

Voice Recognition Browser for the Visually Impaired Learners (Mg Sys VISI)

HALIMAH BADIOZE ZAMAN, DAVID KENNEDY, CHOO W.O., AZLINA AHMAD
AND AIDANISMAH YAHYA

Abstract: Learning through the use of web technology or web based learning has become an important media in the education revolution of the 21st century. The Internet particularly, has become an important tool for learners to acquire information and knowledge that encompasses various elements such as text, graphic, numeric, and animation for their learning process. Learners soon learn that the links in the Internet can lead them to various web pages that can lead them to more information that have a link with one another or to other information that has no link at all with the previous information. However, the visually impaired learners who actually represent a substantial proportion of the world's population living in certain parts of the world have no access at all to this tool nor can it be easily taught to them as they are not able to see the links in the web pages. There is a need to democratize education as this is the basic human right and a way to achieve world peace. This paper hopes to highlight the Mg Sys VISI system to enable the visually impaired learners experience the world of the Internet, which comprises of five modules: Automatic Speech Recognition (ASR), Text-to-Speech (TTS), Search engine, Print (Text-Braille) and Translation (Braille-to -Text) module. Initial testing of the system indicates very positive results.

Keywords: learning through voice browser, visually impaired learners, voice pattern recognition, voice recognition, voice recognition browser, voice recognition system.

I. INTRODUCTION

Malaysia is aspiring to be a developed nation in 2020. As she works towards achieving this aspiration, she has to encounter the changes that is rapidly taking place in the world: changes in technology, values, culture, world view and in the way the country competes and allocates its intellectual resources. It is therefore imperative that Malaysia responds to these changes particularly in the area of education. The country had made substantial achievements in providing opportunities for all types of learners (including the learners with learning disabilities like the visually impaired) to maximize their potential in the past decades. Malaysia feels that there should be democratization of education for all, to ensure that everyone of its population have access to education which is basic human right and the answer to achieving world peace. It is with this in mind that a group of researchers through a grant obtained from the Ministry of Science, Technology and Innovation (MOSTI), conducted a research in collaboration with the Malaysian Association of

the Blind (MAB) to develop a voice recognition browser called Mg Sys VISI for the visually impaired learners.

II. STATEMENT OF THE PROBLEM

Learning through the web or web-based learning has become an important teaching and learning media in the education revolution that is taking place throughout the world. Learners today, have to acquire specific skills in order to use the browser effectively when browsing the Internet in the process of conducting their learning tasks. The Internet includes all types of information and knowledge based on the various elements: text, graphic, numeric and to a lesser extent, audio, video and animation. However, the visually impaired learners are deprived of this very important tool of learning. The visually impaired learners are left to learn using the conventional method of 'talk and braille' by the teacher.

They cannot use the functionalities like 'linking' of web pages through the available browsers such like the Internet Explorer and Netscape Navigator, to acquire the various information and knowledge from the digital libraries and other sources, required for their work independently [1] as they are not able to see the screens and the necessary links. They sometimes also has a problem in reading particularly with word decoding and phonological processing [2]. The visually impaired learners are also deprived of enjoying services in the Internet like sending and receiving e-mails unlike their other normal friends. Due to the fact that the conventional browser available is developed for normal users that enables them to control the functionalities designed in the browser, it is thus not suitable for the visually impaired learners. It is just too complex for them to learn to use the functionalities physically or cognitively. Atkins & Collins [3] said that the use of multimedia through voice recognition can help the visually impaired learn more effectively.

Due to the fact that some teachers are not proficient with the Braille dots, they request that the system be able to convert braille back to text so that teachers can mark their assignments without having to change assignments written in braille into text manually as they were doing then.

Manuscript Received April 20, 2007; Revised June 1, 2007

III. DESIGN AND DEVELOPMENT OF MG SYS VISI

In order to overcome the problems faced by the visually impaired learners, a solution in the form of a specialised voice recognition browser called the Mg Sys VISI was designed and developed. The objectives of the research were as follows:

- To design and develop an Internet browser that enables them to browse the Internet through a voice recognition system.
- To develop the browser using a voice recognition application that allows the visually impaired learners to send and receive e-mails.
- To develop a browser that allows the visually impaired learners to search for articles through a search engine and print the desired articles in braille.
- To develop a translation module that is able to convert braille back to text to help teachers mark assignments of the visually impaired.

