ISSN 2347 - 3983
Volume 7, No. 9 September 2019
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter10792019.pdf
https://doi.org/10.30534/ijeter/2019/10792019

Filipino to Chinese Speech-to-speech Translator Using Neural Network with Database System

Mark Renier M. Bailon[1], Mark Albert C. De Silva[2], Reyann Jhorel P. Lapuz[3], John Larry A. Tinio[4], Ted Bryan K. Yu[5], and Reggie C. Gustilo[6]

[1]ECE Department, De La Salle University, Manila, Philippines, markrenierbailon@yahoo.com
[2]ECE Department, De La Salle University, Manila, Philippines, markalbertdesilva@yahoo.com
[3]ECE Department, De La Salle University, Manila, Philippines, rj_lapuz29@yahoo.com
[4]ECE Department, De La Salle University, Manila, Philippines, johnlarrytinio@yahoo.com
[5]ECE Department, De La Salle University, Manila, Philippines, Student, ted_bryan_yu@yahoo.com
[6]ECE Department, De La Salle University, Manila, Philippines, reggie.gustilo@dlsu.edu.ph

ABSTRACT

Technology nowadays becomes convenient because conveying information between computers and other electronic devices is done in a speech-to-speech environment through speech implementation. There are different technologies that can aid in speech implementation and one of which is the Neural Network System (NNS). A Neural Network System is composed of many simple processors combined together that recognizes pattern through series of trainings. The aim of the research is to produce computer software that uses an algorithm in order to translate Filipino speech into Chinese speech with the use of an artificial neural network aided by LabVIEW. The use of LabVIEW in implementing the whole speech-to-speech system organized the details through three stages: speech-to-text stage, text-to-text stage, and text-to-speech stage. The results showed that increasing the number of training of words increases the success rate of the system in recognizing the words. Some words used are nearly similar in pronunciation with other word received lower success rates as compared to unique words. Training has a success rate of 93.4% while testing yielded 91.59 percent. Speech implementation was conducted successfully that applied speech recognition, speech synthesis, and database system. The results of the speech-to-speech translation of Neural Network System in LabVIEW presented herewith such that they can be used as an aid to improvement of speech implementation.

Key Words – Neural Network System, LabVIEW, Database System, Speech Recognition, Training

1. INTRODUCTION

Advancements in computer technology has played a very important role in simplifying systems analysis and design whether they are used in robotics, medical technology or pure communications systems design [1], [2], [3], [4], [5], [6], [7]. Prior to actual systems implementation, the design and analysis is done using computer simulations and emulators which significantly reduced design cost and labor cost [8], [9], [10], [11], [12], [13].
The combination of computer technology, systems simulators and emulators are run by a large number of advanced intelligent and information systems that is multi-disciplinary in nature and has great impact and significance in all fields of information technology and engineering [14], [15], [16], [17], [18].

Throughout the years, convenience has been the aim of our country's technology. This is very evident and can be visibly seen in things like minimizing the usage of bigger components, transforming large machines into integrated circuits packed in smaller chips, and improving the speed of devices. Instead of typing words into the computer, an alternative way of conveying information into a certain device is through speech implementation.

Nowadays, there are different technologies that can aid in speech implementation. One of which is the Neural Network System (NNS). This Neural Network System is composed of many simple processors combined together to form the so-called system. Every single bit of this processor contains memories that are interconnected by communications channels, which carry data. [19] In line with this, an artificial neuron in 1943 was invented by Warren McCulloch, a neurophysiologist, and Walter Pits, a logician. Due to limited resources at that time, the producers were not able to have a great deal with the artificial neuron that they had invented.

Speech is simpler when it comes to relaying information or commanding a program to function. Our study is about artificial intelligence based on a Neural Network System which will be the core controller of the system. Basically, the study focuses on transferring input signal in the form of an analog voice waveform into the Neural Network. It will process the information by comparing the input analog waveform to the stored information in the system. Analog signals cannot be compared to digital signals directly, that is why the input signal is first converted into its equivalent digital component. After which, the digital equivalent signal of the input parameter will be compared to the information inside the data base. As the process continues, this digital waveform of signal will then be transformed again to an analog signal, which comprises the translated signal of the input analog waveform. For this to work, an algorithm will also be used to give commands that will translate the input signal to their desired destinations depending on their specific details.

The Neural Network basically needs to be trained in order for it to absorb specific words that are being conveyed by the user. By that, it will help the neural network to have a lesser chance of failing to acquire the desired input. More so, there is a bigger chance of getting the correct words regardless of the weight of their speech patterns. Training the Neutral Network would entail a lot of different speech patterns for it to practice its recognition on the limited words that it can process. This type of reinforcement learning does not only consider the speech of the person but also the environmental noise that it recognizes. When the background or environmental noise is greater than or having a high tone that would distract the system, it will then evaluate the speech pattern to be vague.

