International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

ECHO: Hands-Free Computer Interaction using Speech Recognition System for the Debilitated

Elmer C. Matel, Herchel Aquines, Carl Brian S. de Guzman, Jasond Isaac V. Gustilo, Gregg Victor P. Ibera, Priam D. Jodilla

Abstract*: Computer nowadays is a must-have tool for most people. However, it is not a tool to be used by people with physical disabilities, especially the ones lacking an arm or two. The goal of this paper is to introduce a system that will help computer users perform tasks and make use of computer features and functions despite their physical limitations through the Speech Recognition System (SRS) in the English language. Ideally, this aims to provide users with an alternative way of interacting with the computer system and navigate through its functions using SRS in place of peripheral devices. It can be used to navigate through menus, open and manage applications, open certain websites, browse the internet, and type words, letters, numbers, and symbols using the dictation mode. For the testing phase, the following test cases were used: Functionality Testing, Stress Testing, and Compatibility. The testing phase yielded a result of 94.79% for the functional, 100% for stress, and 100% for compatibility, effectively ensuring that the software is working as intended. The evaluation results conforming to the standards of the ISO/IEC 9126-1 yielded a mean of 3.57 with a standard deviation of 0.52 interpreted as 'Highly Acceptable', which means that the software can be used as an effective alternative to peripheral devices and can even be used to complement its usage.

***Index Terms*: Speech recognition, Intelligent Personal Assistant, Voice command, Assistive technology**

I. INTRODUCTION

"An easier way to do things" this is what most people are looking for, and the answer for that is computers. But not all people are capable of using computers. This is especially true for people with disabilities. These things, whether temporary or permanent, are something that only serves as hindrances for a person. It is also what prevents someone to perform a certain task [1] and [2]. A basic computer setup has two vital parts that are required for it to be utilized; the mouse and the keyboard. Without these two, not even the most basic functions of the computer can be used. This becomes a disadvantage for people with physical disabilities, especially the ones lacking a limb or two. Without the body part to do so, the individual would not be able to do anything productive with a computer.Fifteen percent (15%) of the population around the globe, which is approximately 1 billion people, experience some form of disability [3]. One of the most prevalent disabilities is the lack of limb, and it is said to have an incidence of one and five-tenths per 1,000 people. The world population of people lacking a limb or amputees is 10 million. There are one billion people in this world having difficulties doing regular activities and using types of machinery like computers, and 10 million are probably incapable of using it at all [4]. An intelligent personal assistant (IPA) on the other hand is a software agent that helps people to perform tasks and can be used in conjunction with the SRS to provide people with technology-driven ways of performing tasks. This way, the IPA can do the task for the individual, instead of just providing help or assistance. Most of the tasks that can be done with the computer require the user to make use of a peripheral device like a mouse or a keyboard. This requires the user to use hands and arms to fully utilize these devices. There are also instances where the user must also make use of certain shortcuts or key combinations to fulfill the conditions for a certain task to be done. Examples of both these problems include the management and manipulation of applications. Simply opening and closing an application already requires three to four left clicks to be done. Even the very process of taking a screenshot takes a lot of steps before the actual screenshot is saved. All of these are impossible to accomplish if the user is debilitated. The performance of a speech recognition system is normally influenced by the audio input system. It degrades due to noise from other outer sources. The accuracy and reliability of the system are affected by unwanted input and a low rate of results. Another challenge in the implementation of the Speech Recognition System (SRS) is the user responsiveness which happens when the resources are not yet available and users start to speak command that leads to a problem of synchronization of all data with multiple applications [5]. Several types of research were conducted in the same area to further improve the lives of those debilitated and others with physical limitations. N. Aktar et. al [6] proposed for a voice recognition-based intelligent wheelchair and GPS tracking system for physically handicapped people who are incapable of driving the wheelchair by hand.

Revised Manuscript Received on 30 July 2019.

