Vowels as Human Computer Interface Towards Web Navigation

Mohamed Fezari, N. Benouioua, Ahmed Al Dahoud
Badji Mokhtar Annaba University, Algeria
Bradford University, UK

Abstract

In this paper we experiment the use of vowels to activate the movement of mouse pointer on the screen. The control of the windows icon mouse pointer (WIMP) by voice command is currently based on using vowel utterances, this category of letters is easy to recognize and to be pronounced, especially for individuals with speech pathology disorder and hands gestures r. So this type of Human-Computer interface could be of great help for this category of persons. Moreover, vowels are quite easy to model by automatic speech recognition (ASR) systems. In this work we represent the design of a system for the control of mouse cursor based on voice command, using the pronunciation of certain vowels and sylables. The Mel Frequency Cepstral Coefficients (MFCCs), fundamental frequency (F_0) and Formants (F_1,F_2) are selected as features. The TDW with Euclidian Distance and Hidden Markov Models (HMMs) have been tested as classifiers for matching components (vowels and short words). Comparison between different features and classifiers were tested and results are presented on tables, finally a GUI has been designed for user applications, then an example on Web navigation has been presented.

1. Introduction

Existing human-computer interfaces are not suited to individuals with upper limb motor impairments. Recently, a lot of interest is put on improving all aspects of the interaction between human and computer especially for this category of persons, however these devices are generally more expensive example sip-and-switches [1] eye-gas and eye tracking devices[4], headmouse [2,3] chin joystick[5] and tongue switches [6]. Here is some related works on human computer interaction, based on voice activation or control, which can be invested for individuals with motor impairments. Most of concepts of vocal commands are built on the pronunciation of vowels [5, 6, and 7], where the particularity of vowels used is the simple and the regular pronunciation of these phonemes. Many vocal characteristics are exploited in several works, but the most used are: energy [1, 2, 3 and 5], pitch and vowel quality [9,10] speech rate (number of syllables per second) and volume level [7]. However, Mel Frequency Cepstral Coefficients (MFCCs) [11, 12 and 13] are used significantly of speech processing as bio-inspired feature for automatic speech recognition of isolated words [15-16].
The paper is organized as follows: in section 2, presentation of an overview on related works of mouse cursor control based on voice control and commands. In section 3, we showed LPC and MFCC computation and use as features extraction techniques. Then we describe used classifiers: DTW then HMM in section 4. In section 5, we present tests and results. And finally, we provide graphic user interface as an application.

2. Related Works

We describe some related works with vocal command system in the literature review. Voice recognition allows you to provide input to an application with your voice. In the basic protocol, each vowel is associated to one direction for pointer motion [1].This technique is useful in situations where the user cannot use his or her hands for controlling applications because of permanent physical disability or temporal task-induced disability. The limitation of this technique is that it requires an unnatural way of using the voice [5] [6]. Control by Continuous Voice: In this interface, the user's voice works as an on/off button. When the user is continuously producing vocal sound, the system responds as if the button is being pressed. When the user stops the sound, the system recognizes that the button is released. For example, one can say "Volume up, ahhhhhh", and the volume of a TV set continues to increases while the "ahhh" continues. The advantage of this technique compared with traditional approach of saying "Volume up twenty" or something is that the user can continuously observes the immediate feedback during the interaction. One can also use voiceless, breathed sound [6].

In [4] The HeadMouse Extreme is an infrared optical sensor that tracks side-to-side and up-and-down head movements. It then filters and transforms these movements to control the position of the mouse pointer on the computer screen. HeadMouse Extreme operates with very low input power (1 watt) that is

usually supplied directly from the computer or hub through a standard USB cable. HeadMouse Extreme uses infrared light to track a tiny reflective dot that is placed on the user's forehead or glasses.

When used with mouse button software, such as Origin Instruments' Dragger, mouse clicks are performed by positioning the mouse pointer and dwelling for a selectable period of time. Alternately, mouse clicks can be performed with an adaptive switch, such as the Origin Instruments Sip/Puff Switch. One or two adaptive switches can be connected directly to HeadMouse Extreme through a 3.5 mm stereo jack located next to the USB connector.

