

Mouse Cursor's Movements using Voice Controlled Mouse Pointer

Sarita
Department of Computer
Science and Engineering
Lovely Professional University
Phagwara, India

Kiran Kumar Kaki
Department of Computer
Science and Engineering
Lovely Professional University
Phagwara, India

ABSTRACT

In this era, doing everything fast is the main requirement. Computers play a vital role to perform various tasks with more speed. For this reason this paper presented a system called as "Voice Controlled Mouse Pointer" (VCMP), which enables the users to perform computer's operations with their voice. VCMP will provide mouse cursors movement with voice in eight directions, which will be helpful for the people with disabilities. VCMP uses Linear Predictive Coding (LPC) for speech features extractions and features matching will be done by calculating minimum distance between the extracted features of voice and stored features in the features library. VCMP does not require the user's voice database; it extracts the voice features online during training and stores them into the features library. This makes VCMP speaker independent. VCMP will provide instantaneous mouse cursor's movement on the screen as soon as speech is recognized by the system. VCMP also provides voice commanding, where user can command the computer to open various operating systems' programs with voice.

Keywords

Human computer interaction, Speech recognition, Linear Predictive Coding, Artificial Neural Network, Voice controlled mouse pointer

1. INTRODUCTION

Human computer interaction is a field which focuses on providing a means of interaction between humans and computers [2]. Controlling the mouse pointer is one of the best ways to provide a meaningful interaction. Human computer interaction is not only evolved in the area of mouse control using voice or in simple words in the area of speech recognition, but also in other areas as:- Face recognition and Gesture recognition. All these areas have their own pros and cons but speech recognition is cheap and best way to make interactions and it is helpful for the people with disabilities [20]. This is not good only for the people who can't move their hand or body but also for the people who have vision impairments. The thing that is missing in current speech-based technology is to rapidly and incrementally execute operations using spoken commands whose impact on the object of interest is immediately visible. To operate entire computer through speech several things are required

- Textual information: ability to input text entries accurately and fast.
- Giving Commands: ability to execute all commands.

- Direct manipulation: ability to manipulate objects and perform mouse-like operations fluidly [10].

First two functions have been supported through advancements in the field of automated speech recognition (ASR) [10]. Text entry and commanding is supported by two main commercial products, Windows Speech Recognizer and Dragon Naturally Speaking, but there remains operating system's significant portion and functionalities that are still not accessible using such commands. Direct manipulation means rapidly executing operations, is still virtually unaddressed. The standard spoken commands are ideal for discrete but not good for continuous operations. For example, use discrete command based system to move the cursor from bottom-left to the upper right then a number of "up" and "right" commands are needed, which is inefficient. None of the method till now takes advantage of full continuous nature of the human voice. There are various limitations of traditional systems:

- Traditional voice based system are speaker dependent, they need to store user's voices in the database before using the system.
- Recognition capability of the traditional systems is very low.
- Existing automatic speech recognizer (ASR) systems are not perfectly robust to various speaking conditions, accented speaker, and noise, so they are still not universally adopted as a dominant human computer interface.

Using above mentioned criteria we demonstrate our thesis as:

"Non speech vocal characteristics can be used alone and in combination with other input methods for people with motor impairments to control mouse pointer effectively."

System can be commanded to do some operations without using mouse with hands. Here we will focus on hands free computing which is useful for the disable people. By doing this people with disabilities can use the computer systems normally as the normal users use it. Current systems work well when they are operated by one user only, in other words they are speaker dependent. But VCMP is speaker independent and it can be embedded in any of the application because it does not require the predefined database, it makes the database online during the training phase. It is not like that voice based interaction is best among all the alternatives of hands free computing but main point is that voice can be used as the supplement to many of the devices. Voice

based interaction because it does not require any costly hardware setup, only a good quality microphone is required whose costs is much less than the hardware used in other methods of the hands free computing. Also in human computer interaction use of the speech characteristics is still unexploited.

2. PROBLEM FORMULATION

VCMP is a system which will take human voice as input and further this feature can be used to perform computer operations such as moving mouse pointer diagonally (top-left, bottom-left, top right, bottom right) and in the four directions (Left, right, up, down).VCMP will also allow the user to open various files and to perform operating system's operations for example open notepad, open paint, open calculator etc. VCMP will provide online training which results in the speaker independent system, so VCMP can be embedded in any of the application.

