Asian Journal of Information Technology 5 (2) : 139-144, 2006

Hybrid Technique to Enhance Voice Command for an Electric Wheelchair

Mohamed Fezari, Mounir Bousbia-salah and Mouldi Bedda
Department of Electronics Laboratory of Automatic and Signal,
Annaba Faculty of Engineering, Badji Mokhtar University,
Annaba BP.12, Annaba, 23000, Algeria

Abstarct: In this study we propose a simple approach to small vocabulary word recognition applied to control a wheelchair. The methodology adopted is based on hybrid techniques used in speech recognition which are zero crossing and extremes with dynamic time warping followed by a decision system based on independent methods test results. To test the approach on a real application, a PC interface was designed to control the movement of a wheelchair for a handicapped person by simple vocal messages. Tests showed that the decision system outputs follow a logic in line with the words uttered. The experiment aimed to find any problems for running tests in outdoor environments. Also, it was checked whether the basic commands used would be enough to control the vehicle. Some experiments are tested by actually running a powered-toy vehicle.

Key words: Speech recognition, hybrid methods, DTW, voice command

INTRODUCTION

Speech recognition has a key role in many application fields[1-4]. Various studies made in the last few years have given good results in both research and commercial applications[5-8]. This study proposes a new approach to the problem of the recognition of isolated words, using a set of traditional pattern recognition approaches and a decision system based on test results of classical methods[9-13] in order to increase the rate of recognition. The increase in complexity as compared to the use of only traditional approach is negligible, but the system achieves considerable improvement in the matching phase, thus facilitating the final decision and reducing the number of errors in decision taken by the voice command guided system.

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.

The recent sophisticated methods for intelligent signal processing have led to the development of various applications of hybrid techniques in the field of telecommunications as well as automatic speech recognition. 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: Crossing Zero and Extremes (EXZCR), Linear Dynamic Time Warping (DTW), linear predictive coefficient (LPC) parameters , Parameters correlation (Parcor), and cepstral coefficients. This study is part of a specific application concerning system control by simple voice commands. The application uses six commands in Arabic words. It has to be implemented on a DSP[14] and has to be robust to any background noise confronted by the system. Thus a simple approach is a better choice[5,7].

The aim of this study is therefore the recognition of isolated words from a limited vocabulary in the presence of background noise. The application is speaker-dependent. It should, however, be pointed out that this limit does not depend on the overall approach but only on the method with which the reference patterns were chosen. So by leaving the approach unaltered and choosing the reference patterns appropriately, this application can be made speaker-independent[8].

As application, a vocal command for a handicapped person wheelchair (HPWC) is chosen. A wheelchair is an important vehicle for physically handicapped persons. However, for the injuries who suffer from spasms and paralysis of extremities, the joystick is a useless device as a manipulating tool. It therefore involves the use of voice or head to control the movement of the wheelchair. Voice command needs the recognition of isolated words from a limited vocabulary used in AGV (Automatic Guided Vehicle) system[15-22].

Corresponding Author: Mohamed Fezari, Department of Electronics Laboratory of Automatic and Signal, Annaba Faculty of Engineering, Badji Mokhtar University, Annaba BP.12, Annaba, 23000, Algeria

DESIGNED APPLICATION DESCRIPTION

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

The Handicapped Person Wheelchair (HPWC) specifications are limited to six commands which are necessary to control the movement of the wheelchair: switching on and off the engine 'motor', forward movement, backward movement, stop, turn left and turn right. The vocabulary chosen to control the system contains a total of six words. The number of words in the vocabulary was kept to a minimum both to make the application simpler and easier for the user.

The six words in the Arabic vocabulary are the following:

- "Mouharek": This switches the motor on or off at average speed.
- "Ameme": This makes the movement upward.
- "Wara": This makes the movement backward.
- "Kif": This command stops the movement.
- "Yamine": This makes the turn right;
- "Yassar": This makes the turn left.

The system is by nature in movement within the wheelchair. Thus it is affected by external noise. In designing the application, account was taken to reduce the affecting noise on the system at various movements. To do so, the external noise was recorded and spectral analysis was performed to study how to limit its effects in the recognition phase[7]. This is just done within the experience area.