* Correspondence Author

Elmer C. Matel*, College of Engineering, Computer Studies and Architecture / Lyceum of the Philippines University - Cavite, Philippines

Herchel Aquines, College of Engineering, Computer Studies and Architecture / Lyceum of the Philippines University - Cavite, Philippines

Carl Brian S. de Guzman, College of Engineering, Computer Studies and Architecture / Lyceum of the Philippines University - Cavite, Philippines

Jasond Isaac V. Gustilo, College of Engineering, Computer Studies and Architecture / Lyceum of the Philippines University - Cavite, Philippines

Gregg Victor P. Ibera, College of Engineering, Computer Studies and Architecture / Lyceum of the Philippines University - Cavite, Philippines

Priam D. Jodilla, College of Engineering, Computer Studies and Architecture / Lyceum of the Philippines University - Cavite, Philippines

The location of the patient can be tracked using the GPS module through a mobile app. The voice module is used to recognize the voice commands of the patient as well as recording the patient's voice. In the study made by M. Matić et. al [7], possibilities were explored in using available cloud-based voice recognition services, to enable voice control in the smart home automation system. The authors identified the appropriate architecture that provides better scalability of the voice command interface. One problem in the speech recognition system is noise and other unwanted sounds. The focus of the proposed study made by T. Athanaselis, et. al [8] is to improve the speech recognition in the presence of noise when a parametric method of signal enhancement is used. M. Hawley, et. al [9] made a study on voice-input voice-output communication aid for people with severe speech impairment. This is a new form of augmentative and alternative communication technique which intends to recognize the disordered speech of the user and builds messages which are then converted into synthetic speech. "Silent Speech Recognition as an Alternative Communication Device for Persons with Laryngectomy," was proposed by G.S. Meltzner, et. al [10]. This study focuses on patients who underwent surgical removal of the larynx due to trauma or diseases and thereby required an alternative voice source or assistive device to verbally communicate.

II. OVERVIEW

ECHO is a hands-free computer interaction model that utilizes SRS for the debilitated, which is an IPA software that helps users to perform tasks and navigate through the computer's functions through voice commands [11]. This paper promotes the great potential of SRS and addresses the issues concerning the capabilities of those physically challenged individuals to prove their worth through the use of this system. ECHO utilizes the Visual Studio framework and can be referenced with other projects, thus, making the system easy to be integrated with other existing SRS. Through the dictation and spell-out features, letters, words, numbers, and symbols can be typed out on a word-processing software through voice commands [12] and [13]. These functions can also be used when searching for keywords on the Internet. The preferred browser in using the search function of the software is Google Chrome. The printing of documents coming from any document-viewing software can also be done through voice commands. The software is also able to provide information about the current time and date. Features of the Shutdown Windows menu are also possible for use through voice commands. Taking screenshots is also made easy and the shot taken is automatically saved in a convenient location, which is the desktop. A Help menu is provided for users wanting to view all of the commands available for use. There is also a registration function that prompts the user to voice out their name or spell it out if the software can't recognize it. Users can also enter their gender and the software will address them as 'he' if the user is male and 'she' if the user is female. It can be run with a mid-spec computer or laptop as long as it is Windows-based because it does not take up that much space and does not exactly consume a huge amount of memory. The only exceptions are those running in Windows Microsoft Vista or below. Other exceptions include the voicing of several commands at a very fast rate, or when there are a lot of applications being run at the same time, both of these instances induce a strain on the computer's memory and thus might result in some delays and lags [14]. Devices running on Windows 7, 8, and 10 should have no problems in using the software. Still, it is strongly recommended to use the Windows 10 OS to ensure that of all the software's functions, since some commands will only work for Windows 10 users, namely features that are present in the Windows 10 OS but not on Windows 7 and 8, can be used. There are also some features that require an Internet connection to work. It is recommended to use the software with an earphone or a headphone device with a microphone to increase the software's accuracy in interpreting commands. Users should also avoid using the software in very loud or crowded areas because the software may not be able to properly interpret what the user is voicing out. There are also cases when the software will not understand a particular command or even execute a different, but similar-sounding one [15],[16] and [17]. This is true for most speech recognition software, because of the more words in the vocabulary the higher the chances of having words that sounds the same, thus, decreasing the accuracy of detection and recognition.