The HeadMouse Extreme also contains an integrated infrared receiver for use with wireless switches. The optional Beam™ Wireless Switch Transmitter supports wireless connections for up to three adaptive switches. See figure 1, on how to use this moduel.

Figure 1. HeadMouse Extrem mounted on a desktop display. Fromdatasheet

Alex Olwal et al. [7] have been experimenting with non verbal features in a prototype system in which the cursor speed and direction are controlled by speech commands. In one approach, speech commands provide the direction (right, left, up and down) and speech rate controls the cursor speed. Mapping speech rate to cursor speed is easy to understand and allows the user to execute slow. The cursor's speed can be changed while it is moving, by reissuing the command at a different pace. One limitation of using speech features is that they are normally used to convey emotion, rather than for interaction control.

The detection of gestures is based on discrete pre-designated symbol sets, which are manually labeled during the training phase. The gesture-speech correlation is modeled by examining the co-occurring speech and gesture patterns. This correlation can be used to fuse gesture and speech modalities for edutainment applications (i.e. video games, 3-D animations) where natural gestures of talking avatars are animated from speech [7] [8].

Bilmes J. et al. [9] have been developed a portable modular library (the Vocal Joystick"VJ" engine) that can be incorporated into a variety of applications such as mouse and menu control, or robotic arm manipulation. Our design goal is to be modular, low-latency, and as computationally efficient as possible. The first of those, localized acoustic energy is used for voice activity detection, and it is normalized relatively to the current detected vowel, and is used by our mouse application to control the velocity of cursor movement. The second parameter, "pitch", is not used currently but it is left for the future use. The third parameter: "vowel quality", where the vowels are characterized by high energetic level. The classification of vowels is realized by extraction of two first formants frequencies, tongue height and tongue advancement [9, 10]. Thus, the VJ research has focused on real time extraction of continuous parameters since that is less like standard ASR technology [9]. The main advantage of VJ is the reaction of the system in real time.

In [14], Thiang et al., described the implementation of speech recognition system on a mobile robot for controlling movement of the robot. The methods used for speech recognition system are Linear Predictive Coding (LPC) and Artificial Neural Network (ANN). LPC method is used for extracting feature of a voice signal and ANN is used as the recognition method. Backpropagation method is used to train the ANN. Experimental results show that the highest recognition rate that can be achieved by this system is 91.4%. This result is obtained by using 25 samples per word, 1 hidden layer, 5 neurons for each hidden layer, and learning rate 0.1.

3. Main Features Selection

In order to implement the HMI application on embedded system in future, and to get good results in automatic speech recognition is to select better and easy to compute features, so the features would be robust and fast to compute. The LPC, MFCC with energy and derivatives were selected based on literature reviews [15, 16] and [17], Figure 1 presents the synoptic to compute these features.

Figure 1. Ilustration of flowcahrt to compute LPC and MFCC features

3.1. MFCC Feature extraction [11]

The extraction of the best parametric representation of acoustic signals is an important task to produce a better recognition performance. The efficiency of this phase is important for the next phase since it affects its behavior. MFCC is based on human hearing perceptions which cannot perceive frequencies over 1Khz. In other words, in MFCC is based on known variation of the human ear's critical bandwidth with frequency. MFCC has two types of filter which are spaced linearly at low frequency below 1000 Hz and logarithmic spacing above 1000Hz. A subjective pitch is present on Mel Frequency Scale to capture important characteristic of phonetic in speech. The overall process of the MFCC can be presented in the following steps:

1. After the pre-emphasis filter, the speech signal is first divided into fixed-size windows distributed uniformly along the signal.

2. The FFT (Fast Fourier Transform) of the frame is calculated. Then the energy is calculated by squaring the value of the FFT. The energy is then passed through each filter Mel. S_k: is the energy of the signal at the output of the filter K, we have now m_p (number of filters) S_k parameters.

3. The logarithm of S_k is calculated.

4. Finally, the coefficients are calculated using the DCT (Discrete Cosine Transform).

$$c_i = \sqrt{\frac{2}{m_p}} \left\{ \sum_{k=1}^{m_p} \log\left(S_k\right) \cos\left[i \left(k - \frac{1}{2} \right) \frac{\pi}{m_p} \right] \right\} \quad (1)$$

$$pour\ i = 1 \ldots\ldots N$$

N: is the number of MFCC coefficients.