The system was designed based on the software engineering participatory method, whereby the designers and developers work very closely with the visually impaired learner at the Malaysian Association of the Blind (MAB). The process was iterative and the system was improved each time the learner used a certain module. The module is improved and the next module is developed. Throughout the process, the user is an active collaborator. The methodology is as indicated in Figure 1.

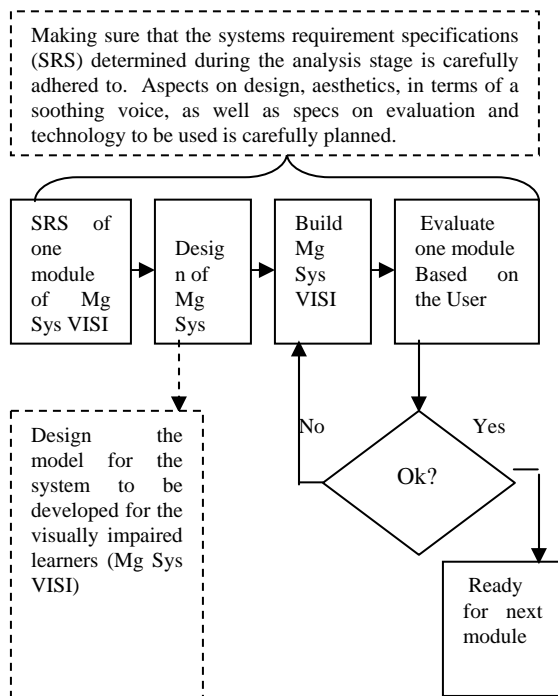


Fig. 1. Iterative Participatory Software Engineering Model of Development for Mg Sys Visi

The system was designed based on the Holistic Cognitive Voice Haptic Architecture model based on Cognitive theories [4] as shown in Figure 2. Based on this model, it can be observed that the visually impaired learner can interact with the system through the use of the microphone. Graevenitz [5], [6] found from their research that a voice recognition system can be designed to contain features that can be speaker dependent, adaptive or independent. Mg Sys Visi was built based on the independent feature.

Based on the HCVHA model, six modules were developed and they are as follows:

- Automatic Speech Recognition Module (ASR)
- Text-to-Voice Module (TTV)
- Search Engine (SE)
- Text-to-Braille (TTB)
- Braille-to-Text (BTT)

A. Model and Modules of Mg Sys Visi

The Automatic Speech Recognition (ASR) is the first module found in the Mg Sys Visi for the visually impaired learners.

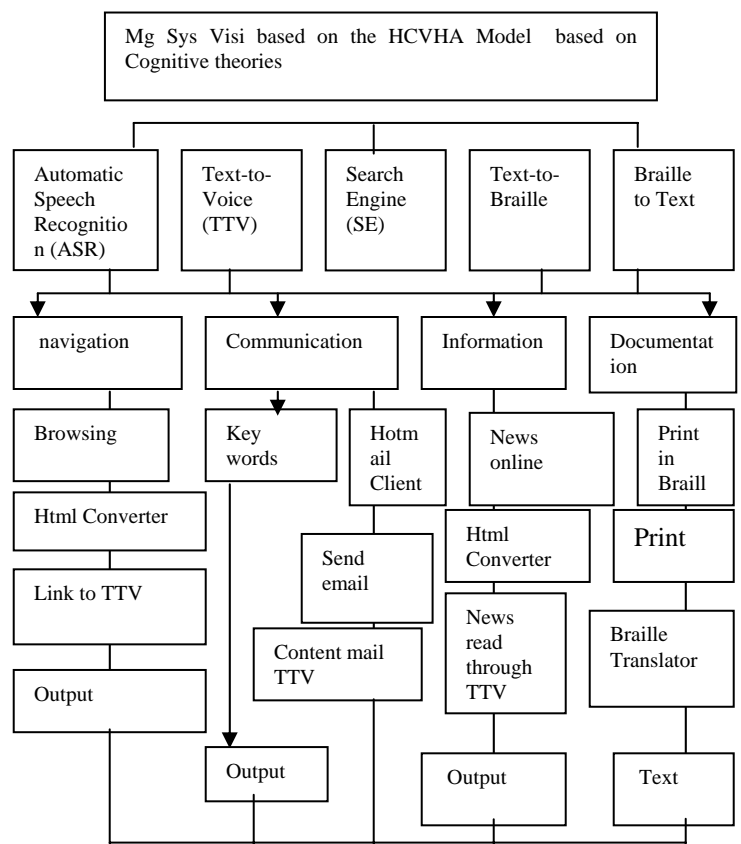


Fig. 2. Holistic Cognitive Voice Haptic Architecture (HCVHA) Model

The most important factor that needs to be taken into consideration with the ASR is the intersession variability and the variability over time [7], [8]. The changes can occur from the user themselves when they are recording their voice in a

different place or the noise in the background of the place where the voice of the user is being recorded. The speaker normally is not able to repeat in the same manner with the same intonation at all times. However, it always better when a recording is conducted straight at one time as compared to when recording is interrupted and continued at a different time. Two types normalisation exist that is: the parameter domain normalisation and the Equalisation domain technique.

