchair

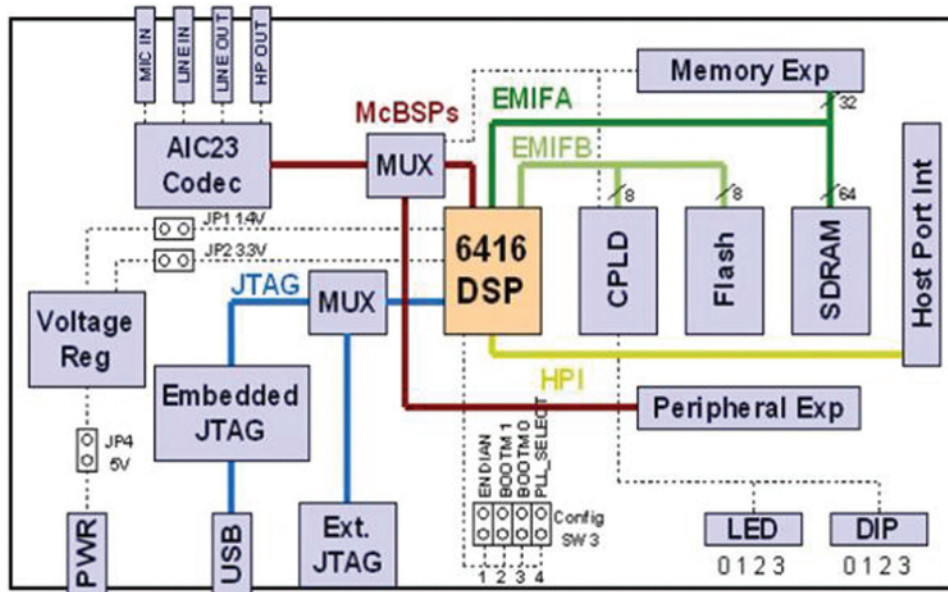
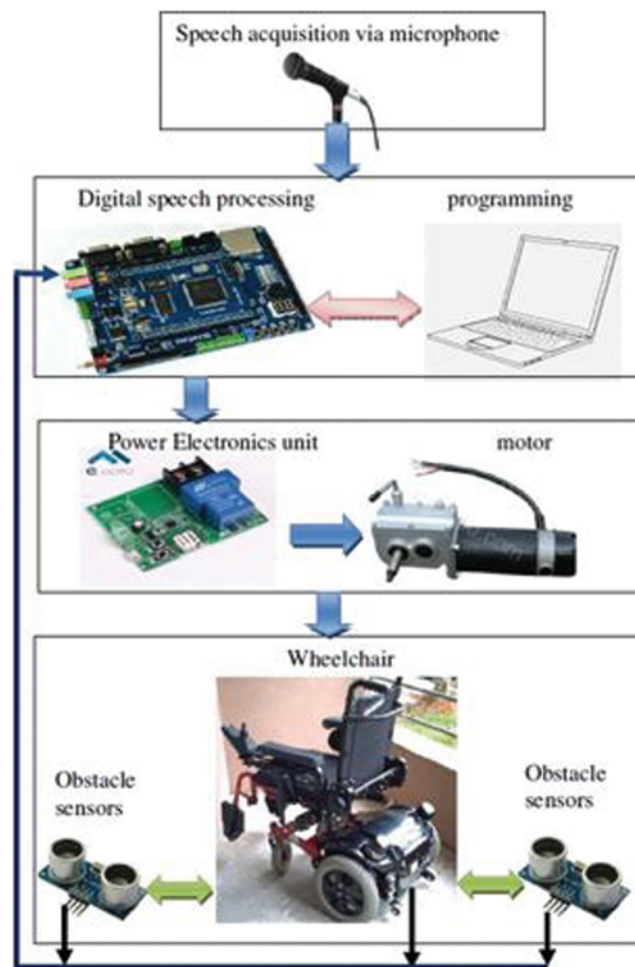


Figure 12: DSP structure

### 4.2 Validation

In order to evaluate the performance performances, we used in Tab. 7 two objective criteria: the recognition ratio RR [15] and the word Error Rate WER, which is defined as [16, 17]:



**Figure 13:** Recognized speech word

**Table 7:** Performances evaluation of RR and WER

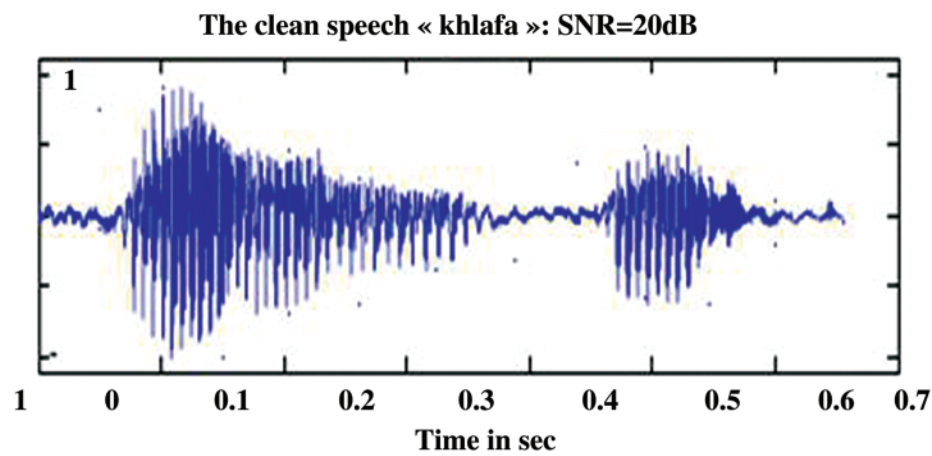
	User1				User2			
Control words	Test1	Test2	Test3	Test4	Test1	Test2	Test3	Test4
go	●	●	●	●	●	●	●	●
forward	●	●	●	●	●	●	●	●
accelerate	●	●	●	●	●	●	●	●
left	●	●	●	●	●	●	●	●
right	●	●	●	●	●	●	●	●
backward	●	●	●	●	●	●	●	●
stop	●	●	●	●	●	●	●	●
start	●	●	●	●	●	●	●	●
back	●	●	●	●	●	●	●	●
repeat	●	●	●	●	●	●	●	●
Number of recognized words	7/10	10/10	7/10	6/10	8/10	10/10	7/10	6/10
WER (%)	30	0	30	40	20	0	30	40

● Correct Action    ● false action    ● no action

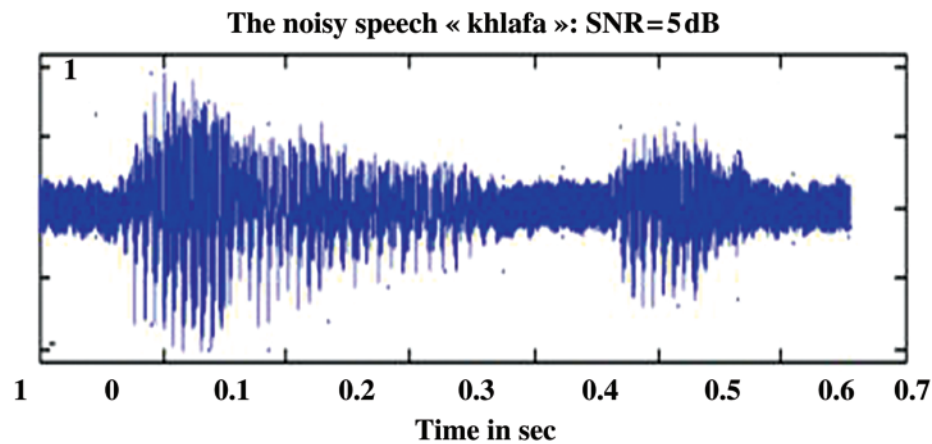
$$WER = \frac{S + D + I}{N} \quad (5)$$

where:  $N$  is the number of words,  $S$  is the number of substitutions (wrong words after recognition),  $D$  is the number of suppressions (forgotten words after transcription) and  $I$  is the number of insertions (added words along the recognition step).

In order to validate the tests, the recognition system and the speech control interface, we illustrated in Fig. 14 and Fig. 15 the experimental speech signals control of the recognized control word “khalfa” for several SNR values (noise, clean, moderate).