3. OBJECTIVES

The objective of this research is to develop VCMP, which works on the human voice and perform hands free mouse pointer movement and various operating system's operations. Its main objectives are:

1) Movement in Four Directions

Mouse pointer should start moving in the left, right, up, down as soon as VCMP recognizes the words.

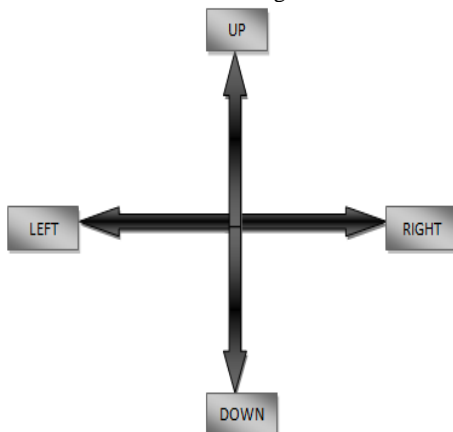


Figure 1: Mouse Pointer Movement in four directions

2) Movement in Four Diagonals

Mouse pointer should start moving in the Top-left, Top-right, Bottom-left, Bottom-Right as soon as VCMP recognizes the words.

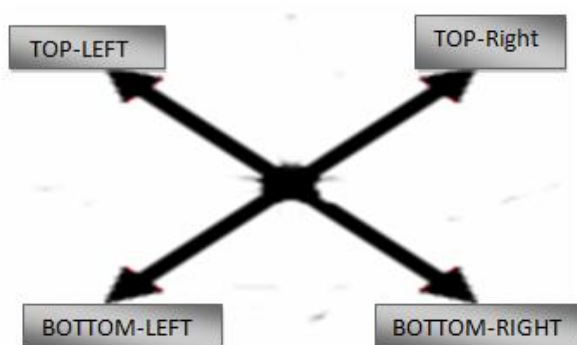


Figure 2: Mouse Pointer Movement in diagonals

3) Make more efficient human computer interaction for people with disabilities.

Mainly our system will be for motor impaired users, who can't access the computer systems as the normal users do. So here, developed system can be efficiently used by the abnormal users. Hands free computing will be provided with speech recognition system.

4) Make more efficient human computer interaction for general users.

This system will not be beneficial for only people with disabilities of any kind, but our goal will be to make current human computer interaction system more usable and user friendly so that normal user can perform work as fast as possible. Some speech recognition systems are also available but our system will provide enhanced features available in the current systems.

5) Performing Operations

VCMP also provides functionality to perform various operations. User needs to speak only the simple commands and corresponding task will be performed. VCMP will open the notepad using the 'notepad' command, user needs to speak notepad whenever he/she wants to open notepad, same like this other program like 'vlc', 'paint' & 'calculator' should also open using human voice.

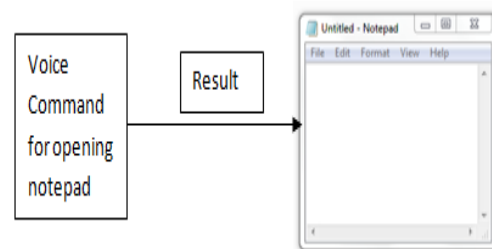


Figure 3: Opening notepad using 'Notepad' command

4. SIGNIFICANCE

There are various features in the Voice Controlled Mouse Pointer that makes it different from other existing systems. These all features are given below:

1. Voice Controlled Mouse Pointer (VCMP) is a speaker independent system; it does not require any kind of voice database. Features matrix will be calculated online whenever user wants to use the system.
2. VCMP is based on the word recognition, which is very much correct in VCMP than the other systems. 80% of the time recognition of the system is correct, rest 20% is due to the noise, system generated sounds etc.
3. Costly equipments are not required by VCMP, it requires a good quality microphone only.
4. This system is speaker independent so it can be used anywhere and also can be embedded in other applications also.
5. VCMP can be used by any kind of impaired users because only the ability to vocalization is required.
6. VCMP is highly interactive so it's working is very easy and can be easily understood by the normal computer users.
7. No supporting files are required so it is platform independent and can be use anywhere.

