Journal of Innovation & Development Strategy (JIDS)

(J. Innov. Dev. Strategy)

HANDS-FREE COMPUTING AND BANGLA TEXT WRITING USING SPEECH COMMANDS FOR DISABLED PEOPLE

M.M. ISLAM, M.N.R. PROTIK, F. KHATUN, M.Z. SULTAN AND S. SARKAR

HANDS-FREE COMPUTING AND BANGLA TEXT WRITING USING SPEECH COMMANDS FOR DISABLED PEOPLE

M.M. ISLAM[1]*, M.N.R. PROTIK[1], F. KHATUN[1], M.Z. SULTAN[2] AND S. SARKAR[1]

[1]Telecommunication and Electronic Engineering Department, Hajee Mohammad Danesh Science and Technology University, Dinajpur -5200, Bangladesh; [2]Electrical and Electronic Engineering Department, Hajee Mohammad Danesh Science and Technology University, Dinajpur -5200, Bangladesh.

*Corresponding author & address: Md. Mehedi Islam, E-mail: mehedi_eap@yahoo.com

ABSTRACT

Islam MM, Protik MNR, Khatun F, Sultan MZ, Sarkar S (2015) Hands-free computing and Bangla text writing using speech commands for disabled people. *J. Innov. Dev. Strategy.* 9(3), 7-12.

This work describes about a system which has been developed especially for disabled people who can at least speak. This system includes two main functions Hands-free computing and Bangla text writing for disabled people through speech recognition system. Hands-free computing is for controlling or operating computers such as for cursor movement and clicking operations of the mouse pointer on the screen without the use of the hands interfacing with commonly used human interface devices such as the mouse and keyboard. Another function Bangla text writing is for writing Bangla text from the English words which is one of the main targets as well as advantage of my system. This advantage includes the features, example as for creating Bangla document or for writing Bangla text in any application. Speech dependent this software can be a great choice over HCI (Human Computer Interface) devices for disabled people.

Key words: accessibility to disabled users, bangla text writing, bangla phonetic, hands-free computing, isolated-word speech recognition

INTRODUCTION

Hands-free computing is a term used to describe a configuration of computers so that they can be used by persons without the use of the hands interfacing. Hands-free computing is important because it is useful to both able and disabled users. Speech recognition systems can be trained to recognize specific commands and upon confirmation of correctness instructions can be given to systems without the use of hands. This may be useful while driving or to an inspector or engineer in a factory environment. Likewise disabled persons may find hands-free computing important in their everyday lives (Islam *et al.* 2011).

Speech recognition is the process of conversion of speech to text (Lau and O´Leary, 1993). The objective of human speech is not merely to transfer words form one person to another person, but rather to communicate understanding – a thought, concept or idea. In computer speech recognition, a person speaks into a microphone or telephone and the computer listens. This however, doesn't mean that the computer understands the meaning of the words it has heard. More often, the computer simply attempts to transcribe the speech into a textual representation (Lau and O´Leary, 1993).

Bangla is an important language with a rich heritage and is spoken by approximately 8% of the world population (Hasanat *et al.* 2002 and Rashid *et al.* 1998). Among the world countries, Bangladesh is one of the developing countries of the South Asian region. The population near about 122 million of which 3% people are intellectually disabled (according to WHO assumption). Persons with Intellectual Disability are classified as: who's IQ

(Intelligent Quotient) is below the normal range, or has lost intellectual balance or is damaged, partly or wholly (Haider and Convenor, 2002). The purpose of this paper is to introduce with this system to assist disable persons in their interaction with computing devices. It gives the results of a survey as to how people with limited or no hand or finger movement interact with computers (Rabiner *et al.* 1979).

Speech recognition software (also known as speech to text software) allows an individual to use their speech instead of typing on a keyboard. Speech recognition may be used to dictate text into the computer or to give commands to the computer (such as opening application programs, pulling down menus, or saving work). Speech recognition software allows for a quick method of writing onto a computer. It is mainly useful for people with disabilities who find it difficult to use the keyboard.

