TaAl-Lip: Application of Artificial Neural Network in Recognition of Select Tagalog Alphabet through Lip Reading

Oliver M. Membrere, Reynaldo M. Supan, Jackilyn C. Magtoto, Miles Kevin B. Galario, Benilda Eleonor V. Comendador, and Ranil M. Montaril

***Abstract*—The paper introduces TaAl–Lip: Application of Artificial Neural Network in Recognition of Select Tagalog Alphabet through Lip Reading. It is an experimental study designed to determine the accuracy of the recognition of select Tagalog alphabet or also known as the ABAKADA Alphabet. It utilized Artificial Neural Network (ANN), Computer Vision (CV), Macropixelling, and Image Processing (IP) to develop a tool that can recognize mouth's movement by means of lip reading.**

After the tool was developed the degree of accuracy of the recognition of the application was evaluated by the proponents according to the: (a) light orientation; (b) viewing angle and (c) the user's distance from the camera. Based on the experiment conducted, the researchers concluded that the mouth's movement is most recognizable with a front side light orientation with an average of 70.34%.

***Index Terms*—Lip reading, artificial neural network, macropixelling, image processing, computer vision.**

I. Introduction

There are two aspects of human speech: audio and visual. The audio speech signal refers to the acoustic that uses the articulatory organs together with the muscles a speaker produce speech. The problem of acoustic signal is that the accuracy of speech recognition technique is not good enough in noisy condition. [1] The process of identifying the words or letters from the movement of lips is called lip-reading. Lip-reading can solve this problem because lip-reading uses only visual features, which are not affected by noise. The lip movements contain enough information for the visual categorization of speech but these features vary from one language to another because the pronunciation is different for different languages.

As for spoken language learning for hearing impaired people, aside from residual listening, lip reading is an important channel to understand the information. Lip reading or speech reading is a tool applicable for person with hearing disability. This is a way of understanding speeches by interpreting the movement of the lips and tongue when there is noise interference or normal sound is not available. With the complexity of the Artificial Neural Network (ANN), nowadays, many computer systems use ANN for it is developed with a systematic step-by-step procedure which optimizes a criterion commonly known as the learning rule.

Artificial Neural Network refers to computing systems whose central theme is borrowed from the analogy of biological neural networks. The use of the term "Artificial" is to provide an imitation of the real thing by the use of computer technology. [2] A neural network's ability to perform computation is based on hope that we can reproduce some of the flexibility and power of human brain by artificial means. Basically, [3] a neural network is machined that is designed to model the way in which the brain performs a particular task or function of interest. The network is usually implemented by using electronic components or is simulated in software on a digital computer.

This paper is closely related to the existing study entitled "Lip-Reading using Neural Networks" wherein [4] the researchers of this study implemented this project using an evaluation version of the software NeuroSolutions5. They have used the evaluation version of the software; they got the maximum accuracy of 52%. This research project is the first attempts to use Neural Networks classification (clustering) for addressing this challenging problem that combines two different application domains of classification and predicting which brings out the much desired output. However, none of these include the effect of the variable extraction of the lips, distance, viewing angle and the light orientation while recognizing the speech.

The developed tool is an application that recognizes the select alphabet by interpreting the movement of the lips and tongue where noise interference and normal sound is a not a factor. [5] Filipinos speak nine more indigenous languages all belonging to the Malayo-Polynesian group namely: Tagalog, Cebuano, Ilocano, Ilonggo, Bikolano, Waray, Visayan, Pampango and Pangasinan. Each of the nine has a number of dialects; hence, there are 87 dialects in all. Some dialects of the same language are mutually unintelligible to each other. Each of the nine languages has its own extensive literature. The oldest and the richest is Tagalog, the language extensively used in Central and Southern Luzon, and considered the basis of the national language of the Philippines. The ABAKADA alphabet is the traditional Filipino alphabet that is being used as guide in writing and speaking of Filipino words. Since the said alphabet is the starting point of learners herein the Philippines, the researchers decided to make a lip reading system using this

Manuscript received December 29, 2013; revised June 20, 2014. This work was supported in part by the Polytechnic University of the Philippines.