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

SPEECH RECOGNITION SYSTEM BLOCKS

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

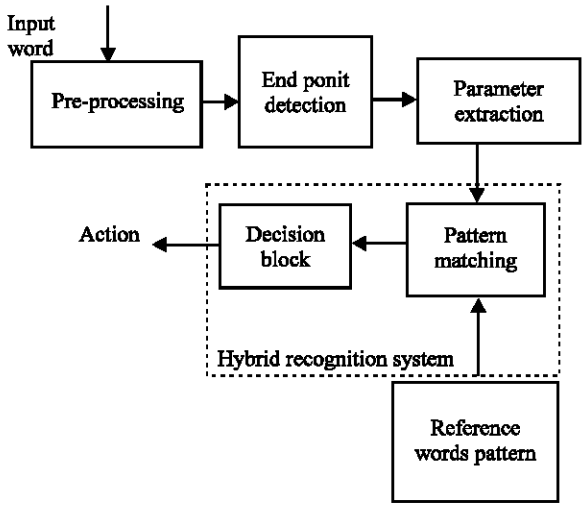

Fig. 1: Hybrid recognition system block diagram.

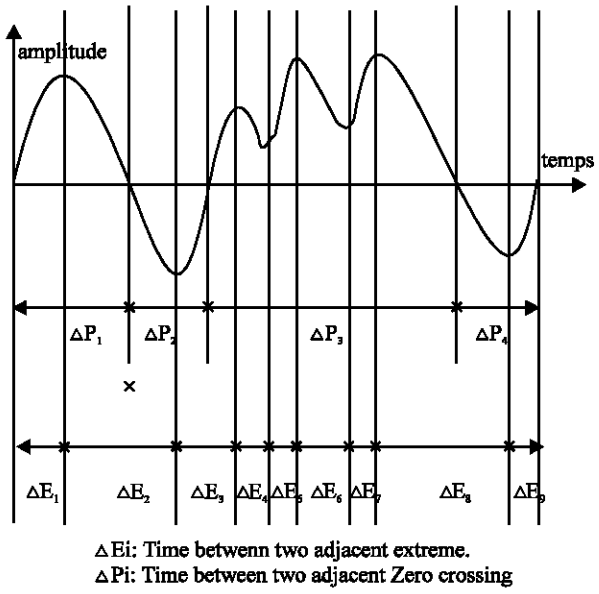

Fig. 2: Extremun and zero crossing

To eliminate the motor sound when it is on, the sound signal is recorded and then it is suppressed from the composite signal which includes the command word and motor sound. This makes it possible to increase the signal-to-noise ratio, with obvious benefits for the application.

The End Point Detector (EDP) 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 generated by engine. Its output is a vector of samples of the word (i.e.: those included between the endpoints detected).

The End Point Detector implemented bases detection on analysis of crossing zero points and energy of the signal[4], the linear prediction mean square error computation helps in limiting the beginning and the end of a word; this makes it computationally quite simple. Simulations have also shown that it performs well in typical use situations.

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 10-mili seconds frames, at 100 samples per frame. Five types of parameters are extracted: Normalized Extremes Rate with Normalized Zero Crossing Rate (EXZCR) (Fig. 2), linear DTW with Euclidian distance (DTWE), LPC coefficients (Ai), Parcor Coefficients (Ki) and cepstral parameters (Ci)[1].

These parameters were chosen for computational simplicity reasons (EXZCR), robustness to background noise (10 Cepstral and Parcor parameters) and robustness to speaker rhythm variation (DTWE). The parameter extraction and ordering tool made the task simpler and more efficient. In addition, calculation of the cepstral parameters does not create an additional computational load, because they are obtained from the autocorrelation coefficients previously calculated by the Parcor parameters.

The reference pattern block is created during the training phase of the application, where the user is asked to enter five times each command word. For each word and based on the five repetition, five sets of parameters are extracted and stored.

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.

HYBRID RECOGNITION SYSTEM

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. Fig. 3 shows the elements making up the main blocks for the hybrid recognition system (HRS). The input of HRS block is a set of five values representing the parameters of the input word obtained from the five methods. The Hybrid Recognition block compares the input parameters with the references parameters of the six words. It then generates a vector of five values; the elements are the recognized word number. If the input word is "ameme" and the EXZCR method recognizes the word then it generates the number 2 (which means de second word).