3.2 Fundamental Frequency and formants extraction

Linear predictive analysis of speech has become the predominant technique for estimating the basic parameters of speech. Linear predictive analysis provides both an accurate estimate of the speech parameters and also an efficient computational model of speech.

The basic idea behind linear predictive analysis is that a specific speech sample at the current time can be approximated as a linear combination of past speech samples. Through minimizing the sum of squared differences (over a finite interval) between the actual speech samples and linear predicted values a unique set of parameters or predictor coefficients can be determined.

LPC computation basic steps can be presented as follow [14]:

a) Pre-emphasis: The digitized speech signal, *s*(*n*), is put through a low order digital system, to spectrally flatten the signal and to make it less susceptible to finite precision effects later in the signal processing.

b) Frame Blocking: The output of pre-emphasis step $\tilde{s}(n)$ is blocked into frames of *N* samples, with adjacent frames being separated by *M* samples. If $x_l(n)$ is the l^{th} frame of speech, and there are *L* frames within entire speech signal.

c) Windowing: After frame blocking, the next step is to window each individual frame so as to minimize the signal discontinuities at the beginning and end of each frame. If we define the window as *w* (*n*), $0 \leq n \leq N-1$, then the result of windowing is the signal:

$$\tilde{x}_l(n) = x_l(n) w(n)$$

d) Autocorrelation Analysis: The next step is to auto correlate each frame of windowed signal in order to give:

$$r_l(m) = \sum_{n=0}^{N-1-m} \tilde{x}_l(n) \tilde{x}_l(n+m) \quad (2) \qquad (3)$$

$$m = 0, 1, \ldots p$$

e) LPC Analysis: which converts each frame of *p* + 1 autocorrelations into LPC parameter set by using Durbin's method.

f) LPC Parameter Conversion to Cepstral Coefficients: LPC cepstral coefficients, is a very important LPC parameter set, which can be derived directly from the LPC coefficient set. The recursion used is:

(1)

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) . c_k . a_{m-k} \qquad 1 \leq m \leq p \quad (3)$$

And:

$$c_m = \sum_{k=m-p}^{m-1} \left(\frac{k}{m} \right) . c_k . a_{m-k} \qquad (4)$$

$$m > p$$

The LPC cepstral coefficients are the features that are extracted from voice signal and these coefficients are used as the input data for the classifier (Euclidian Distance or DTW). In this system, voice signal is sampled using sampling frequency of 8 kHz and the signal is sampled within 1.5 seconds, therefore, the sampling process results 1200 data. Because we choose LPC parameter N = 200, m = 100, and LPC order = 10 then there are 119 vector data of LPC cepstral coefficients.

4. Presenting Selected Classifiers

In pattern recognition in general, automatic speech recognition, speaker Identification, image or shape recognition we need some how an algorithm to classify.

4.1 DTW (Dynamic Time Warping)

DTW algorithm is based on Dynamic Programming techniques .This algorithm is for measuring similarity between two time series which may vary in time or speed. This technique also used to find the optimal alignment between two times series if one time series may be "warped" non-linearly by stretching or shrinking it along its time axis.

This warping between two time series can then be used to find corresponding regions between the two time series or to determine the similarity between the two time series.

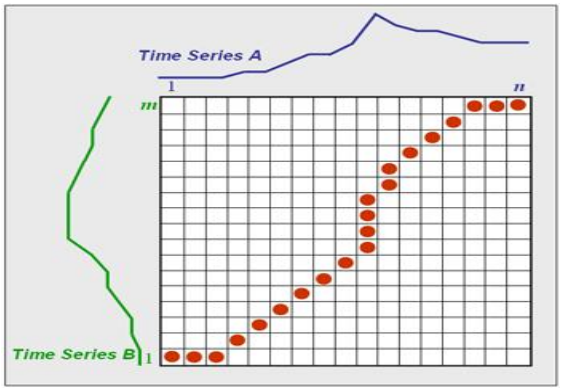

Figure 2. The optimal warping path from [22]