The authors are with the College of Computer and Information Sciences, Polytechnic University of the Philippines, Sta. Mesa, Philippines (e-mail: olivermembrere@yahoo.com, reysupan7@yahoo.com, jackilynmagtoto@yahoo.com, Miles_kev21@yahoo.com, bennycomendador@yahoo.com, rmmontaril@yahoo.com).

DOI: 10.7763/JACN.2014.V2.128

language as the basis of the training ground for the experiment. The researchers used algorithms such as ANN and macropixelling algorithm and processes to be able to come up with the desired variables presented in this paper. [6] Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an expert in the category of information it has been given to analyze. Neural network is applied in Lip Reading, one of the easiest ways to recognize the speech. It is one of the latest techniques widely preferred for speech recognition. [7] We describe a lip reading system that uses both, shape information from the lip contours and intensity information from the mouth area. (Shape information is obtained by tracking and parameterizing the inner and outer lip boundary in an image sequence. Intensity information is extracted from a grey level model, based on principal component analysis. [8] In comparison to other approaches, the intensity area deforms with the shape model to ensure that similar object features are represented after non-rigid deformation of the lips. We describe speaker independent recognition experiments based on these features.

II. The Developed System

The system is developed using MATLAB and Microsoft Visual C#. It is not a web-based application therefore, it is intended to run on standalone computer. Since it is an experimental study, it will only yield on the ABAKADA alphabet and the researcher alone will be the respondent to evaluate the developed application.

A. System Architecture

In the initialization of the system, it will capture a video of the speaker by using a webcam. The captured video will undergo the process of image acquisition; this process will acquire the image frames from the video. The images will be pre-processed; cropping, gray-scaling and noise reduction will be implemented. Then the pre-processed images will be converted to image matrices. The system will then manipulate the gathered data and will compare it to the trained data using the Artificial Neural Network algorithm to give the optimized solution. Once the process was done it will give the optimized output selected as recognized word based on the A-BA-KA-DA and will display the frames of images of the mouth of the speaker. The System Architecture is illustrated in Fig. 1.

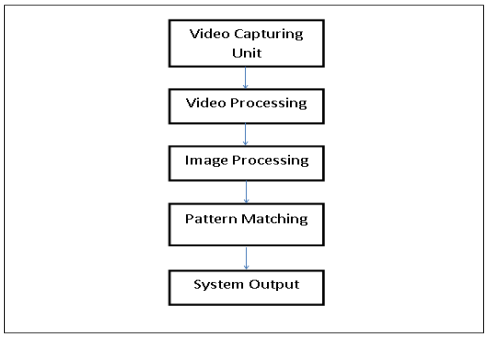

Fig. 1. System architecture.

Fig. 2 shows the composition of image processing architecture. The sequence of images captured from the video will be the input for image processing. This unit includes techniques in image processing or editing like recognition of point of interest, cropping, gray-scaling and noise reduction of the images. Then the pre-processed images will be generated to image matrices.

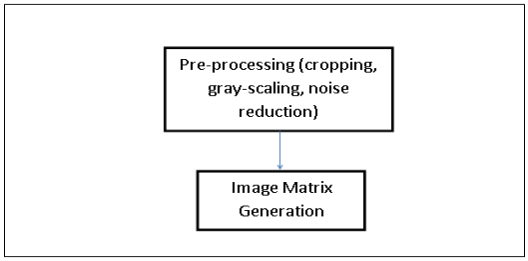

Fig. 2. Image processing architecture.

Fig. 3 shows the artificial neural network architecture. The image matrices are the input and will undergo the artificial neural network algorithm that will determine if the pattern generated is matched to the trained data. The process will output the matched data recognized by the system.

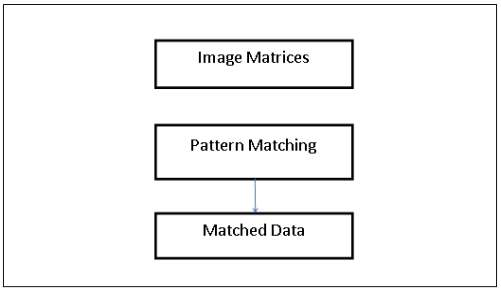

Fig. 3.Artificial neural network/ Pattern matching architecture.

B. Software

The researcher provides an interface wherein the user can interact with the system. Fig. 4 shows the prototype of the interface where there is only one option on it.

Fig. 4. System interface.

The aim of the system is to capture the lips and identify the letter that is been mentioned.