There are several speech recognition software programs available on the market now. Unfortunately, with other speech recognition products, user cannot write Bangla text, create Bangla document, also cannot write Bangla text in any application, where my system gives the advantages to overcome the limitation on Bangla text writing. For fulfill the purpose, isolated-word speech recognition is used here to do so. An isolated-word speech recognition system requires that the speaker pause briefly between words. A sentence constructed from connected words does not represent real speech as it is actually concatenation of isolated words. For Isolated word the assumption is that the speech to be recognized comprises a single word or phrase and to be recognized as complete entity with no explicit knowledge or regard for the phonetic content of the word or phrase (Hasnat *et al.* 2007). In this technology words are used as a unit for isolated speech recognition. Word used as a unit for better detection.

7

J. Innov. Dev. Strategy 9(3):December 2015

MATERIALS AND METHODS

A. Low Signal to Noise Ratio

The program needs to "hear" the words spoken distinctly, and any extra noise introduced into the sound will interfere with this. The noise can come from a number of sources. Users should work in a quiet room with a quality microphone positioned as close to their mouths as possible. Low-quality sound cards, which provide the input for the microphone to send the signal to the computer, often do not have enough shielding from the electrical signals produced by other computer components. They can introduce hum or hiss into the signal.

B. Overlapping Speech

Current systems have difficulty separating simultaneous speech from multiple users. "If you try to employ recognition technology in conversations or meetings where people frequently interrupt each other or talk over one another, you're likely to get extremely poor results," says John Garofolo.

C. Intensive Use of Computer Power

Running the statistical models needed for speech recognition requires the computer's processor to do a lot of heavy work. One reason for this is the need to remember each stage of the word-recognition search in case the system needs to backtrack to come up with the right word. The fastest personal computers in use today can still have difficulties with complicated commands or phrases, slowing down the response time significantly. The vocabularies needed by the programs also take up a large amount of hard drive space. Please do not revise any of the current designations.

D. Homonyms Factor

Our computer system only knows the English word and their pronunciation. A computer speech recognition system only detect the correct pronunciation which is better done by those who are possesses English language as mother language. But it is naturally seen that correct pronunciation is not perform by us as well as for Bangladeshi people. For a grammar is created which contains some word. Word detection is better and nearer to 100% from this grammar which contains limited word. The grammar which contains huge word, detection is so poor there.

For the grammar which contains so many words detect as below-

Nine (speech) ⟶ Line, Time(detect wrong word)
Ten (speech) ⟶ Then, Men(detect wrong word)

This is also known as "Homonyms". Homonyms are two words that are spelled differently and have different meanings but sound the same. "There" and "their," "air" and "heir," "be" and "bee" are all examples. There is no way for a speech recognition program to tell the difference between these words based on sound alone. However, extensive training of systems and statistical models that take into account word context has greatly improved their performance. For the grammar which contains limited words, detection as below-

Nine (speech) ⟶ Nine (detect correct word)

In limited words grammar fault detection is not possible because words are so unique as well as their pronunciation is also so unique. It is better suit for those who are disable, also they are my main target. Enable or normal people usually does not prefer this because writing Bangla, with other system is so first then this system. But it is nearer of 100% fast for disable people comparing them. Please do not revise any of the current designations.

CREATING GRAMMER

Isolated-word speech recognition is used here for detecting speech word uniquely for performing two tasks. One task is implemented here for hands-free computing and other task for writing Bangla text in various applications. Here my main target is the interaction between disable people and computer. Those disable people who have lost intellectual balance or are damaged, partly or wholly, but speak. Approach towards the isolated speech recognition is quite simple, for this simple dictionary is used which contains only the input and no language model is necessary. Also regular grammar is used which has the small vocabulary. The task Grammar which defines all of the legal word explicitly. Step begins by labeling the speech data for each word of the created grammar. The label is same as the text that represents the spoken word.