4.2 Euclidean distance formula

The Euclidean distance between points p and q is the length of the line segment connecting them ($\overline{\mathbf{pq}}$). In Cartesian coordinates, if $\mathbf{p} = (p_1, p_2,..., p_n)$ and $\mathbf{q} = (q_1, q_2,..., q_n)$ are two points in Euclidean *n*-space, then the distance (d) from **p** to **q**, or from **q** to **p** is given by the Pythagorean formula:

$$\text{dist}((x, y), (a, b)) = ((x - a)^2 + (y - b)^2)^{1/2} \quad (5)$$

$$\text{Dist}(q_i,p_i) = \text{Sum}(q_i - p_i)^2 \quad \text{for } i=1..n \quad (6)$$

The position of a point in a Euclidean *n*-space is a Euclidean vector. So, p and q are Euclidean vectors, starting from the origin of the space, and their tips indicate two points. The Euclidean norm, or Euclidean length, or magnitude of a vector measures the length of the vector:

$$\|\mathbf{p}\| = \sqrt{p_1^2 + p_2^2 + \cdots + p_n^2} = \sqrt{\mathbf{p} \cdot \mathbf{p}} \quad (7)$$

where the last equation involves the dot product.

A vector can be described as a directed line segment from the origin of the Euclidean space (vector tail), to a point in that space (vector tip). If we consider that its length is actually the distance from its tail to its tip, it becomes clear that the Euclidean norm of a vector is just a special case of Euclidean distance: the Euclidean distance between its tail and its tip.

The distance between points **p** and **q** may have a direction (e.g. from **p** to **q**), so it may be represented by another vector, given by

$$\mathbf{q} - \mathbf{p} = (q_1 - p_1, q_2 - p_2, \cdots, q_n - p_n) \quad (8)$$

If D(x,y) is the Euclidean distance between frame x of the speech sample and frame y of the reference template, and if C(x,y) is the cumulative score along an optimal alignment path that leads to (x,y), then:

$$C(x,y)=\text{MIN}(C(x-1,y),C(x-1,y-1),C(x,y-1))+D(x,y) \quad (9)$$

4.3 HMM Classifier

Hidden Markov Models have been widely applied in several models like pattern, or speech recognition. To use HMM in automatic recognition, we need a training phase and a test phase. For the training stage, we usually work with the Baum-Welch algorithm to estimate the parameters (π,A,B) for the HMM. This method is based on the maximum likelihood criterion. To compute the most probable state sequence, the Viterbi algorithm is the most suitable [13].

An HMM model is basically a stochastic finite state automaton, which generates an observation string, that is, the sequence of observation vectors, O = O1, .. Ot, ... ,OT. Thus, a HMM model consists of a number of N states S={Si} and of the observation string produced as a result of emitting a vector 'Ot' for each successive transitions from one state Si to a state Sj. 'Ot' is d dimension and in the discrete case takes its values in a library of M symbols.

The state transition probability distribution between state Si to Sj is A={aij}, and the observation probability distribution of emitting any vector 'Ot' at state Sj is given by B={bj(Ot)}. The probability distribution of initial state is Π={ πi}

$$a_{ij} = P(q_{t+1} = \frac{S_j}{q_t} = S_i) \quad (10)$$

$$B = \{b_j(O_t)\} \quad (11)$$

$$\pi_i = P(q_0 = S_i) \quad (12)$$

Given an observation O and a HMM model λ=(A,B,Π), the probability of the observed sequence by the forward-backward procedure P(O/λ) can be computed. Consequently, the forward variable is defined as the probability of the partial observation

sequence O1, O2, ...,Ot (until time t) and the state S at time t, with the model λ as α(i). and the backward variable is defined as the probability of the partial observation sequence from t+1 to the end, given state S at time t and the model λ as β(i). The probability of the observation sequence is computed as follow:

$$p(o/\lambda) = \sum_{i=1}^{N} \alpha_t(i) * \beta_t(i) = \sum_{i=1}^{N} \alpha_T(i) \quad (13)$$

And the probability of being in state I at time t, given the observation sequence O and the model λ is computed as in (13).

5. New Interface Description

The application is designed to control the mouse cursor by using the pronunciation of certain phonemes and words, which we chose as vocabulary: " aaa", " ooh", " iii"," eeu"', "ou", "uu", " Clic" and " stop".