C. Research Methods and Techniques

In Ref. [9] an experimental method is an attempt by the researcher to maintain control over all factors that may affect the result of an experiment it is cause and effect relationship based on the manipulated variables. In doing this, the researcher attempts to determine or predict what may occur. This method is design to test the degree of accuracy in recognition of the Tagalog Alphabet.

The researchers gathered data through experiments. In each

experiment, there is controlled variable in the lip-reading process. These experiments determine the sample amount different parameters to be used. The following are the amount of each independent variable used by the proponents. [10] RSS is used to calculate the aggregate accuracy of the measurement when the accuracies of the all measuring devices are known. The average accuracy is not merely the arithmetic average of the accuracies (or uncertainties), nor is it the sum of them. The following is the formula for computing the RSS.

$$\mathrm{RSS} = \sqrt{\sum x^2}$$

where x = individual average per parameter, and mean

$$\overline{x} = \frac{\sum x}{n}$$

where n = total number of trials, x = individual results of trials.

III. Result and Conclusion

After the tool was developed the degree of accuracy of the recognition of the application was evaluated by the proponents according to (a) light orientation; (b) viewing angle and (c) the user's distance from the camera. The light orientation determines the shadow it creates and how it is affected by a particular light source. It also identifies an object's shadow as the region from which a light source is not visible because the object obstructs the line of sight.

Table I shows the Summary of findings in experiment paper for varying light orientation. It illustrates that the mouth's movement are most recognizable with front side light orientation with an average of 69.66%, followed by 45 ° right hand side light orientation with an average of 69.42% and lastly by 45 °left hand side light orientation with 68.99%. These findings only show that the most ideal light orientation to use for the system to get the highest accuracy of recognition is the front side.

On the other hand, the viewing angle assesses the degree of accuracy in the recognition of Select Tagalog Alphabet through lip reading with respect to the viewing angle. We tested the tool and set the independent parameters of distance to 60cm, light orientation to front and answered the formulated experiment paper. The proponents chose 45 degrees left side view, front view and 45 degrees right side view as the various settings for the viewing angle to test with.

The researchers summed up the precision rate of the tool for matched, mismatched and unrecognized outputs on the different viewing angle settings. The results of the tests are shown it Table I. It depicts that mouth's movement are most recognizable with front side light orientation with an average of 69.66%, followed by 45 ° right hand side light orientation with an average of 69.42% and lastly by 45 ° left hand side light orientation with 68.99%. These findings only show that the most ideal light orientation to use for the system to get the highest accuracy of recognition is the front side.

TABLE I: Summary of Findings in Experiment Paper for Varying Light Orientation

Alphabet	Accuracy Rate (%) 45 °Left Hand Side	Accuracy Rate (%) Front Side	Accuracy Rate(%)45 °Right Hand Side
A	70	78.89	76.67
E	88.89	76.67	82.22
I	77.78	84.44	77.78
O	75.56	88.89	78.89
U	77.87	77.78	70.00
Ba	68.89	82.22	70.00
Be	66.67	63.33	67.78
Bi	62.22	63.33	58.89
Bo	73.33	75.56	84.44
Bu	63.33	63.33	64.44
Da	64.44	60.00	67.78
De	67.78	68.89	58.89
Di	64.44	83.33	63.33
Do	68.89	83.33	63.33
Du	71.11	71.11	68.89
Wa	68.89	64.44	66.67
We	70.00	53.33	70.00
Wi	67.78	54.44	72.22
Wo	61.11	57.78	66.67
Ya	64.44	63.33	65.56
Ye	65.56	61.11	66.67
Yo	65.56	68.89	67.78
Yu	62.22	60	67.78
Average	68.99	69.66	69.42