COMPUTER INTERFACING & BANGLA WRITING

Define Speech is input via microphone and its analog waveform is digitized. The job of the recognition system is to derive necessary information from the waveform needed to make the correct decision. For recognition process 10 words are used to make grammar. These words are controlling the whole process.

Recognition process is used here for two operations:
- For computer interfacing with speech command
- For writing Bangla text.

For computer interfacing, the user normally needs to operate computer are-
- Cursor movement
- Double click to open any directory or any file
- Single click for selecting any file or folder and also access webpage content
- Right click for getting option menu to do other related task

For cursor movement the used words are-left, right, up, down.
- For double click the used word is-"Enter"
- For single click the used word is-"go"
- For Right click the used word is-"press"

For writing Bangla the process is followed-
- Get command through speech which gives input to text area. Bangla Phonetic is used to write English to Bangla text by the speech command.

English writing also included here at default mode.

SPEECH RECOGNITION & PROCESSING ALGORITHM

In this software the recognizer can recognize only the words defined in the grammar. The number of words are defined in grammar is so finite level. For this, better detection is possible.

A. For Computer Interfacing With Speech Command

For better detection and performing command "left, right, up, down" words are used for cursor movement and for clicking event enter, go, press for double, single and right click are used.
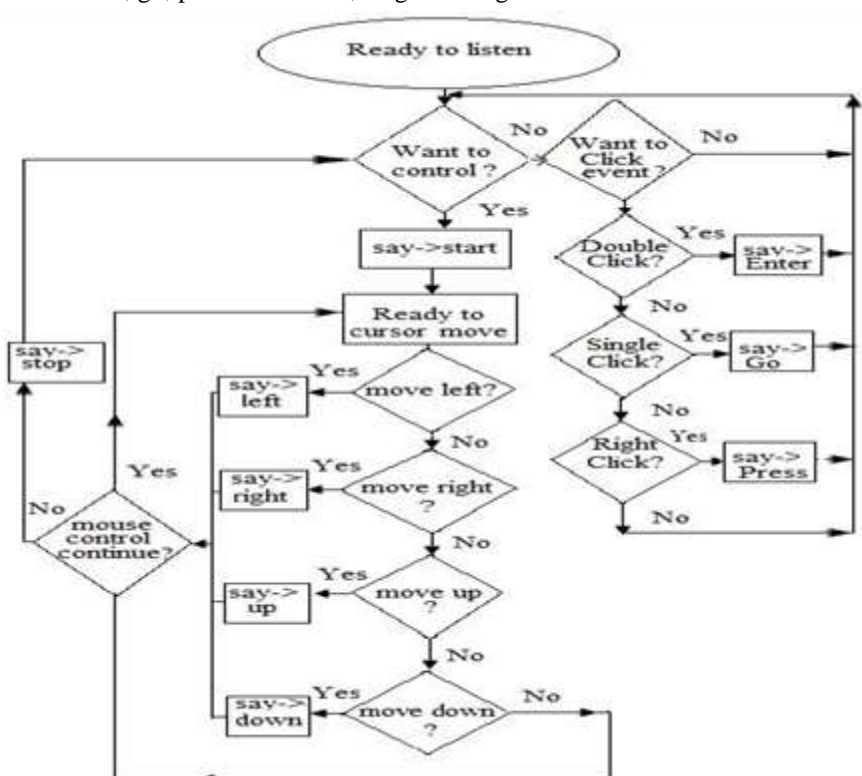

Fig. 1. Flow diagram for controlling PC by speech commands

"Start" command is used for starting cursor movement event then it listen "left, right, up, down" commands for cursor moving in their meaning direction. "Stop" command is used for stopping the cursor movement and ready to listen the command for clicking event. Figure 1 shows the computer controlling algorithm by speech commands.

B. Default Mode English Writing

In default mode English writing same algorithm is used liked Bangla text writing but Bangla phonetics is not used here. This helps user to write English in any application.