The choice of these vowels and short words is based on the following criteria:

• Easy to learn.

• Easy to pronounce.

• Can be pronounced persons with voice disorder.

• Easy to recognize by automatic speech recognition system.

5.1. DataBase Description

The database consists of 10 women (age 20 to 50 years), 10 men (age 20 to 60 years), and 5 children (age from 5 to 14 years) and category of persons with voice disorder from German database of the PTSD Putzer's voice in [18], each speaker had: 5 trials for each phoneme or word. Collection of the database is performed in a quiet room without noise.

5.2. The parametrization

According to the tests, we found that the parameters more robust to noise than other parameters are the LPC coefficients and Mel Frequency Cepstral Coefficients (MFCCs). The input signal is segmented by a window of 25 ms overlapping 10ms, from each segment parameters were extracted by both methods LPC (the order of the prediction: 10) then MFCC (42 coefficients: Energy and derivative and second derivatives).

5.3. Used Classification

For this moment, we have tested two classifier, first one has been used for simplicity in order to be implemented in future on DSP circuit of microcontroller: Dynamic Time Warping (DTW) with Euclidian distance and Hidden Markov chains (HMM) for classification phase.

For Hidden Markov models, in our system, we utilize left-to-right HMM structures with 3 states and 3 mixtures are used to model MFCCs coefficients.

5.4. Application

Our application is used to control the mouse cursor by voice, pronouncing a vowel or short words above. Based on just formats frequencies of vowels, we distinguish a discrimination as shown in figure 3.1. The vowels are mapped to directions of movement cursor and push buttons on mouse as follow and presented in figure 3.2:

- Up:" ooh"
- Down:" aah"
- To the right:" iii"
- Left:" eeu"
- To double-click (open):" click" or "eke"
- To exit the application by voice command:" stop" or"abe".
- Left-Click : "ou"
- Right-Click:" "uu"

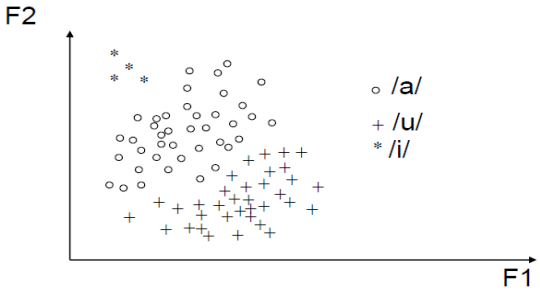

Figure. 3.1. Discrimination between vowels 'a', 'u' and 'I' using Formants F1 and F2

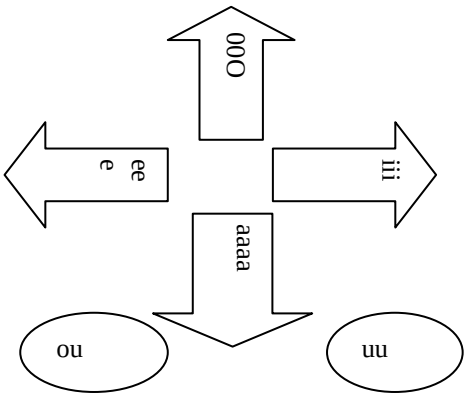

Figure. 3.2. Directions of cursor mouse mapping from vowel

The flowchart of the developed system application is presented in Figure 4, where two basic

classifiers were presented, i.e. DTW and Euclidian Distance (DE) with two types of features. Based on results obtained in Table 3, we notified that even if the DE classifier is easy to compute, the results are weak, then we intended to improve that result by using HMM classifier.

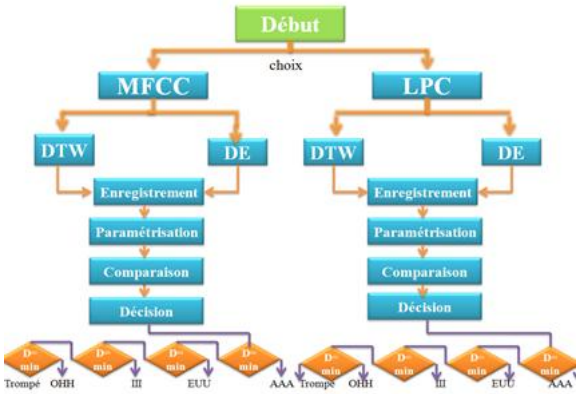

Figure 4. Flowchart of DTW and DE classifier

6. Tests, Results and Discussion

For the testing phase, 20% of recorded sounds are selected for each vowel or short word from the vocabulary.