The former known as spectral equalisation is also known as the blind equalisation method [9]. This technique can reduce the linear channel and the long-term spectral variation. This method is most effective for speaker dependent application where the speaker involved normally uses long sentences. This technique is not suitable for a speaker recognition that uses short words. The latter, however uses the likelihood ratio approach. The likelihood ratio is defined based on two situations to measure the style of speech spoken. The first likelihood is the acoustic data that could contribute to the recognition of the speaker and the second likelihood is the speaker could be an imposter. The normalisation based on the posterior likelihood was also studied. The difference between the normalisation based on the ratio likelihood and the posterior likelihood is to ensure that the claimed speaker is recorded for normalisation [10]. Based on experiments carried out, both of the methods obtain almost the same level of effectiveness. Both methods allows the system to lessen the need to depend on only the claimed speaker.

The Text-to Voice Module (TTV) includes the technology to synthesise speech of the speaker whereby the input in the form of text is converted automatically to the audio output in the form of voice [11],[12]. In the case of Mg Sys Visi, the TTV converts the html text codes through the html converter into voice codes. The main function of the TTV engine then, is to process the input text and converts it to the audio output and then played to the user. The level of vocabulary output of the TTV is based on the set of words spoken into the engine. The TTV engine comprises of two components: the text processor and the synthesiser. When a text is put into the engine, the text processor analyses the words, phrase or sentences that is put in. The system has to make rough semantic meaning. The text that has been normalised will be sent to the synthesiser to produce the sound heard by the visually impaired learner.

The html converter changes the html codes into text. This so that the engine Text-to-Voice (TTV) can read the web content and meaningfully for the visually impaired learner. The converter will remove all the tags in the html codes and leave it as text only. For example:

Input: <html>
 <h1> This is an example of the conversion </h1>
 </html>

Output:
 This is an example of the conversion

Apart from removing all the taggings, the role of the converter is also screen the pictures and graphics that cannot be converted into text.

The Hotmail is an e-mail service that is popular in the Internet world. The Hotmail Client module incorporated into the Mg Sys Visi for the visually impaired learners allows them to send e-mail and receive e-mails from their friends through a voice recognition technology. To retrieve hotmail from the Internet, the visually impaired users of Mg Sys Visi has to follow certain protocol. The protocol used by the system is Httpmail. This is almost similar to the XML structure where it is used to put elements that is desired to find.

For authentication, Http headers were used. To build the client, two components were required and they were as follows:

- i. One proxy that can 'talk' Httpmail and can recognise HTTP.
- ii. The real client that access the Hotmail through the proxy uses the Xpath for parsing response.

The proxy is responsible to send send queries or requests HTTP to the Hotmail server and receives the response from the client.

The search engine (SE) of Mg Sys Visi helps the visually impaired learners to acquire and search for information required for their learning process. The engine allows the visually impaired to say keywords into the microphone and the SE will search for the information required. To search for the information in the millions of websites, SE uses robots known as "*spiders*" to build a list of words that are found in the websites. untuk membina satu senarai perkataan The information found from the Internet