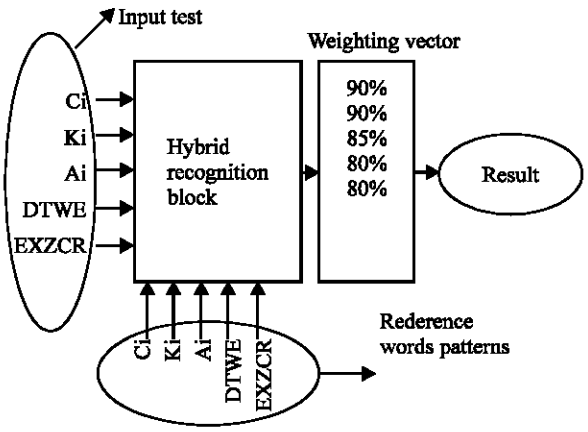

Fig. 3: HRS block structure

However it may recognize another word from the used vocabulary or none of them, in this case it generates either the number of the recognized word or a Zero.

The last block in the HRS is the weighting block. It is made of a threshold system which discriminates which of the six words was pronounced by the user on the basis of the information provided by the HRS. The resulting vector is multiplied by the weighting block that contains five rates. These rates have been fixed based on results given by each method. The best word to fit the input word is chosen.

INTERFACE CARD FOR IMULATION

A parallel port interface was designed to show the real-time commands. It is based on two TTL 74HCT245 buffers and 16 light emitting diodes (LED), six green LED to indicate each recognized command and a red LED to indicate wrong or no recognized word. The other LED's were added for future insertion of new command word in the vocabulary example the words: "Faster", "slower" and "light" as shown in Fig. 4.

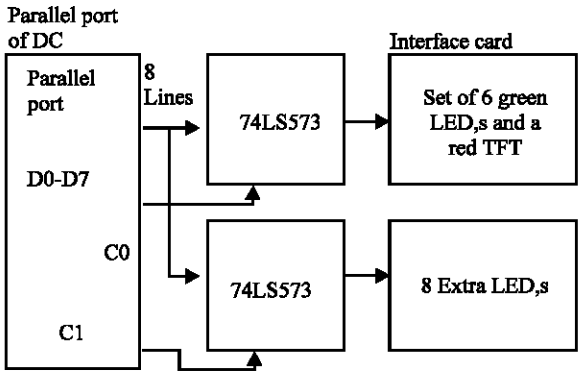

Fig. 4 : Parallel port interface

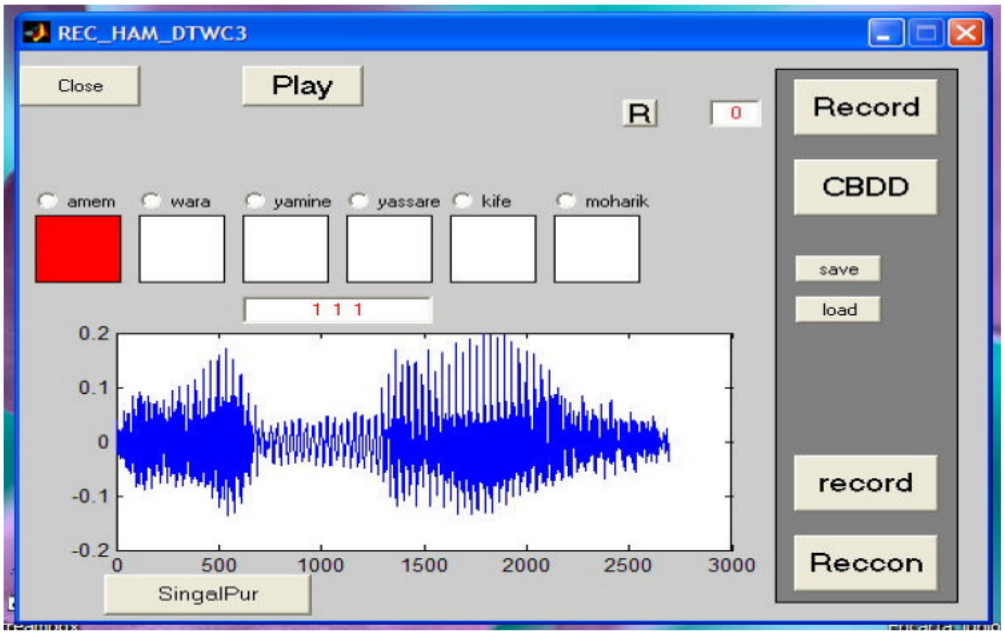

Fig. 5: GUI of the application.