In order to see the effect of training and making the system speaker independent, different scenarios for the tests were done, where we choose the results of recognition of three users out of database.

Some vowels and short words were correctly classified with some confusion, where a phoneme (or word) test classified as another phoneme (or word), the misclassification is presented in the tables 1 and 2 below. And it is obvious that the confusion is higher in LPC features with DTW classifier while it is reduced using MFCC with HMM classifier.

Table 1.Confusion Table using (mfcc/hmm)

Pronounced Vowel	**Classified as:**						
	aaa	***ooh***	***eeu***	***iii***	***clic***	***stop***	***ou***
aaa	o	x	-	-	-	x	-
ooh	x	o	-	-	-	x	-
eeu	-	x	o	x	-	-	x
iii	-	-	-	o	x	-	-
ou	-	x	x	-	-	-	o
uu	-	-	x	-	-	-	x
Clic or “eke”	-	-	-	-	o	x	-

x: means that pronounced phoneme classified as an other

Table 2.Confusion Table using (LPC/DTW)

Pronounced Vowel	**Classified as:**						
	aaa	***ooh***	***eeu***	***iii***	***clic***	***stop***	***ou***
aaa	o	x	-	-	x	x	
ooh	x	o	x	-	-	-	x
eeu	x	x	o	x	-	-	x
iii	x	-	x	o	x	-	-
ou	-	x	-	-	-	x	o
uu	-	x	x	-	-	-	x
Clic or “eke”	-	-	x	x	o	-	-
Stop or “ebe”	-	x	x	-	x	o	x

x: means that pronounced phoneme classified as an other

Table 3. Classification using LPCs, MFCC as Features and DTW , DE as Classifier

Vowel	**LPC DTW**	**MFCC DTW**	**LPC-DE**	**MFCC -DE**
aaa	76	81	65	72
ooh	58.33	62	54	63
eeu	57	59	51	56.5
iii	61	73	53	67
Clic or “eke”	54.55	79	49	61
Stop or “ebe”	55.56	81	48.5	65.5

Table 4. Classification using LPC,MFCCs and HMM as Classifier

Vowel	**LPC (%)**	**MFCC (%)**
aaa	85	92
ooh	78	83
eeu	74	84
iii	79	87
clic	87	94
stop	90	95

According to the results presented above (Tables: 3 and 4), the recognition rates using MFCCs parameterization classification with DTW or HMM classification is better than: LPCs and MFCCs with DTW. So we can say that the MFCCs / HMM system is partially independent of the speaker.

Results using MFCC and HMM, on German database vowels (sounds) for persons with chronic inflammation of the larynx and vocal fold nodules [19], are presented in Table 5.

Table 5. Classification using LPC and MFCC Using HMM for Voweld form German DB [18]

Vowel	LPC (%)	MFCC (%)
aaa	55	67
ooh	42	53
eeu	43	49
iii	53	72
Clic or “eke”	57	64
Stop or “ebe”	54	70

We can see that the recognition rate is little bit lower than for healthy persons, we conclude that in this case other special features might be necessary to include on the application.

In addition, we must consider the preprocessing for noise in future work, as well as the database training models need from the category of children.

7. Graphic User Interface (GUI) for Mouse Cursor Movements.

We then developed an easy to use graphic user interface under Matlab Version 6, the main elements are presented in Figure 5.