A. Viewing Angle

TABLE II: Summary of Findings in Experiment Paper for Varying Viewing Angle

Alphabet	Accuracy Rate (%) 60 °Left Hand Side	Accuracy Rate (%) Front Side	Accuracy Rate (%) 60 °Right Hand Side
A	77.78	81.11	71.11
E	71.11	77.78	65.56
I	70	81.11	66.67
O	85.56	78.89	71.11
U	63.33	76.67	84.44
Ba	77.78	82.22	71.11
Be	66.67	63.33	77.78
Bi	67.78	64.44	70.00
Bo	68.89	75.56	81.11
Bu	70	61.11	66.67
Da	66.67	60.00	70.00
De	66.67	68.89	70.00
Di	70	83.33	72.22
Do	75.56	90.0	73.33
Du	70.00	71.11	68.89
Wa	63.33	74.44	66.67
We	65.56	53.33	66.67
Wi	63.33	54.44	67.78
Wo	68.89	57.78	67.78
Ya	72.22	63.33	67.78
Ye	70.00	66.67	66.67
Yo	64.44	68.89	74.44
Yu	65.56	60.00	72.22
Average	69.11	70.19	70.87

To assess the degree of accuracy in the recognition of the developed tool with respect to the viewing angle, we tested the tool and set the independent parameters of distance to 60cm, light orientation to front and answered the proposed experiment paper. The researchers chose 45 degrees left side

view, front view and 45 degrees right side view as the various settings for the viewing angle to test with.

The researchers summed up the precision rate of the tool for matched, mismatched and unrecognized outputs on the different viewing angle settings as shown in Table II.

The Table II shows that mouth's movement is most recognizable with right side view viewing angle with an average of 70.87%, followed by 60 ° front view viewing angle with an average of 70.19% and lastly by 60 ° left side view viewing angle with 69.11%. These findings only show that the most ideal viewing to use for the system to get the highest accuracy of recognition is the 60 ° left side viewing angle.

B. Distance of the Speaker

To assess the degree of accuracy in the recognition of the developed tool with respect to the distance of the speaker, we tested the tool and set the independent parameters of light orientation to front, viewing angle to front view and answered the proposed experiment paper. The researchers chose 50cm, 60cm and 70 cm as the various settings for the distance of the speaker to test with.

The researchers summed up the precision rate of the tool for matched, mismatched and unrecognized outputs on the different distance of the speaker settings. The results of the tests are shown in Table III.

TABLE III: SUMMARY OF FINDINGS IN EXPERIMENT PAPER FOR VARYING DISTANCE OF THE SPEAKER

Alphabet	Accuracy Rate (%) 50 centimeters	Accuracy Rate (%) 60 centimetres	Accuracy Rate (%) 70 centimetres
A	83.33	85.56	74.44
E	80	83.33	83.33
I	90	77.78	82.22
O	90	83.33	86.67
U	80	84.44	71.11
Ba	81.11	82.22	73.33
Be	70	65.56	62.22
Bi	68.89	67.78	56.67
Bo	80	82.22	70
Bu	67.78	73.33	73.33
Da	81.11	65.56	67.78
De	70	68.89	60
Di	83.33	60	58.89
Do	75.56	71.11	72.22
Du	76.67	76.67	64.44
Wa	74.44	64.44	62.22
We	64.44	58.89	60
Wi	67.78	60	61.11
Wo	70	65.56	71.11
Ya	63.33	70	63.33
Ye	66.67	58.89	56.67
Yo	68.89	56.67	63.33
Yu	66.67	56.67	67.78
Average	74.78	70.39	67.92

The Table III shows that mouth's movement is most recognizable with 50cm distance with an average of 74.78%, followed by 60cm distance with an average of 70.39% and lastly by 70cm distance with 67.92%. These findings only show that the most ideal distance to use for the system to get the highest accuracy of recognition is the 50cm.

Table IV summarized the findings in experiment paper for varying viewing angle The table shows that mouth's movement is most recognizable with right side viewing angle with an average of 70.87%, followed by 60 ° front viewing angle with an average of 70.19% and lastly by 60 ° left side viewing angle with 69.11%. These findings only show that the most ideal viewing to use for the system to get the highest accuracy of recognition is the 60 ° left side viewing angle.