The application was developed with MATLAB language under windows XP as operating system.

The graphic user interface (GUI), where the training and recognition phases are presented in the right, the six LED for the six words to recognize are presented by squares, the red square is the recognized word and the pronounced0 word is plotted in blue as shown in Fig. 5.

The training phase is the first step in which the database is created. In this phase, the speaker repeats ten times each word, the parameters are produced and saved in a file.

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 corresponding box on the GUI is changed to red colour.

RESULTS OF SIMULATION

First, some tests were done on each method and the rate of recognition was registered as shown on Fig. 6.a. In the EXZCR method, the rate is lower; however the rate is speaker dependant. DTWE gives better results specially if there is any distortion in locution rhythm. LPC, Parcor and Cepstral coefficients give better rates than those cited earlier. And based on the tests for each method separately, we fixed the weighting vector.

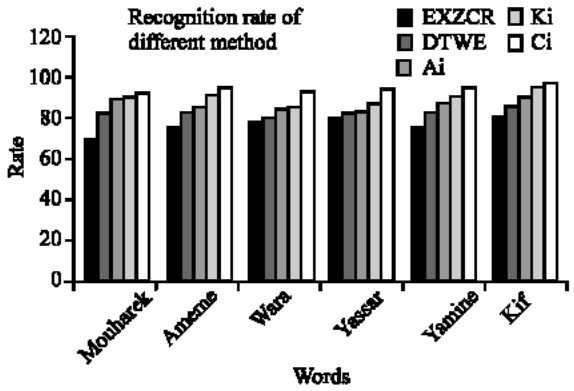

Fig. 6a: Recognition rate for the six command words using different methods.

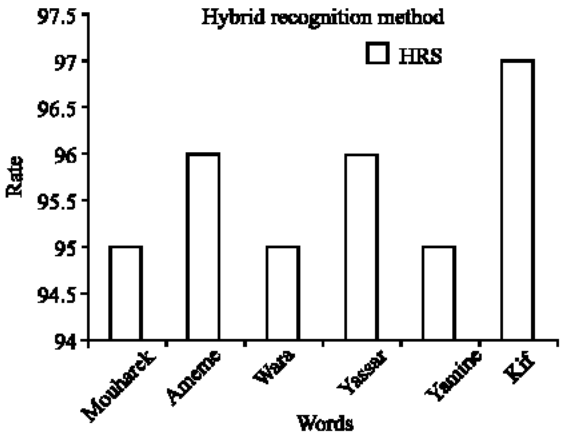

Fig. 6b: Recognition rate for the six command words using Hybrid method.

The second tests were made using the hybrid method, the recognition rate is improved as it is shown if Fig. 6b.

CONCLUSION

This study has presented a new methodological approach to the implementation of an isolated word recognition system based on a mixture of techniques used in speech recognition. This can be implemented easily within a DSP or a CMOS RISC microcontroller.

The use of hybrid technique based on classical recognition methods makes it easier to separate the class represented by the various words, thus simplifying the task of the final decision block. Tests carried out have shown an improvement in performance, in terms of misclassification of the words pronounced by the user. The increase in computational 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. The idea can be implemented easily within a hybrid design using a DSP with a microcontroller since it does not need too much memory capacity and the system need to be portable , thus not bulky and easy to control. Advances have been made in the technology of smart wheelchairs during the development of the HPWC. Performance of the HPWC has demonstrated its potential as an effective approach to providing independent mobility to a wide range of users who can not independently operate a powered wheelchair system. This speech command system can be enhanced by eliminating other kind of outdoor environment noises. The future works in the improvement of the HPWC will allow for different operating levels ranging from simple obstacle avoidance to fully autonomous navigation. Additionally, the HPWC will provide a means for development and testing of shared control methods, where a human operator and machine share control of a system. The reinforcement learning algorithms will be investigated. Results from research and development efforts in this area should have application to a broad range of assistive technology systems. Finally 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 robot movements or robot-hand control.