III. METHODOLOGY

After discussing some of the basic concepts used in this paper, the system flow and setup are presented in this section. This provides several diagrams and representations of the inner workings of ECHO.

A. ECHO System Flow

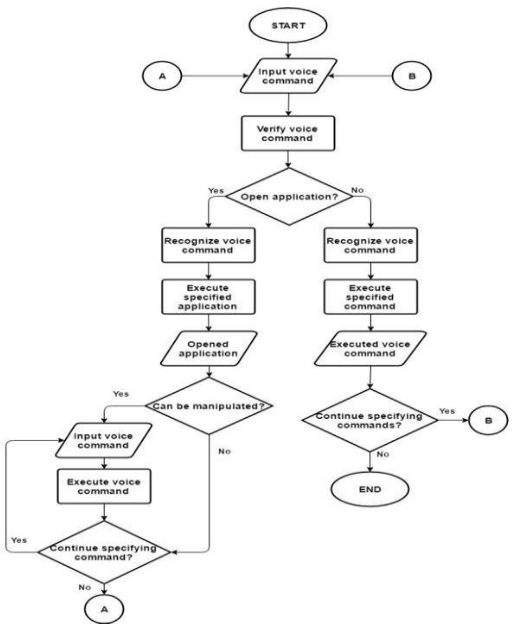

Figure 1. System Flow

The system flowchart, as illustrated in Figure 1, is the general process of how ECHO worked. First, the user would have to dictate or state a particular command. The software will then proceed to identify and verify the said command. Since each command has a different function, this is a vital requirement.

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Afterward, the software will check whether the stated command requires it to open or execute a particular application, if it needs to open an application, it will then proceed to execute the specified application. Afterward, it will check if the executed application can be further manipulated with the use of other commands. If so, the user may voice out specific commands that can be done in the application that is currently opened. If not, then the software will proceed to stand-by and wait for further commands. If the user chose to close the specified application or did not specify commands that the application can execute, then the software will revert to its initial state and begin waiting for generally used commands, not just application-specific ones.

If the initial command voiced out by the user did not correspond to any application at all, it simply pertains to commands for interactivity purposes or manipulating the other features of the computer. The software will then proceed to recognize the specified command, execute it, and act based on what command has been specified.

B. ECHO Class Diagram

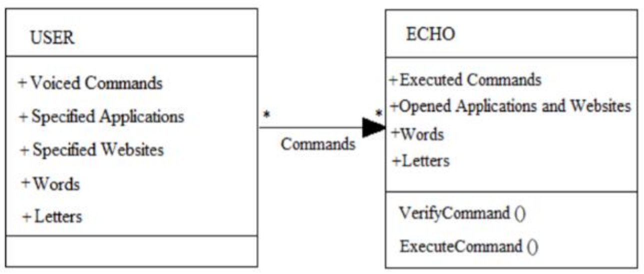

Figure 2. Class diagram of ECHO

Illustrated above in Figure 2. is ECHO's class diagram that depicts the structure of the user and the system interaction. The classes contain all the commands that are available for use. Both the user and ECHO classes are considered public. The user class is associated with ECHO through predetermined commands. Both the user and ECHO reference to each other in various instances. ECHO's classes are mostly comprised of the outputs, which are the executed commands, based on what the user commanded it to do. If it receives the command, it will proceed to verify and execute it first before actually returning an output.