C. Writing Bangla Text

Speech command "Bangla" is used for starting the listening process for Bangla writing. Here "left, right, up, down" words are used as commands for choosing or selecting the English letter, and "ok" command is used for write the choosing letter to the input area. Figure 2 shows the Bangla text writing algorithm by speech commands.

Such as from Figure 3, at the choosing letter "A", "Ok" command write the value "a" to the input area. The remaining letters of the word "amar" are also written according to same process.

Selection "A" + speech command "Ok" = write "a"
Selection "M" + speech command "Ok" = write "m" Selection "A" + speech command "Ok" = write "a" Selection "R" + speech command "Ok" = write "r" Total result="amar".
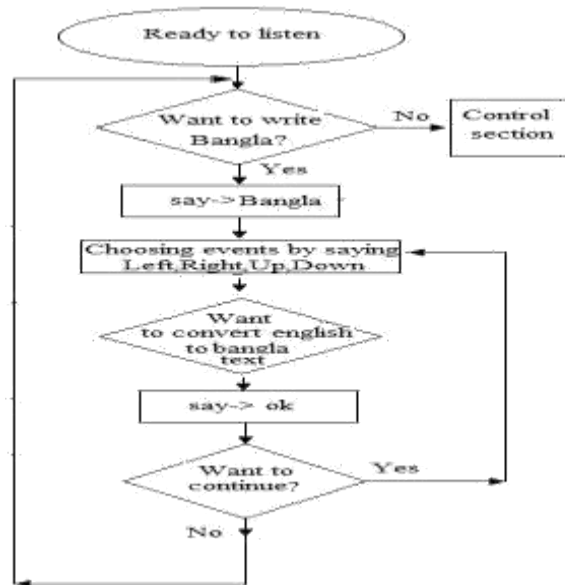

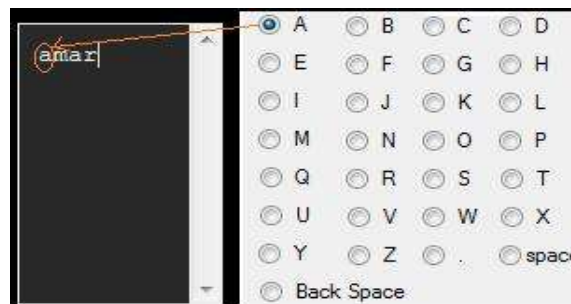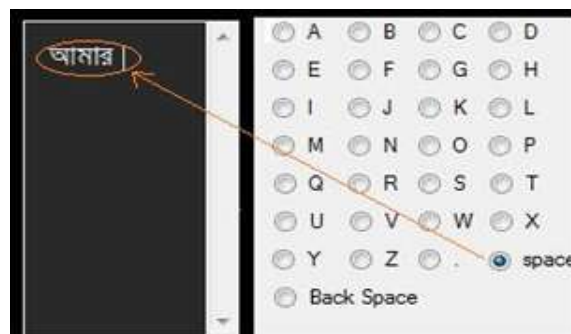Fig. 2. Flow diagram for writing Bangla text

Fig. 3. Write input value from the speech command

After writing the required word for Bangla text, this word is converted into Bangla text using Bangla phonetics by "Ok" command at the choosing point of "space" which is showed in Figure 4.

Selection "space" + speech command "Ok" = write "আমার"

Fig. 4. Converting English text into Bangla by speech command

Two features are included here:-

- Users can save Bangla text in office document format (.doc or .docx).
- Users can send Bangla text in any application(such as Facebook post, writing URL, etc in any application)

Fig. 5. Bangla phonetic layout

Both are done by the speech commands frequently. Users need to learn a few tricks to show proper Bangla letters, even when he/she is writing English. Example of some Phonetic layout shows in Figure 5.

EXPERIMENTAL RESULT & DISCUSSION

For developing entire algorithm, Microsoft Visual Studio 2010, object oriented programming language C#, Microsoft.Office.Interop.Word.dll, User32.dll, NET 3.5, office.dll are used. User computer was a 2.0 GHz Dual Core with 2 GB RAM and operating system was windows 7.