REFERENCES

1. Suhm, B., B. Myers and A. Waibel, 2001. Multimodal error correction for speech user interfaces.,ACM Transactions on Computer-Human Interaction, 8: 60-98.
2. Nishimoto, T., 1993. Improving human interface in drawing tool using speech, mouse and Key-board, Proceedings of the 4th IEEE International Workshop on Robot and Human Communication, Tokyo, Japan, pp: 107-112.
3. Ono, Y., H. Uchiyama and W. Potter, 2004. A mobile robot for corridor navigation: A multi-agent approach, ACM Southeast Regional Conference archive Proceedings of the 42nd annual Southeast regional conference table of contents Huntsville, Alabama, pp: 379 - 384.
4. Shlomot, E., V. Cuperman and A. Gersho, 1997. Hybrid coding of speech at 4 Kbps, Proc. IEEE Workshop on Speech Coding, Pocono Manor, PA,. pp: 37-38.
5. Hagen, A., A. Morris and H. Bourlard, 1990. Different weighting schemes in the full combination sub-bands approach in noise robust ASR, In: Proceedings of the ESCA Workshop on Robust Methods for Speech Recognition in Adverse Conditions, pp: 199-202.
6. Godin, C. and P. Lockwood, 1989. DTW schemes for continuous speech recognition: A unified view, Computer Speech and Language, 3: 169-198.
7. Barker, J., M. Cooke and P. Green, 2001. Robust ASR based on clean speech models: An evaluation of missing data techniques for connected digit recognition in noise, In Proc. of 7th European Conference on Speech Communication Technology (EUROSPEECH-01), 1: 213-216.
8. Wang, T. and V. Cuperman, 1998. Robust Voicing Estimation with Dynamic Time Warping, Proceedings IEEE ICASSP, pp: 533-536.
9. Farrell, K.R. and R.P. Rama, 1998. An analysis of data Fusion methods for speaker verification, IEEE Proc. ICASSP, pp: 1129-1132.
10. Glotin, H. and F. Berthommier, 2000. Test of several external posterior weighting functions fur multiband full combination ASR, In: ICSLP, pp: 333-336.
11. Fioretti S., T. Leo and S. Longhi, 2000. A navigation system for increasing the autonomy and security of powered wheelchair, IEEE Trans. On Rehabilitation Engineering, 8: 490-498.
12. Byrne, W., P. Beyerlein and J.M. Huerta, 2000. Towards Language Independent Acoustic Modeling, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Istanbul, pp: 1029-1032.

13. Suontausta, J. and J. Hakkinen, 2000. Decision tree based text-to-phoneme mapping for speech recognition, Proceedings of the Intl. Conf. Spoken Lang. Processing, Beijing, pp: 199-202.
14. Gaici´c, A., M. Puschel and J.M.F. Moura, 2003. Fast Automatic Implementation of FIR Filters, In IEEE Proc. ICASSP, 2: 541-544.
15. Brent, M.R. and J.M. Siskind, 2001. The role of exposure to isolated words in early vocabulary development, Cognition, 81: 33-44.
16. Parikh, S.P., R.S. Rao, S.H. Jung and V. Kumar, 2003. Human robot interaction and usability studies for a smart wheelchair, Proc. of IEEE/RSJ, International Conference on Robots and Systems. Las Vegas, Nevada, pp: 3206-3211.
17. Bourhis, G. and Y. Agostini, 1998. The Vahm robotized Wheelchair: System architecture and human-machine interaction. J. Intelligent and Robotic Syst., 22: 39-50.
18. Perzanowski, D., A. Schultz, W. Adams and E. Marsh, 2000. Using a natural language and gesture interface for unmanned vehicles, in the Proc. of Unmanned Ground Vehicles II, Aerosense, pp: 341-347.
19. Cooper, Rory A., 1999. Engineering manual and electric powered wheelchairs, Critical Reviews in Biomedical Engineering, 27: 27-73.
20. Graf, B., 2001. Reactive Navigation of an Intelligent Robotic Walking Aid, In proceeding of Roman, pp: 353-358.
21. Fong, T., I. Nourbakhsh and K. Dautenhahn, 2003. A Survey of socially interactive robots., Robotics and Autonomous Systems, 42: 143-166.
22. Oviatt, S.L. and P.R. Cohen, 2000. Multimodal interfaces that process what comes naturally, Communication of ACM, 43: 45-53.