C. Speech Recognition System Architecture

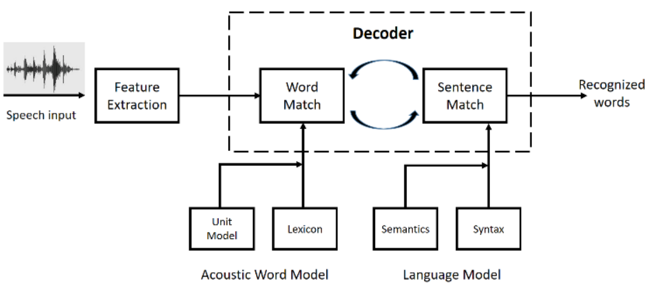

Figure 3. Speech Recognition System Architecture

Figure 3. shows the block diagram of a typical integrated continuous speech recognition system available in the market. Surprisingly, this diagram can work with any speech recognition system that has been developed. The feature extraction module provides the acoustic feature vectors used to characterize the spectral properties of the time-varying speech signal. The word acoustic match module checks the similarities between the input speech and the set of acoustic word models in the vocabulary to determine which words were mostly to match [18].

The sentence matching module uses the language model comprising of the syntax and semantics models to determine the sequence of words in the input speech. Syntax and semantic rules are specified manually. The decoder block searches the best match of word sequence for the input acoustic features based on the acoustic model, lexicon and language model [19].

D. Basic Speech Recognition Processes

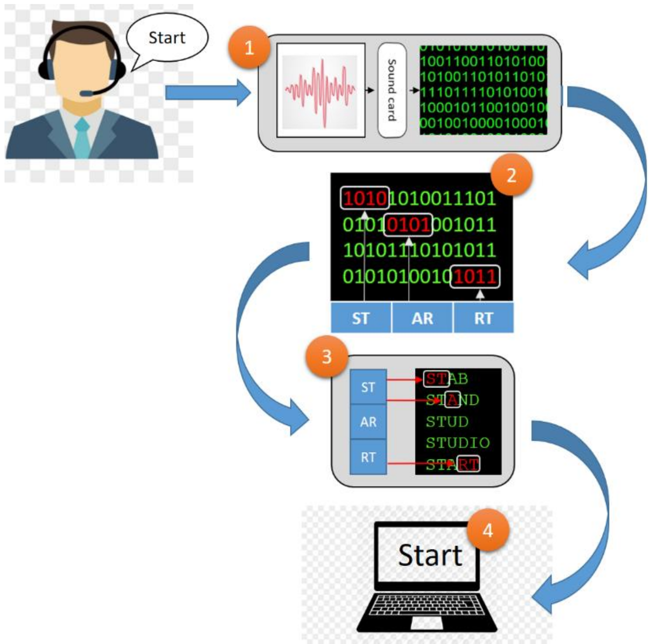

Figure 4. Basic Speech Recognition Processes

Initially as illustrated in Figure 4, the spoken command (i.e. "Start") is received by the microphone for better data capturing. The sound card inside a computer system converts the analog waves into its equivalent digital format. The software acoustical model breaks the word into phonemes which is any of the perceptually distinct units of sound in a specified language that distinguish one word from another [20] (i.e. ST AR RT). The software Language Model compares the phonemes to words in a built-in dictionary. The software decodes what it thinks the spoken words were and display the best match [21] and [22].

IV. Results and Discussion

The testing and evaluation phase are where the software's functions and components were verified. The results were evaluated in this part, to identify what parts were fully functional and what needs improvement.

A. Test Results

This is the overview of data that was gathered during the testing phase. The tests conducted were functionality testing, compatibility testing, and stress testing.