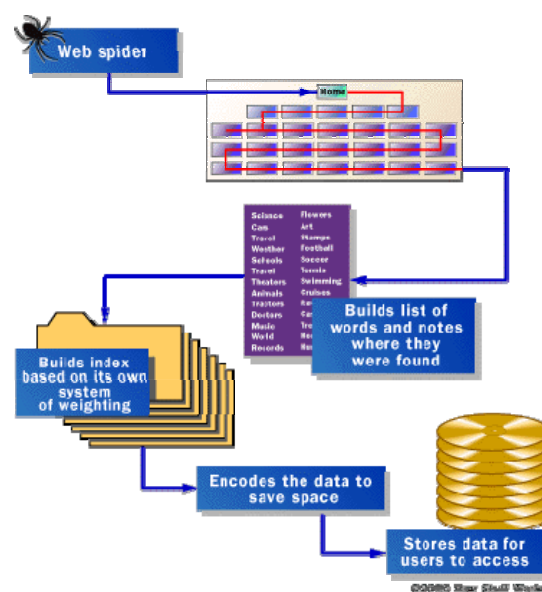


Fig. 3. Spider Process in SE

will be listed and the headings of the articles are read out based on no. 1, 2, 3, ...The spider process can be observed in Figure 3. When the visually impaired learner requires a desired article, the number is mentioned through the microphone and the system reads out the abstract of the article. If the abstract is satisfactory, and the visually impaired learner wants the full article, the learner can say 'print' and the system will convert the text into braille through the Text-to-Braille (TTB) module for the user.

When the system was first developed, the four modules were thought enough to fulfill the needs of the visually impaired learners. However, when further discussions were held with teachers and students at the Malaysian Association of the Blind (MAB), they indicated another problem that the teachers face. Currently, the teachers have to convert the assignments written in Braille into text manually before they can mark the assignments. This was tedious and took a lot of their time.

Thus, the translation module (BTT) was the fifth module developed in the Mg Sys Visi system for the visually impaired learners. With this module, the assignments of the students done in Braille can be automatically converted back to text. This saves a lot of the teachers' time.

B. Data Flow Diagram of Mg Sys Visi

Below are examples of the data flow diagrams of Mg Sys Visi and the end-user. Figures 4 and 5 show examples of the context diagram of Mg Sys Visi. The former shows the scope and boundary of the response between the system and the end user; whilst the latter shows a more detailed interaction between the system and the user. The navigational sub module of the Mg Sys Visi, functions as a web browser and capable of searching information online based on keyword searching.

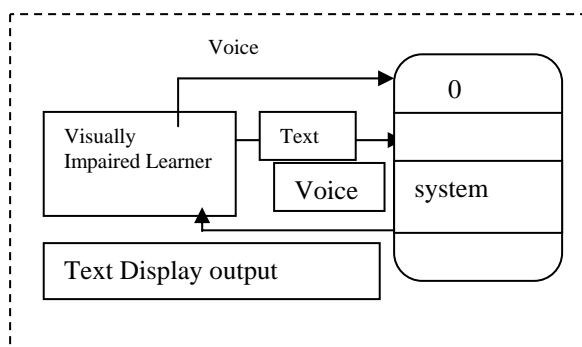


Fig. 4. Context Diagram for Mg Sys Visi at Level 0

The system reads the text (article) to the visually impaired learner by applying the text-to-voice technology. The information sub module allows the visually impaired learners to hear news being read to them online. The documentation module allows the visually impaired to read the article converted to Braille by the system. Figure 6 on the otherhand shows the level 1 data flow of Mg Sys Visi: From Text to Voice