TABLE IV: SUMMARY OF FINDINGS IN EXPERIMENT PAPER FOR VARYING VIEWING ANGLE

Alphabet	Accuracy Rate (%) 60 ° Left Hand Side	Accuracy Rate (%) Front Side	Accuracy Rate (%) 60 ° Right Hand Side
A	77.78	81.11	71.11
E	71.11	77.78	65.56
I	70	81.11	66.67
O	85.56	78.89	71.11
U	63.33	76.67	84.44
Ba	77.78	82.22	71.11
Be	66.67	63.33	77.78
Bi	67.78	64.44	70.00
Bo	68.89	75.56	81.11
Bu	70	61.11	66.67
Da	66.67	60.00	70.00
De	66.67	68.89	70.00
Di	70	83.33	72.22
Do	75.56	90.0	73.33
Du	70.00	71.11	68.89
Wa	63.33	74.44	66.67
We	65.56	53.33	66.67
Wi	63.33	54.44	67.78
Wo	68.89	57.78	67.78
Ya	72.22	63.33	67.78
Ye	70.00	66.67	66.67
Yo	64.44	68.89	74.44
Yu	65.56	60.00	72.22
Average	69.11	70.19	70.87

TABLE V: SUMMARY OF PROPONENTS' FINDINGS WITH REGARDS TO VARYING DISTANCE OF THE SPEAKER TO THE CAMERA

Alphabet	Accuracy Rate (%) 50 centimetres	Accuracy Rate (%) 60 centimetres	Accuracy Rate (%) 70 centimetres
A	83.33	85.56	74.44
E	80	83.33	83.33
I	90	77.78	82.22
O	90	83.33	86.67
U	80	84.44	71.11
Ba	81.11	82.22	73.33
Be	70	65.56	62.22
Bi	68.89	67.78	56.67
Bo	80	82.22	70
Bu	67.78	73.33	73.33
Da	81.11	65.56	67.78
De	70	68.89	60
Di	83.33	60	58.89
Do	75.56	71.11	72.22
Du	76.67	76.67	64.44
Wa	74.44	64.44	62.22
We	64.44	58.89	60
Wi	67.78	60	61.11
Wo	70	65.56	71.11
Ya	63.33	70	63.33
Ye	66.67	58.89	56.67
Yo	68.89	56.67	63.33
Yu	66.67	56.67	67.78
Average	74.78	70.39	67.92

Apparently, the distance of the speaker assess the degree of accuracy in the recognition of Tagalog alphabet through lip reading with respect to the distance of the speaker, we tested the tool and set the independent parameters of light orientation to front, viewing angle to front view and answered the proposed experiment paper. The researchers chose 50cm, 60cm and 70 cm as the various settings for the distance of the speaker to test with.

The proponents summed up the precision rate of the tool for matched, mismatched and unrecognized outputs on the different distance of the speaker settings. The results of the tests are shown below.

Table V summarizes the findings with regards to varying distance of the speaker to the camera. It shows that the mouth's movement is most recognizable with 50cm distance with an average of 74.78%, followed by 60cm distance with an average of 70.39% and lastly by 70cm distance with 67.92%. These findings only show that the most ideal distance to use for the system to get the highest accuracy of recognition is the 50cm.

IV. Conclusion

Based on the thorough and critical analysis of the implemented and evaluated study the researchers concluded that mouth's movement is most recognizable with a front side light orientation with an average of 70.34%. It is followed by 45 ° right hand side light orientation with an average of 69.28% and lastly by 45 °left hand side light orientation with 69.13%. However, the mouth's movement is most recognizable with a right side view viewing angle with an average of 70.87%. Then followed by 60 °front viewing angle with an average of 70.19% and lastly by 60 °left side viewing angle with 69.11%. On the contrary, the mouth's movement is most recognizable with a 50cm distance with an average of 74.78%, followed by 60cm distance with an average of 70.39% and lastly by 70cm distance with 67.92%. During the tool implementation, the proponents discovered that the output of the tool may vary due to different environment settings such as the amount of light luminance. Furthermore, the chosen algorithms used by the proponents to develop the tool was able to properly recognized the pattern of the lip's (mouth's) movement.

V. Recommendation and Future Works

In the future, the researchers would like to extend their work by using other algorithms for the better performance and higher accuracy of the system. They will add movement of the tongue and mouth for higher accuracy in lip reading. Moreover, they will develop a real-time application for Lip reading, with ABAKADA alphabet as spoken language and they will train more samples of mouth in order to get more patterns for comparison in recognition of the images.

Through this study, it was proven that lip reading can be a new way of recognizing speech using computer vision and ANN. Thus, recommending the future researchers to use other methods and algorithms to maximize the performance of the system.

Reference

[1] M. Tamal, "Recognition of vowels from facial images using lip reading technique," Jadavpur University, Kolkata, 2010.