TABLE 1. Summarized Test Result for ECHO

System Testing	Pass	Fail	Test Cases	Percentage
Functionality Testing	91	5	96	94.79%
Compatibility Testing	12	0	12	100%
Stress Testing	5	0	5	100%

The summarized result for the testing phase of ECHO – Hands-free Computer Interaction using Speech Recognition System for the debilitated conducted by the technical adviser and other developers was shown in Table 1. The functionality testing had a total of 96 test cases and was able to pass 91 out of the 96 test cases garnering a percentage of 94.79%. As expected, the test case that failed during testing was the functions related to the dictation mode, which could be considered as the weakness of the system due to it not utilizing the neural network or the Internet for its vocabulary. Other test cases that failed were associated with the level of interactions, where the software failed to respond to some of the commands that were specified. Another notable function that failed was the commands 'Echo, Show Yourself' and 'Echo, Hide Yourself'. The compatibility testing had a total of 12 test cases that all passed in testing. Although it had a 100% percentage, some of the commands and features did not function when Windows 7 and 8 OS were used. It was confirmed to work best with Windows 10. This can be attributed to the portability issue. The stress testing passed 5 out of 5 test cases, garnering a percentage of 100%. The software was able to work even in areas with some tolerable noise but advisable to be used in some noise-free surroundings. In moderately busy areas, it is also able to function well with minimal errors encountered. In extremely busy or loud areas, the software had a hard time recognizing commands but was resolved by the usage of a microphone and a speaker.

B. Evaluation Results

This section provided an overview of the evaluation results that were gathered through the evaluation instrument that was used to evaluate the software.

TABLE 2. Over-All Summary of Evaluation

Level of Acceptability	Mean	SD	Interpretation
Functionality	3.52	0.55	Highly acceptable
Reliability	3.64	0.54	Highly acceptable
Usability	3.65	0.47	Highly acceptable
Efficiency	3.58	0.47	Highly acceptable
Maintainability	3.58	0.52	Highly acceptable
Portability	3.69	0.46	Highly acceptable

The Over-all result of evaluation as shown in Table 2 that Portability got the first rank because the system is capable of adaption to different environments like different Windows OS. Next is maintainability which is capable of diagnosis of deficiencies or causes of failures. The system is valid according to the client's needs. The 3.5th was reliability and usability. The system can maintain a specified level of performance in case of software faults. Next was Efficiency because the system can bear user response on using the system with an ample amount of time and effort. Lastly, the functionality, although it is on the last rank, it is still interpreted as Highly Acceptable, because voice command can hinder if the pronunciation is not accurate or if the place is loud.

V. Summary of Findings

The following are the list of findings and information that were gathered based on the results of the evaluation that was conducted: In terms of system functionality, the respondents were satisfied with the functions of the software, being rated as "Highly Acceptable" with a mean of 3.48 because it was able to perform the tasks it was commanded to do accurately and efficiently. The evaluators rated the software "Highly Acceptable" with a mean of 3.56 which meant that the software was able to maintain its performance regardless of the situation and re-establish its level of performance in case of system failure. When it comes to usability, this criterion was rated "Highly Acceptable" with a mean of 3.56 because it was easy-to-learn and it did not have further complications upon use. With regards to efficiency, the software was rated efficient and handles resources well, thus, the evaluators rated as "Highly Acceptable" with a mean of 3.52. The maintainability of the software was rated as "Highly Acceptable" with a mean of 3.60 in this criterion because it was able to adapt based on the user's needs and diagnose errors caused by failures. About the portability of the system, the evaluators rated as "Highly Acceptable" with a mean of 3.67 because the software was able to function when used in different environments and could be ported to other devices as well.

VI. CONCLUSION

This paper entitled ECHO – Hands-free Computer Interaction using Speech Recognition System for the Debilitated was able to implement the Speech recognition system to interact with the computer. A person with a physical limitation can manage desktop applications, navigate through different computer functions and features, manipulate Media Player functions, like playing or stopping what is currently being played, fast-forwarding, going into full-screen or normal screen, and hide and unhide tools boxes. Users can also take Screenshots using voice commands and have them saved in a convenient location. The use of voice commands in opening certain websites is very effective. A pseudo-registration feature allows the software to address the user using the appropriate name or gender that was entered. The dictation feature worked well especially with the use of a microphone. ECHO performed well with other commercially available software applications. The system was tested using Functional testing, Stress testing, and Compatibility testing, and got the result of 94.79% for the functional, 100% for stress, and 100% for compatibility. The system was evaluated using the ISO/IEC 9126-1 with 45 respondents, garnering an overall mean of 3.57 and a standard deviation of 0.52, interpreted as "Highly Acceptable".