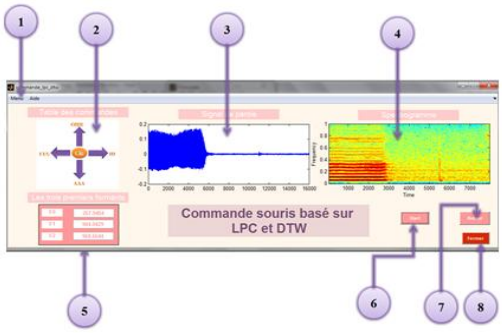

Figure 5. GUI used for test of the Application

1- Menu Tool-bar contiens 2 onglets « Menu » et « Aide »
2- A Command table for user , a guide to better selection of vocabulary.
3- Graphic zone to show wave signal form acquired from Mic.
4- Spectral presentation of acquired signal
5- « edit text » show fundamental frequency and 2 harmonics 6- « Push button» Start to start registration of audio signal.
6- «Push button» Retour allows to go back to main menu.
7- «Push button» Fermer allows to close the GUI LQWHUIDFH

8. Some Results on Application of GUI

User test, on how to activate web page by using our system based on just vowels.

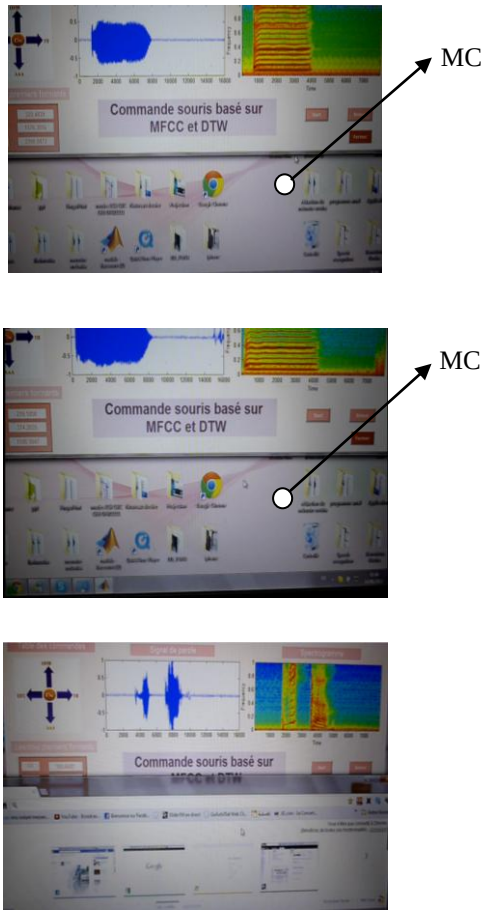

Figure 6. Illustration of Movement of Mouse Cursor (MC)

9. Conclusions

According to the results, we note that the classification using HMM is better than the DTW, and the decision based on MFCC coefficients is more certain than the coefficients LPCs.

From experimental results, it can be concluded that MFCC features and HMM as classifier can recognize the speech signal well. Where the highest recognition rate that can be achieved in the last scenario. This result is achieved by using MFCCs and HMM. Moreover, we need to get better features to improve classification of vowel and short words pronounced from voice disabled persons; in fact this can be resolved by inserting Jitter and Shimmer as features.

The GUI allowed us to make some tests by different users on combined methods, using PLC, MFCC and frequencies : F0, F1, f2 as features ; then

we used DE and DTW as Classifiers, the results are presented in tables.

We also presented the results of peoples with voice pathologies disorders, the signals were collected form Manfred Putzer & Jacques Koreman '' A german databse for a pattern for vacal fold vibration" in [18].

We notified that the variety of signals, collected for database from different age and gender, the recording conditions and the environment, have a considerable impact in classification results.

10. Acknowledgements

We would like to thank members of the laboratory of Automatic and Signals at Badji Mokhtar Annaba University and students of Electronics Department, for their contribution in experimental tests.

11. References

[1] Pride Mobility Products Corporation, "Pride mobility products group sip-n-puff system/head array control", 2005, http://pridemobility.com (Access date: 10 January 2015).

[2] Felzer T., Fischer R., Gr¨onsfelder T., and Nordmann R., "Alternative control system for operating a PC using intentional muscle contractions only," in *Online-Proc. CSUN Conf. 2005*, 2005.Available: http://www.csun.edu/cod/conf/2005/proceedings/2121.htm (Access Date: 20 February 2016).

[3] Felzer T. and Freisleben B., "HaWCoS: The "Hands-free" Wheelchair Control System," in *Proc. ASSETS 2002*. ACM Press, 2002, pp. 127– 134.