8. In future new module of continuous recognition can be added to it to perform mouse cursor movements with more speed and to perform the operations which need the continuous input for example drag, scroll, zoom etc. All these above features make VCMP a highly interactive and efficient system to perform mouse cursor movements and various operating systems' operations.

5. METHODOLOGY

VCMP will perform various steps to process the voice, output of the previous step becomes the input for the next. These steps are given below:

Step 1: Speech Input

Step2: Feature Extraction

Step3: Pattern Recognition

Step4: Motion Control

Step5: Mouse Movements or Operations

STEP1: Speech Input

This is the very first step of VCMP which will take human voice as input. This step does not require any costly hardware; it only requires a good quality microphone.

STEP2: Feature extraction

Goal of this step is to extract the features of the human voice; In VCMP Linear Predictive Coding (LPC) has been used to extract the low level acoustic features. For feature extraction given parameters are used:

- **Sampling Frequency**
Fs=10,000 Hz
Speech signal is sampled at this frequency
- **Frame Size**
For Energy=25ms
For Pitch=40ms
- **Frame Step**
For Energy=10ms
For Pitch=10ms

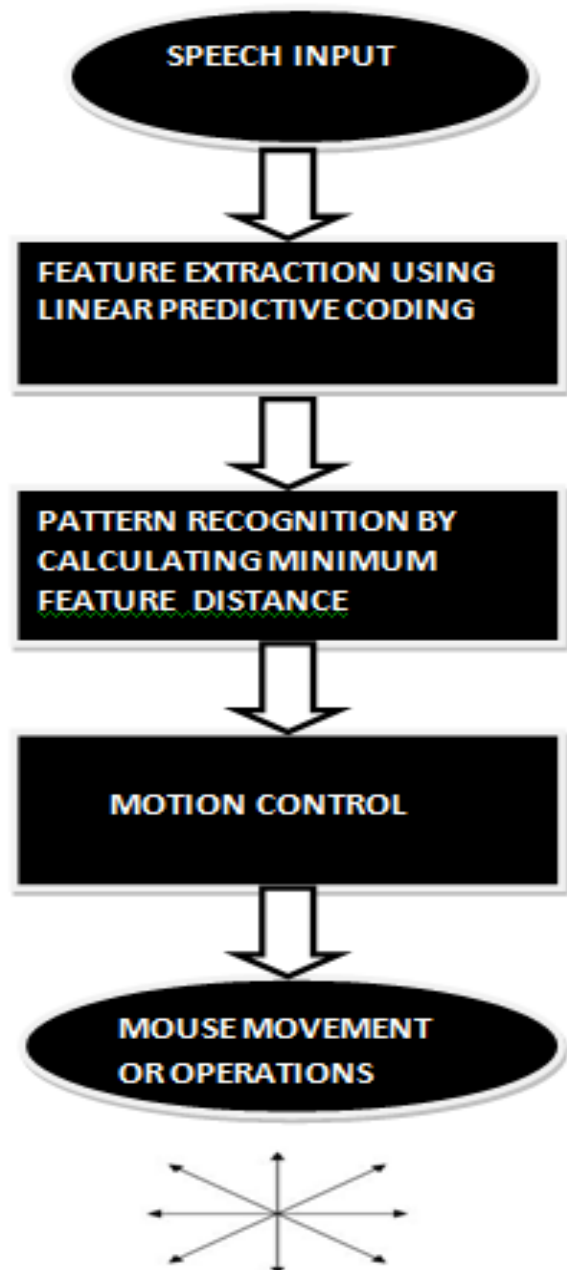


Figure 4: Flow Chart of VCMP

To implement the feature extraction step, VCMP will process the word taken as an input in the first step. VCMP will compute the features of the word and then store these words in the feature matrix. This feature matrix is two- dimensional consisting of 4 rows (for up, down, left, right) and 17 columns. User will have 1 second to speak each word, further the generated library file can be used for future use if the user is same.

STEP 3: Pattern recognition

Extracted features which are the output of the feature extraction step become the input for the pattern recognition step. At this step features stored in the feature matrix F_w during training are matched with the features of the currently spoken word during recognition.