[2] K. Mehrotra *et al*., "Elements of artificial neural networks," Library of Congress Cataloging-in-Publication Data, Massachusetts Institute of Technology, 1997.

[3] S. Nagabhushana, *Computer Vision and Image Processing*, New Age International (P) Ltd., 1st ed. 2005.

[4] A. Bagai, H. Gandhi *et al.*, "Lip-Reading using neural networks," *Int. Journal of Computer Science and Network*, vol. 9, no. 4, 2009.

[5] Filipino Alphabet. (September 18, 2013). [Online]. Available: http://tagaloglang.com/The-Philippines/Language/modern-filipino-alphabet.html

[6] Artificial Neural Network. (September 12, 2013). [Online]. Available: http://www.ukessays.com/essays/computer-science/lip-reading-using-neural-networks-computer-science-essay.php

[7] J. Tebelskis, "Speech recognition using neural networks," School of Computer Science, Carnegie Mellon University, May 1995.

[8] T. W. Lewis and D. M. W. Powers, "Lip feature extraction using red exclusion," Department of Computer Science, School of Informatics and Engineering, Flinders University of South Africa.

[9] Calmorin, L. Paler, and M. A. Calmorin, *Research Methods and Thesis Writing*, 2nd ed. Sampaloc, Manila: Rex Bookstore, 2007.

[10] Root sum square. (September 13, 2013). [Online]. Available: http://klabs.org/richcontent/General_Application_Notes/SDE/RSS.pdf

Benilda Eleonor V. Comendador was a grantee of the Japanese Grant Aid for Human Resource Development Scholarship (JDS) from April 2008 to September 2010. She obtained her master of science in global information telecommunication studies, majored in project research at Waseda University, Tokyo Japan in 2010. She was commended for her exemplary performance in completing the said degree from JDS. She finished her master of science in information technology at Ateneo Information Technology Institute, Philippines in 2002.

Presently, she is the chief of the Open University Learning Management System and the chairperson of the master of science in information technology of the Graduate School of the Polytechnic University of the Philippines. She is an assistant professor and was the former chairperson of the Department of Information Technology of the College of Computer Management and Information Technology of PUP.

Ranil M. Montaril received his undergraduate degree in electronics and communications engineering at the Polytechnic University of the Philippines, Manila, Philippines in 2004 and continued with an MS degree in electronics and communications engineering from the Bulacan State University, Philippines in 2008. His research interests include digital signal processing, power electronics, robotics, computational intelligence and evolutionary computation. He also joined Emerson Network Power in 2007 as an electrical design engineer up to the present.

Oliver M. Membrere is a senior student from Quezon City currently is taking up BS degree in computer science at the Polytechnic University of the Philippines. He is knowledgeable in various programming language (C, Java, C#), database programming (SQL, MySQL, MS Access) and web programming (PHP, HTML, Javascript, CSS). He worked as a student trainee at Optimize Solution, Pasig City for the on-the-job training required by the college and accomplished 200 hours of work.

Jackilyn C. Magtoto is a senior student from Tondo, Manila currently taking up BS degree in computer science at the Polytechnic University of the Philippines (PUP). She is knowledgeable in various programming language (C#, Java), advance networking (CCNA-Module 2) and web programming (PHP, HTML, Javascript, CSS). She worked as a student trainee at Synapse Practice Management- The

Healthcare Acapella Pasig City for the on-the-job training required by the college and accomplished 200 hours of work.

Miles Kevin B. Galario is a senior student from San Jose Del Monte, Bulacan. He is currently taking up BS degree in computer science at the Polytechnic University of the Philippines (PUP). He is knowledgeable in various programming language (C, Java, C#), database programming (SQL, MySQL, MS Access) and web programming (PHP, HTML, Javascript, CSS). He worked as a student trainee at Orion Solutions Inc., Philippines for the on-the-job training required by the college and accomplished 200 hours of work.

Reynaldo M. Supan is a senior student from Mexico, Pampanga. He is currently taking up BS degree in computer science at the Polytechnic University of the Philippines (PUP). He is knowledgeable in various programming language (C, Java, C#), database programming (SQL, MySQL, MS Access) and web programming (PHP, HTML, Javascript, CSS). He worked as a student trainee at Orion Solutions Inc., Philippines for the on-the-job training required by the college and accomplished 200 hours of work.

Cloud Technology and Application Development