Retrieval Number B1017078219/19©BEIESP
DOI: 10.35940/ijrte.B1017.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

REFERENCES

1. B. Cui, and T. Xue, "Design and realization of an intelligent access control system based on voice recognition," 2009 ISECS International Colloquium on Computing, Communication, Control, and Management, Sanya, 2009, pp. 229-232.
2. B. Azvine, N. Azarmi, and D. Nauck, 2006. Intelligent Systems and Soft Computing: Prospects, Tools and Applications. Germany: Springer
3. S. A. R. Che Yong Yeo, R. Al-Haddad and C. K. Ng, "Animal voice recognition for identification (ID) detection system," 2011 IEEE 7th International Colloquium on Signal Processing and its Applications, Penang, 2011, pp. 198-201.
4. A. Paul, M. Panja, M. Bagchi, N. Das, R.M. Mazumder, and S. Ghosh, "Voice recognition based wireless room automation system," 2016 International Conference on Intelligent Control Power and Instrumentation (ICICPI), Kolkata, 2016, pp. 84-88.
5. H. AlShu'eili, G. S. Gupta and S. Mukhopadhyay, "Voice recognition based wireless home automation system," 2011 4th International Conference on Mechatronics (ICOM), Kuala Lumpur, 2011, pp. 1-6.
6. N. Aktar, I. Jaharr and B. Lala, "Voice Recognition based intelligent Wheelchair and GPS Tracking System," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-6.
7. M. Matić, I. Stefanović, U. Radosavac, and M. Vidaković, "Challenges of integrating smart home automation with cloud-based voice recognition systems," 2017 IEEE 7th International Conference on Consumer Electronics - Berlin (ICCE-Berlin), Berlin, 2017, pp. 248-249.
8. T. Athanaselis, S. Bakamidis, G. Giannopoulos, I. Dologlou, and E. Fotinea, "Robust speech recognition in the presence of noise using medical data," 2008 IEEE International Workshop on Imaging Systems and Techniques, Crete, 2008, pp. 349-352.
9. M. Hawley, et al., "A Voice-Input Voice-Output Communication Aid for People With Severe Speech Impairment," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 21, no. 1, pp. 23-31, Jan. 2013.
10. G. S. Meltzner, J. T. Heaton, Y. Deng, G. De Luca, S. H. Roy, and J. C. Kline, "Silent Speech Recognition as an Alternative Communication Device for Persons With Laryngectomy," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 12, pp. 2386-2398, Dec. 2017.
11. S. Babic, T. Orehovacki and D. Etinger, "Perceived user experience and performance of intelligent personal assistants employed in higher education settings," 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, 2018, pp. 0830-0834.
12. B. Byrne B, M. Finke, S. Khudanpur, J. McDonough, H.M. Nock, H. M. Riley, M. Saraclar, C. Wooters, and Zavaliagkos, G., "Pronunciation modelling for conversational speech recognition: a status report from WS97," 1997 IEEE Workshop on Automatic Speech Recognition and Understanding Proceedings, Santa Barbara, CA, USA, 1997, pp. 26-33.
13. D. Poole, A. K. Mackworth, 2010. Artificial Intelligence: Foundations of Computational Agents. USA: Cambridge University Press
14. W. Tsai, Y. Lian, S. Hsu, Q. Zheng, Y. Su and J. Chen, "An Implementation of Voice Recognition and Control System for Electric Equipment," 2018 International Symposium on Computer, Consumer and Control (IS3C), Taichung, Taiwan, 2018, pp. 356-359.
15. F. Giunchiglia, J. Odell & G. Weiss, 2003. Agent-Oriented Software Engineering III. Germany: Springer
16. G. Czibula, A. M. Guran, I. G. Czibula, G. Cojocar, "IPA – An Intelligent Personal Assistant Agent for Task Performance Support". Intelligent Computer Communication and Processing, 2009. ICCP
17. A. F. Khalifeh, K. A. Darabkh, and A. Kamel, "Performance evaluation of Voice-Controlled Online Systems," International Multi-Conference on Systems, Signals & Devices, Chemnitz, 2012, pp. 1-6.
18. K. Myers, P. Berry, J. Blythe, K. Conley, M. Gervasio, D. McGuinness, D. Morley, A. Pfenner, M. Pollack and M. Tamble, "An Intelligent Personal Assistant for Task and Time Management". Association for the Advancement of Artificial Intelligence, 2016
19. H. Lee, S. Chang, D. Yook and Y. Kim, "A voice trigger system using keyword and speaker recognition for mobile devices," in IEEE Transactions on Consumer Electronics, vol. 55, no. 4, pp. 2377-2384, November 2009. doi: 10.1109/TCE.2009.5373813
20. Y. Zhang, H. Ghenniwa and W. Shen, "Agent-Based Assistance in Collaborative Design Environments". Dept. of Electrical and Computer Engineering, The University of Western Ontario, London, Ontario, Canada, Integrated Manufacturing Technologies Institute, National Research Council Canada, London, Ontario, Canada, 2006.
21. W. Astuti and E. B. W. Riyandwita, "Intelligent automatic starting engine based on voice recognition system," 2016 IEEE Student Conference on Research and Development (SCOReD), Kuala Lumpur, 2016, pp. 1-5. doi: 10.1109/SCORED.2016.7810061
22. N. Chadha, R.C Gangwar and R. Bedi, "Current Challenges and Application of Speech Recognition Process using Natural Language Processing: A Survey", International Journal of Computer Applications (0975 – 8887) Volume 131 – No.11, December 2015