[4] Chin Control- Special Controls-647G371=D-1GB_08.06.pdf. Access date: 15 february 2016.

[5] Craig D. A. and Nguyen,H. T. "Wireless real-time head movement system using a personal digital assistant (PDA) for control of a power wheelchair," in *Proc. IEEE-EMBS 2005*, 2005, pp. 772–775.

[6] De-Mauro C., Gori M., Maggini M., and . Martinelli E, 'A voice device with an application-adapted protocol for Microsoft windows', In Proc. IEEE Int. Conf. on Multimedia Comp. and Systems, vol. 2, pp. 1015–1016, Firenze, Italy, 1999.

[7] Igarashi T. and Hughes J. F., 'Voice as sound: Using non-verbal voice input for interactive control', In ACM UIST 2001, November.

[8] Alex Olwal and Steven Feiner, 'Interaction techniques using prosodic features of speech and audio localization', In IUI '05: Proc. 10th Int. Conf. on Intelligent User Interfaces. New York: NY, USA, 2005. ACM Press, pp. 284– 286.

[9] Sargin,M.E. Aran O.,A. et All. , 'Combined Gesture-Speech Analysis and Speech Driven Gesture Synthesis,' ICME 2006 : IEEE International Conference on Multimedia and Expo, July 2006, pp: 893–896.

[10] Bilmes J., Li X., Malkin Jet all.'he vocal joystick: A voice-based human-computer interface for individuals with motor impairments', in Human Language Technology Conf. and Conf. on Empirical Methods in Natural Language Processing, Vancouver, October 2005.

[11] Harada S., Landay J., J. Malkin, X. Li, J. Bilmes, 'The Vocal Joystick: Evaluation of Voice-based Cursor Control Techniques', *ASSETS'06,* October 2006.

[12] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, 'Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques', Journal of Computing, Volume 2, Issue 3, March 2010, pp : 138-143.

[13] Mahdi Shaneh and Azizollah Taheri, '"Voice Command Recognition System Based on MFCC and VQ Algorithms' ,World Academy of Science, Engineering and Technology 57 2009, pp: 534-538.

[14] Bala A, Kumar A, Birla N - Anjali Bala et al., 'Voice command recognition system based on MFCC and DTW' International Journal of Engineering Science and Technology, Vol. 2 (12), 2010, pp :7335-7342.

[15] Thiang, S. Wijoyo, "Speech recognition using linear predictive coding and artificial neural network for controlling movement of mobile robot", International Conference on Information and Electronics Engineering IPCSIT vol.6, 2011, 179-183.

[16] Snani C., " conception d'un system dereconnaissance de mots isolés à base de l'approchestochastique en temps réel : Application commmande vocale d'une calculatrice ", Mémoire de magister ,Institut d'electronique univ. Badji mokhtar Annaba,2004.

[17] HAdri C., boughazi M and fezari M, "improvement of Arabic digits recognition rate based in the parameters choice", in proceedings of international conf. CISA Annaba, june 2008.

[18] Fezari M. and Al-dahoud A., 'An Approach For: Improving Voice Command processor Based On Better Features and Classifiers Selection,' pp. 1–5. The 13th International Arab Conference on Information Technology ACIT'2012 Dec.10-13 ,2012.

[19] Manfred Putzer & Jacques Koreman '' A german databse for a pattern for vacal fold vibration '' Phonus 3, Institute of Phonetics, University of the Saarland, 1997, 143-153.

[19] El Emary I.M. M.' Fezari M., Amara F.,'Towards Developing a Voice Pathologies Detection System', in Jouranl of Electronics and Comunication, 2014 Elsevier.

[20] Eamonn J. Keogh, Michael J. Pazzani, 'Derivative Dynamic Time Warping' In Proc. Of the 1st SIAM Int.Conf. on Data Mining (SDM- 2001).

[21] Sakoe H., Chiba S., 'Dynamic programming algorithm optimization for spoken word recognition'. IEEE Transaction on Acoustics, Speech and Signal Processing, Vol 26, NO1, pp. 43-49. February 1978.

[22] Huang et al "Dynamic Time Warping (DTW) for Single Word and Sentence Recognizers", Chapter 8.2.1; Waibel/Lee, Chapter 4, May 3, 2012.