While the accuracy of voice recognition has improved over the past few years some users still experience problems with accuracy either because of the way they speak or the nature of their voice. At the first experiment, this system was tested on nine normal students. The participants' speech recognition accuracy for Bangla text writing is showed in Table I. The percent accuracy was calculated using the formula:

Percent Accuracy = ((Total detected words - number of errors) / Total detected words) * 100

Fig. 6. Results of nine student's accuracy

Figure 6 shows the accuracy of normal students and it also clear that it gives proper accuracy for the disabled people also. The calculated result of percent accuracy from Figure. 6 is different because speech signal is very much sensitive to the speaker's properties such as age, sex, and emotion, and environment.

Average Accuracy = 92.26%

In another experiment, this system was tested on physically disable people in finding out whether they can use this system properly or not. Ten people of different age with physical disabilities have tried this system to operate computer and overall six of them can successfully use it. The result also indicates that those who have no trouble of brain, tongue and impediment, can operate the computer for various purposes including Bangla text writing. The results are summarized in Table 1.

Table 1. Summary of results for the first ten people with disabilities to try this system

Age	Gender	Condition	Remarks
12	Male	Loss of limbs	No problem, pc control and writing ok
20	Male	Injury to the spinal cord	Problem sometime in proper voice command
22	Male	Paralyzed arms and legs	No problem works satisfactory
21	Female	Cerebral palsy	No severe problem
25	Female	Cerebral palsy	No severe problem
30	Male	Mute with paralyzed arms and legs	Problem arise cause it is voice dependent
10	Male	Mute with loss of limbs	Problem arise cause it is voice dependent
17	Female	Traumatic brain injury	Problem sometime in proper voice command
35	Male	Mute with loss of limbs	Problem arise cause it is voice dependent
45	Female	Loss of limbs	No problem like this system

CONCLUSION

In digital signal processing, speech recognition is a very challenging field. Scientists achieved remarkable success in speech recognition for many languages. Unfortunately in Bengali, works on speech recognition is in the very preliminary stage. In this respect my proposed system can be very helpful and valuable for writing Bangla text as well as hands-free computing for the disable people who have lost physical mobility in their upper extremities due to accidents, or diseases. This paper clearly describes the theory and implementation details of my entire development task. A single- computer version of the system has been developed. Future work will incorporate the independent operating system and the improvement of all the problems of the background pattern.

REFERENCES

Haider MZ, Convenor (2002) "COUNTRY REPORT: BANGLADESH" Care-taker Body, SWID Bangladesh Additional Director, Department of Social Services.

Hasanat A, Karim MR, Rahman MS, Iqbal MZ (2002) "Recognition of Spoken letters in Bangla" 5th International Conference on Computer and Information Technology, East West University, Dhaka, Bangladesh.

Hasnat MA, Mowla J, Khan M (2007) "Isolated and Continuous Bangla Speech Recognition: Implementation, Performance and application perspective" Proc. Statistical Natural Language Processing, Pattaya, Thailand, December.

Islam MO, Mandal AK, Hossain MD (2011) "ThirdHand: A Novel and Cost-effective HCI for Physically Handicapped People" 14th International Conference on Computer and Information Technology (ICCIT 2011), 134–138.

Lau C, O´Leary S (1993) "Comparison of Computer Interface Devices for Persons with Severe Physical Disabilities", The American Journal of Occupational Theraphy, Vol. 47, 1022-1030.

Rabiner LR, Levinson SE, Rosenberg AE, Wilpon JG (1979) "Speaker Independent Recognition of Isolated Words Using Clustering Techniques" IEEE Trans. Acoustics, Speech, Signal Proc., ASSP-27:336-349.

Rashid M, Meftauddin M, Minhaz MN (1998) "Speech Password Security System Using Bangla numerals as Fixed Text" Proceedings of International Conference on Computer and Information Technology, Dhaka.