This matching is done on the basis of the minimum distance between the stored features in the library file and the features of the spoken word. For example features of LEFT, RIGHT, UP, DOWN will be stored in the feature matrix during training, during recognition process when user utters a word let suppose it is LEFT then these features of LEFT will be matched with the features of other four words(LEFT,RIGHT,UP,DOWN) already stores in the feature matrix. Then distance will be calculated between LEFT-LEFT, LEFT-UP, LEFT-DOWN, LEFT-RIGHT. Minimum distance will surely be between LEFT-LEFT, so the mouse pointer will start moving the left direction.

STEP 4: Motion Control

Motion control module will use the patterns recognized by the patterns recognition module and will process them for mouse 2-D movement. Various motion parameters will be used for mouse movements which are given below:-

- XY direction
- Speed
- Acceleration

These all above parameters will be decided on the basis of the speech characteristics such as pitch, loudness and vowel quality.

STEP 5: Mouse Movement

VCMP will perform mouse movement in any of the four directions and diagonals. This movement will be based on the recognized word and all the parameters estimated by the motion control module.

6. RESULTS AND DISCUSSIONS

VCMP is implemented in MATLAB. There are mainly two modules of the system:

- Module 1: Perform Mouse Cursor Movements
- Module 2: Perform

Module 1:

Voice Controlled mouse pointer starts with the interface as shown in figure 5. Users have two options;

- Movement in Four Directions
- Movement in Diagonals

User can choose any of the option.

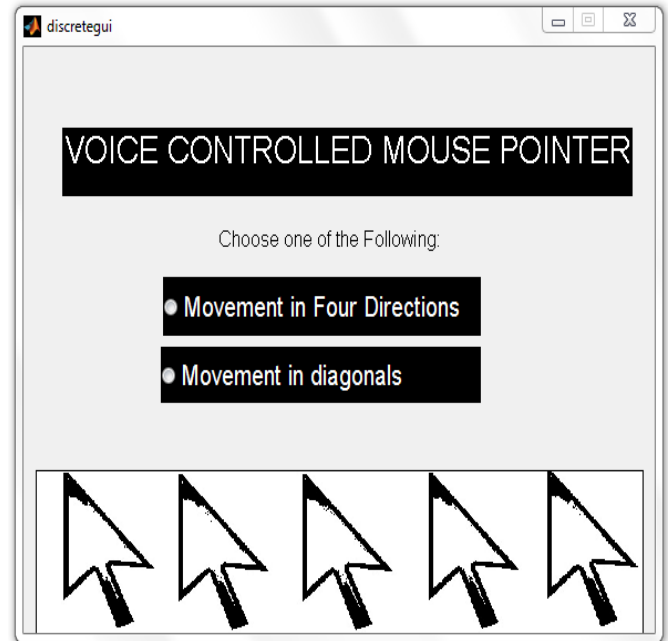


Figure 5: User interface of VCMP



Figure 6: UI when user wants to do cursor movement

EXAMPLE: Spoken Word-RIGHT

Steps of VCMP:

- 1) User will have one second to utter the word.
- 2) User has to press enter and utter "RIGHT" in one second.
- 3) Features of "RIGHT" are computed as shown in figure 7 and then VCMP will compare these features with feature of UP, DOWN, LEFT, RIGHT that are already stored in feature matrix during the training phase.
- 4) Voice Controlled Mouse Pointers will compute distances on the basis of minimum distance technique.
- 5) Calculated distances:-

- a. UP→D1=18.5631
 - b. DOWN→D2=18.3756
 - c. LEFT→D3=11.2962
 - d. RIGHT→D4=11.0180
- 6) Here minimum distance is of word right so RIGHT will be the word spoken by user.
- 7) Output on console window of MATLAB will look like “Word you spoke: right”. FFT of word right is shown in figure 8.
- 8) Then at last mouse pointer will start moving in the left direction as shown in 9.

```
You will have 1 second to say 1 word.  
Press enter and speak immediately:  
  
a =  
  
Columns 1 through 12  
    1.0000   -2.4451    3.0525   -3.2944    4.1793   -4.5557    3.8121   -3.2445    3.0809   -2.3260    1.3405   -1.0716  
  
Columns 13 through 17  
    0.8248   -0.3419    0.1759   16.9189    0.0013  
  
d =  
  
    18.5631    18.3756    11.2962    11.0180  
  
Word you spoke: right
```