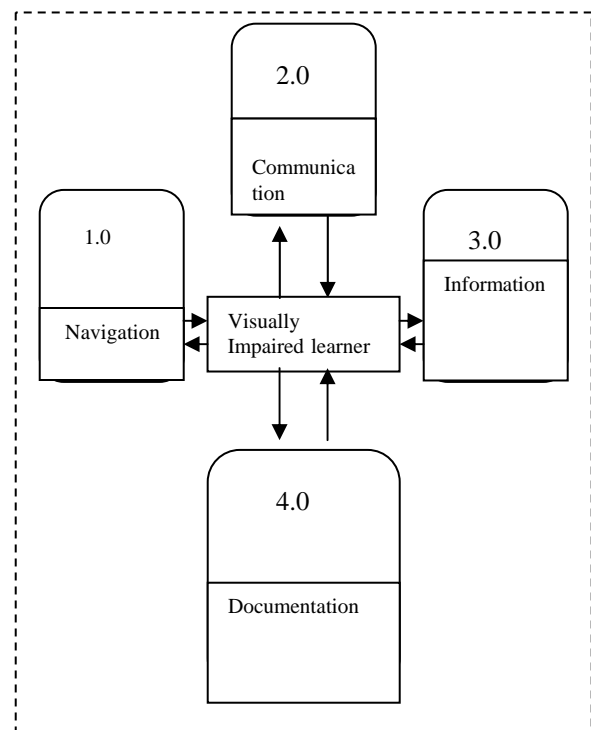


Fig. 5. More Detailed Context Diagram for Mg Sys Visi at Level 0

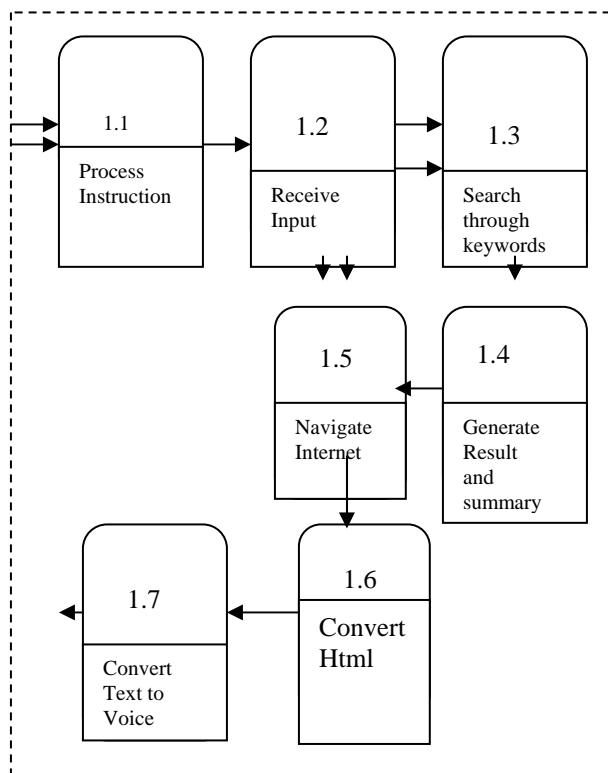


Fig. 6 shows the Level 1 data flow of Mg Sys Visi from text to voice

Based on the Level 1 data flow, it can be observed that system will process the voice instruction received; system will then search through the Intrnet based on the keywords given. System will navigate the Internet, and system will then



Fig. 11 Translator Module of Mg sys Visi

Figure 12 on the other hand, shows how the Braille codes are converted into text codes.



Fig. 12 Braille to Text (BTT) Module of Mg Sys Visi

Figure 13 shows the print screen of the Search Engine (SE) of the system.



Fig. 13 Print Screen of Search Engine Module of Mg Sys Visi

IV. CONCLUSION

The Mg Sys Visi was tested with teachers and students at the Malaysia Association of the Blind (MAB) and the preliminary study carried out with them indicates that Mg Sys Visi is effective in helping visually impaired learners browse the Internet for information necessary to carryout their assignments. The functionality of the system that is able to convert the Html codes to voice codes means that visually impaired learners can now browse the Internet to carryout their assignments.

The study also indicates that teachers were also satisfied with the system as it did not only helped their students, but it also helped them in carrying out their job. This is due the fact that the system can now convert the assignments of their visually impaired students in Braille into text again. This means that they can now mark the assignments of their students without having to convert them to text first manually. The system can now do it for them automatically. This is the first of such a system in Malaysia.

V. REFERENCES

- [1] Bandura, A. Psychological Modeling. New York: Aldine Atherthon. 1971.
- [2] Biggs, M.L. Learning Theories for Teachers. Ed.ke-3. New York: Harper and Row Publishers. 1976.
- [3] Atkins, P.R., & Collins, T. Physical Constraints in Sonar Design, The Journal of the Acoustical Society of America, 109, 5, Part 2, 2001, 2285-2286
- [4] Reed, S.K. Cognition. Fifth edition California: Addison Wesley, California, 2000.
- [5] Graevenitz, G. A. von (n.d.). About Speaker Recognition Technology. Bergdata Biometrics GmbH, Germany
- [6] Rozmiarek, D.J. Continuous Speech Recognition and Computers: A Written Communication Tool for Students with learning disabilities. Delaware: University of Delaware. 1998.
- [7] Dragon Systems Unveils Revolutionary Breakthrough with Continuous Speech Recognition. New York: Dragon Systems. 2001.
- [8] Snaidoo, A. Automatic Voice Recognition System. California: University of California. 2003.
- [9] Ainsworth, W.A. *Speech Recognition by Machine*. London: Peter Peregrinus Ltd on behalf of the IEEE (IEEE Computing Series, 12.) 1988.
- [10] Lamel, L.- Gauvain, J.L. "Speech recognition", in Mitkov, R. (Ed.) *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press. 2003
- [11] National Center for Improved Practice in Special Education (NCIP) Update on Voice/Speech Recognition. New York: NCIP. 2003.
- [12] Wang, Y.Y. Deng, L. Acero, A. "Spoken language understanding", *IEEE Signal Processing Magazine* 22, 5 16-31. 2005