AUTHORS PROFILE

Elmer C. Matel, a professor at the Lyceum of the Philippines University – Cavite Campus is a graduate of Bachelor of Science in Computer Science and has completed his Master in Information Technology at the Technological Institute of the Philippines – Quezon City. He is currently the Manager of the Center for Technology-Enabled Education (e-Learning).

Herchel Aquines, a professor at the Lyceum of the Philippines University – Cavite campus is a graduate of Bachelor of Science in Information Technology at the Technological University of the Philippines - Manila. She is currently the Coordinator of the Department of Computer Studies at the Lyceum of the Philippines University.

Carl Brian S. de Guzman is a graduate of Bachelor of Science in Information Technology at the Lyceum of the Philippine University – Cavite campus. He had his On-The-Job Training at Technomotion Controls International where he worked in the MIS department as a Technical Support Staff.

Jasond Isaac V. Gustilo is a graduate of Bachelor of Science in Information Technology at the Lyceum of the Philippines University – Cavite campus. He is planning to pursue his career related to system analysis. He had his On-The-Job training at the Sangguniang Panlalawigan ng Cavite where he worked as technical support personnel and graphic designer.

Gregg Victor P. Ibera is a graduate of Bachelor of Science in Information Technology at the Lyceum of the Philippines University – Cavite campus. He is planning to pursue a career as a UX Designer to further improve his skills and knowledge. He had his On-The-Job training in the Sangguniang Panlalawigan ng Cavite where he worked as technical support staff and designer.

Priam D. Jodilla is a graduate of Bachelor of Science in Information Technology at the Lyceum of the Philippines University – Cavite campus. He was trained as an intern in the Sangguniang Panlalawigan ng Cavite where he worked as technical support staff and hardware troubleshooting tasks.