**Figure 14:** Speech waveform of the word: “khalfa” with SNR = 20dB



**Figure 15:** Recognized speech word “go-back: khalfa”. Noisy speech: SNR = 5 dB

#### 4.3 Energy and Economic Evaluation

The wheelchair is equipped with 2 motors of 250 Watt each, a rechargeable battery of 30 AH/24 V, a DSP card equipped with microphone, proximity sensors and electronic components.

To evaluate the energy consumption and the cost of the installation, we integrated the costs of each element of the system, which gave the following results:

- Total cost : 750 Dollars
- Energy consumption: 480 WH/day.

## 5 Conclusion

In this paper we have developed an Arabic voice control of wheelchairs intended for handicapped persons. We used an ASR system using HMM with a vocabulary of 39 words (for control). We reached very respectable rates of recognition ratio (RR) around 99%. However, the simulation results show that the RR is very sensitive to noise especially for SNR signals less than 5 dB. In this case, we suggest the addition of a noise suppression procedure such as sub-spectral, wiener or wavelets techniques. Finally, we succeeded to implement a new biometric control on embedded platforms TMS320-C6416 allowing a real time functioning and monitoring. As a work perspective, it is envisaged to make the chair more intelligent to automatically recognize the places of movement (kitchen, bathroom and garden) without guidance by the user. This can be done by a supervised training and geo-location system.

**Funding Statement:** The author(s) received no specific funding for this study.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

1. Benmosbah, B. (2005). *Utilisation de la mémoire de parole pour la reconnaissance: applications pour les personnes handicapées (Ph.D. Thesis)*. Ecole Nationale Supérieure des Télécommunications, Paris, France.
2. Boufif, L. (2010). *A speech coding strategy for cochlea implants (Ph.D. Thesis)*. ENIT, Tunis Manar university.
3. Bruno, J. (2009). *IT tool Markov models management hidden: experiments automatic speech recognition. Ph.D. Thesis* in Information Technology.
4. Saini, P., Parneet, K. (2013). Automatic speech recognition. *International Journal of Engineering Trends and Technology*, 4(2), 132–136.
5. Sandipan, C., Goutam, S. (2009). Improved text-independent speaker identification using fused MFCC and IMFCC feature sets based on Gaussian filter. *International Journal of Information and Communication*, 5, 11–19.
6. Sunny, S. (2012). Development of a speech recognition system for speaker independent isolated Malayalam words. *International Journal of Computer Science Engineering and Technology*, 3(4), 69–75.
7. Rabiner, L., Juang, B. H. (2006). *Fundamentals of speech recognition*. Collections Prentice Hall. Book chapter: ISBN-10: 0130151572.
8. Maucclair, J. (2016). *Mesure de Confiance en Traitement Automatique de la Parole Applications*. Thèse de Doctorat à l'Université du Maine.
9. Sebastien, G. (2006). *Réduction de dimension en apprentissage numérique non-supervise. Ph.D. Thesis* in informatics Université Paris 13, France.
10. Loïc, B. (2008). *Diagnostic pour la combinaison de systèmes de reconnaissance automatique de la parole*. thèse de l'Université d'Avignon.
11. Min, J. W., Lee, K., Lim, K. C., Kwon, D. S. (2002). Human-friendly interfaces of wheelchair robotic system for handicapped persons. *Intelligent Robots and Systems (IEEE/RSJ)*, 1505–1510. DOI 10.1109/IRDS.2002.1043968.
12. Michelle, C., Edward, G., Ivan, G., Owen, C., Joseph, M. (2013). Comparative study of automatic speech recognition. *IET Signal Processing*, 7, 25–46. DOI 10.1049/iet-spr.2012.0151.
13. Tobias, O., Koller, A. (2010). *Automatic speech recognition and identification of African Portuguese (Ph.D. Thesis)*. Univ Lisbon.

14. Bennisr, M. (2016). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27, 113–120.
15. Haton, J. P., Cerisara, C., Fohr, D., Laprie, Y., Smaili, K. (2010). *Reconnaissance automatique de la parole: du signal à son interprétation*. Collections Dunod.
16. Guerig, S. (2016). *Reduction de dimension en Apprentissage numerique non supervisé*. Thèse PhD, Université de Paris 13, France.
17. Cherif, A., Benasr, M. (2017). Real time implementation on DSP of speech enhancement based on kalman filter and wavelet thresholding. *Indian Journal of Science and Technology*, 10(24), 1–7.