Figure 7: Features of the spoken word “RIGHT” & Distance

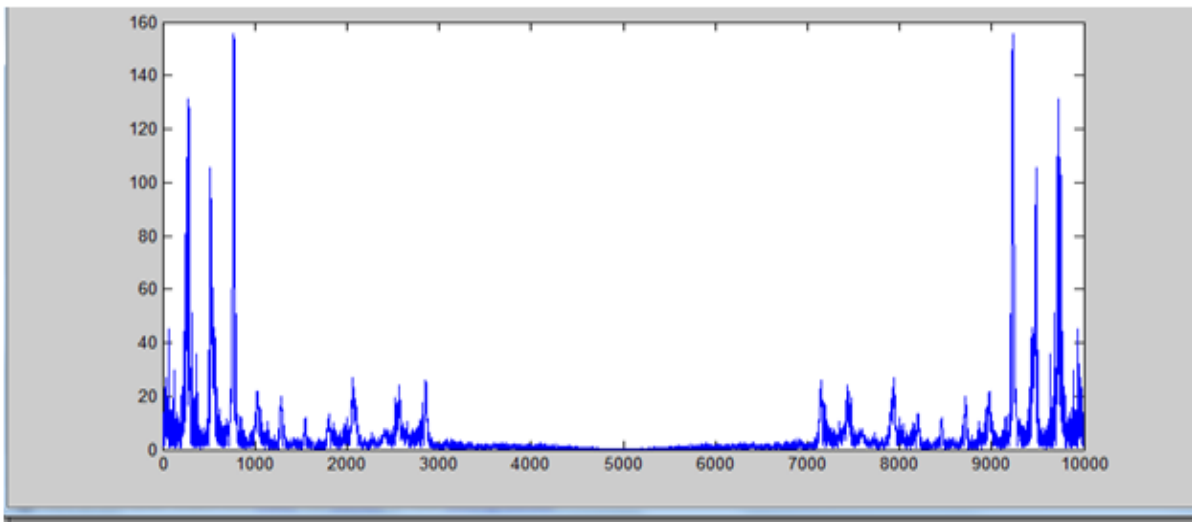


Figure 8: FFT of word “RIGHT”

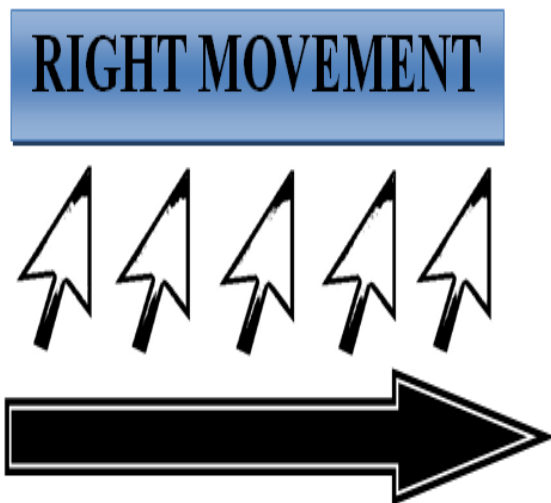


Figure 9: Movement in Right

Module 2:

Voice Controlled Mouse Pointer will move the mouse pointer instantaneously as soon as user utters a word. This will minimize the time to perform various operations because voice input can act as an extra input modality in combination with the standard input devices like mouse and keyboard. To perform operations users have to use the module 2. Its interface is given below:

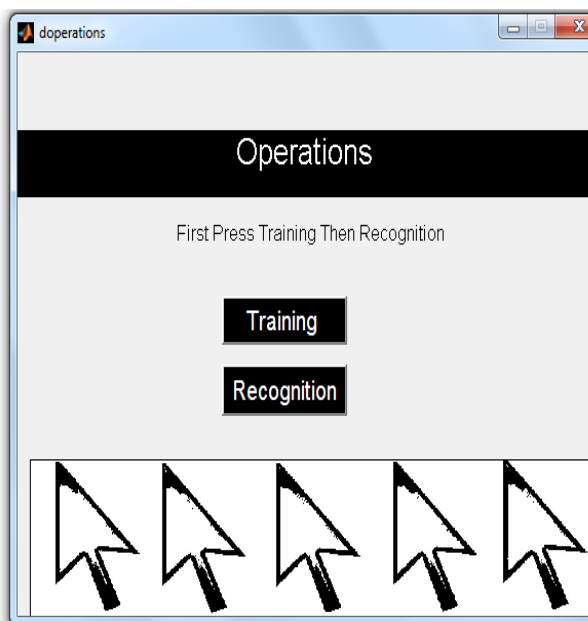


Figure 10: UI to do operations

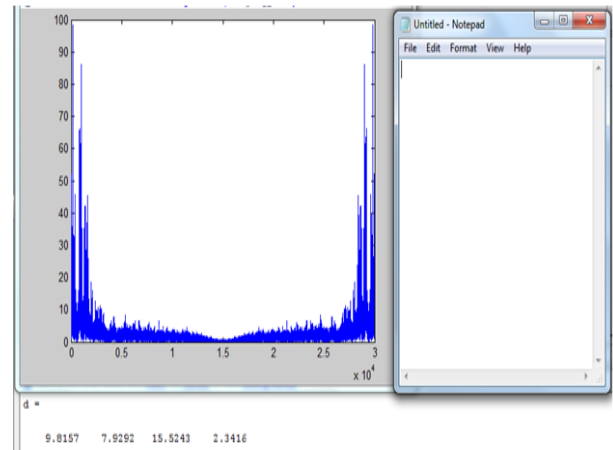


Figure 11: FFT of “notepad” command

7. CONCLUSION AND FUTURE SCOPE

7.1 Conclusion

- **VCMP is User Independent**
 VCMP does not require the predefined database of the user’s voices; it creates the sound’s features library during the training phase. This feature of VCMP is the best feature because due to the user independence this system can be embedded into a number of applications. Only five minutes training is required and user is finished with the work. Features matrix will be calculated online whenever user wants to use the system. Its two different phases of training and recognition allows the users to store the features of the word during training and recognize the spoken word using the minimum feature distance algorithm.
- **VCMP is user friendly**
 Voice Controlled Mouse pointer is much interactive, it can be used easily by the normal computer users. It does not require learning some extra thing to operate the VCMP. VCMP’s graphical user interface is easy to understand. People who have little bit computer’s knowledge will be able to use it efficiently.
- **Fast and correct Recognition**
 Recognition efficiency of Voice Controlled Mouse Pointer is much better than the existing systems. VCMP is based on the word recognition, which is very much correct in VCMP than the other systems. 80% of the time recognition of the system is correct, rest 20% is due to the noise, system generated sounds etc.
- **VCMP is cheap**
 VCMP does not requires extra hardware as required in the gesture recognition, tooth click control, eye gaze system etc. It requires only a good quality microphone. So every human being can use this system.

7.2 Future Scope

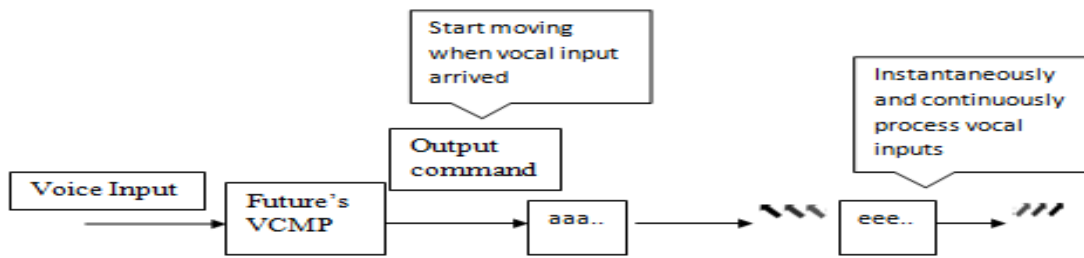


Fig 12: General Idea of the Future's system

- **Continuous Speech Recognition**

Implementation of this system can be continued further by using combination of discrete and continuous commands, which will help to minimize more time to perform operations. VCMP works on the discrete sounds which are efficient for entering information in the form of text or commanding computer. But when user wants to do the operations which needs the continuous operations then discrete command are not efficient. These operations are:

- Double click
- Dragging
- Picking file from one folder and drop to another
- Continuous path following(e.g., in games and drawing)
- Scrolling Etc

Using Continuous speech or vowel sounds mouse pointer movement will start immediately as soon as vowel will be recognized by the user. In the VCMP when entire word is recognized then mouse cursor start movement but in future using continuous sounds mouse will start moving as soon as some vocal input arrives and it will not stops when the vocal input or its characteristics change. The general idea of this type of system is given in figure 12.

- **Handling false Positive**

When user does not want to do recognition but unwanted speech or some system generated sounds results into the executions of unwanted mouse movements or other operations, these types of unwanted operations are called as false positives. Our system is not 100% efficient, external noise or system generated noise results into the unwanted executions. This problem is not new in the speech recognition, because there is not any of the system which 100% correctly recognizes the speech. But this problem can be lower down to certain extent by using filtered human voice. This is one the area where VCMP requires the improvements.

8. REFERENCES

- [1] Abdeen, M., Mohammad, H., & Yagoub, M. C. E., (2008) "An architecture for multi-lingual hands-free desktop control system for PC windows". In Electrical and Computer Engineering, CCECE Canadian Conference on IEEE, pp. 001747-001750.
- [2] Abu Shariah, M. A. M., Aion, R. N., Zainuddin, R., & Khalifa, O. O., (2007) "Human computer interaction using isolated-words speech recognition technology", In Intelligent and Advanced Systems, ICIAS, International Conference, pp. 1173-1178, IEEE.
- [3] Bendjima, B., Rivenq, A., Iyad, D., & Rouvaen, J. M., (1999) "Speech enhancement for vehicle hands-free mobile telephony", In Vehicular Technology Conference, 1999.VTC 1999-Fal., IEEE VTS 50th (Vol. 4, pp. 2168-2172), IEEE.
- [4] Bilmes, J. A., Li, X., Malkin, J., Kilanski, K., Wright, R., Kirchhoff, K., ... & Chizeck, H., (2005) "The Vocal Joystick: A voice-based human-computer interface for individuals with motor impairments", In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 995-1002..
- [5] Dam, H. Q., Nordholm, S., Dam, H. H., & Low, S. Y., (2005) "Adaptive beamformer for hands-free communication system in noisy environments", InCircuits and Systems, ISCAS 2005, IEEE International Symposium, pp. 856-859, IEEE.
- [6] Frank Loewenich and Frederic Maire, (2008) "Motion-Tracking and Speech Recognition for Hands-Free Mouse-Pointer Manipulation", Queensland University of Technology Australia, ISBN: 978-953-7619-29-9, pp. 550.
- [7] Gauvain, J. L., & Lamel, L., (2000) "Large-vocabulary continuous speech recognition: advances and applications", Proceedings of the IEEE, 88(8), pp.1181-1200.
- [8] Gomez, R., Kawahara, T., & Nakadai, K., (2010) "Robust hands-free Automatic Speech Recognition for human-machine interaction", In Humanoid Robots (Humanoids),10th IEEE-RAS International Conference, pp. 138-143.
- [9] Herbordt, W., Horiuchi, T., Fujimoto, M., Jitsuhiro, T., & Nakamura, S., (2005), "Hands-free speech recognition and communication on PDAs using microphone array technology", In *Automatic Speech Recognition and Understanding, IEEE Workshop*, pp. 302-307, IEEE.
- [10] Jeff A. Bilmes, Jonathan Malkin, Xiao Li, Susumu Harada, Kelley Kilanski, KatrinKirchhoff, Richard Wright, AmarnagSubramanya, James A. Landay, Patricia Dowden and Howard Chizeck, (2006), "The vocal Joystick", in *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, pp. 625 – 628.

- [11] Kumar, R., & Singh, M., (2011), “Spoken Isolated Word Recognition of Punjabi Language Using Dynamic Time Warp Technique”, *In Information Systems for Indian Languages, Springer Berlin Heidelberg*, pp. 301-301.
- [12] Liu, J., & Kavakli, M., (2010), “A survey of speech-hand gesture recognition for the development of multimodal interfaces in computer games”, *In Multimedia and Expo (ICME), IEEE International Conference*, pp. 1564-1569.
- [13] Maskeliunas, R., Ratkevicius, K., & Rudzionis, V., (2011), “Voice-based Human-Machine Interaction Modeling for Automated Information Services”, *Electronics and Electrical Engineering*, 110(4), pp.109-112.
- [14] Mellor, B. A., Baber, C., & Tunley, C., (1996), “Evaluating automatic speech recognition as a component of a multi-input device human-computer interface”. *In Spoken Language, ICSLP 96.Proceedings., Fourth International Conference*, Vol. 3, pp. 1668-1671, IEEE.
- [15] M.R., Hasan, M., Jamil, and M.G., SaifurRahman, (2004), “Speaker Identification Using Mel Frequency Cepstral Coefficients”, *3rd International Conference on Electrical and Computer Engineering, Dhaka, Bangladesh*, pp. 565-568.
- [16] M.Z., Bhotto and M.R., Amin, (2004), “Bangali Text Dependent Speaker Identification Using Mel Frequency Cepstrum Coefficient and Vector Quantization”, *3rd International Conference on Electrical and Computer Engineering, Dhaka, Bangladesh*, pp. 569-572.
- [17] Piotr Dalka and Andrzej Czyzewski, (2010), “Human-computer interface based on visual lip movement and gesture recognition”, *International Journal of Computer Science and Applications, Techno mathematics Research Foundation Volume-7, No. 3*, pp. 124 – 139.
- [18] Potamianos, G., Neti, C., Gravier, G., Garg, A., & Senior, A. W., (2003), “Recent advances in the automatic recognition of audiovisual speech”, *Proceedings of the IEEE*, 91(9), pp. 1306-1326.
- [19] R.L.K.Venkateswarlu, R. Vasantha Kumari, G.VaniJayaSri, (2012), “Speaker Independent Recognition System with Mouse Movements”, *International Journal of P2P Network Trends and Technology*, Volume-2, Issue-1, pp. 32-37, ISSN: 2249-2615.
- [20] SandeepKaur, (2012), “Mouse Movement using Speech and Non Speech Characteristics of Human Voice”, *International Journal of Engineering and Advanced Technology (IJEAT)*, ISSN: 2249 – 8958, Volume-1, Issue-5, pp. 368 – 374.
- [21] S Srinivasan, Aichner, W B Kleijn & W Kellermann, (2006), “Multichannel parametric speech enhancement”, *IEEE Signal Processing Letters*, vol 13, May, pp 304-307.
- [22] Su, M. C., & Chung, M. T., (2001), “Voice-controlled human-computer interface for the disabled”, *Computing & Control Engineering Journal*, 12(5), pp 225-230.
- [23] Tyler Simpson, Colin Broughton, Michel J. A. Gauthier, and Arthur Prochazka, (2008), “Tooth-Click Control of a Hands-Free Computer Interface”, *IEEE transaction on biomedical engineering*, Volume-55, No. 8, pp. 2050-2056.
- [24] Urban, M., & Bajcsy, P., (2005), “Fusion of voice, gesture, and human-computer interface controls for remotely operated robot”, *In Information Fusion, 8th International Conference* Vol. 2, pp. 8, IEEE.
- [25] Wald, M., (2005), “Using automatic speech recognition to enhance education for all students: Turning a vision into reality”, *In Frontiers in Education, FIE'05, Proceedings 35th Annual Conference*, pp. S3G-S3G, IEEE.
- [26] Wu, D., Tanaka, M., Chen, R., Olorenshaw, L., Amador, M., & Menendez-Pidal, X., (1999), “A robust speech detection algorithm for speech activated hands-free applications”, *In Acoustics, Speech, and Signal Processing, IEEE International Conference*, Vol. 4, pp. 2407-2410, IEEE.
- [27] Wu, K. G., & Chen, P. C., (2001), “Efficient speech enhancement using spectral subtraction for car hands-free applications”, *In Consumer Electronics, ICCE International Conference*, pp. 220-22, IEEE.