2. PROCESS FLOW

The study is separated into three stages: the Recording stage, the Gathering stage, and the Testing stage. In the recording stage, 25 voice variations of a certain word will be sampled and recorded. This will be done for 500 commonly used Filipino words. The recorded voice variations are programmed in the neural network system of LabVIEW to be able to convert the voice signal to a specific parameter that can be read by the system. For the gathering stage, 500 Filipino words will have a corresponding 500 Chinese words and these data will be stored in the database system of LabVIEW. This stage can also be called as the training stage since in this phase; the recorded spoken word is being absorbed by the Neural Network System for it to train itself to recognize the specific word regardless of their speech patterns and environmental noise.

The speech is recognized by a microphone which will relay its gathered information to the Neural Network. In the testing stage, the prototype will be run like in an actual set-up. 25 testers (people) will articulate a specific word in Filipino. The words articulated will then be converted into corresponding text in Chinese and is compared to the equivalent information in the database. The output equivalent which is a Chinese word is then played in the computer through the speakers.

With the translation, there are some Filipino words that have common translations in Chinese. This will not be a problem because the system made in LabVIEW can call more than one output corresponding to the input being called. In addition to these, the GUI will also aid the users in dealing with the translation. Using the existing GUI in LabVIEW (some of these are created as one constructs the terminal and connections); one can be assisted in the overall operation.

3. DESIGN CONSIDERATIONS

3.1 Speech-to-Text

In this stage, the system will be recognizing the voice as its input. Ideally, the recognized voice must be in Filipino so that it can have a desirable translation in Chinese depending upon its common usage. In a real scenario, any voice input can be fed to the system and there can be external factors that might hinder the system in recognizing the voice. One possibility is that the voice input is not listed in the database. Another will be the voice input is hardly audible because of the distance from the microphone.

3.1.1 Speech Recognizer

In order to help in the training stage, the Windows Speech Recognizer was used for a more accurate recognition of voice. This was first used in the training stage. As the speaker utters the 500 needed adjectives, the windows speech recognizer has the capacity to absorb the voice as accurate as possible because of its characteristics of having fairly high recognition accuracy and it also provides a set of commands that assists in dictation.

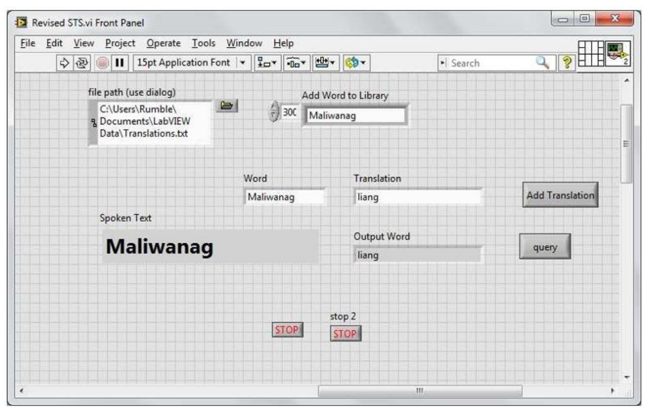

Figure 1: Front panel of the Speech-to-Speech

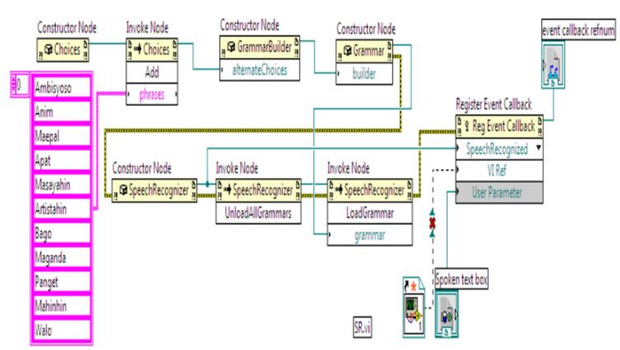

Figure 2: Speech Recognition Block Diagram

With regards to training, Windows Speech Recognition plays a big role. First, the user must supply a word that is needed in the training. Figure 1 shows the front panel of the overall system. The dialog box "Add Word to Library" is responsible for the list of words that one has entered and trained in the system. Once a word was entered into that box, the word is already ready for training, and here is where the Speech Recognizer can also be used. One just has to go to the train, and then input the different voice variations of the word needed to be trained. The Labview program for speech recognition is shown in figure 2 above.

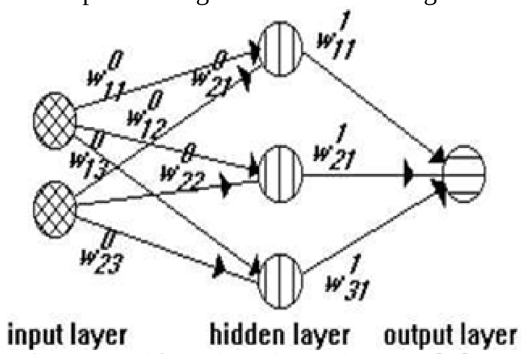

Figure 3: Feed forward neural network structure

Figure 3 above shows the multilayer feed forward (FF) network structure. This multilayer FF consists of many single layer FF that were combined together to form the hidden layers, which comprise the hidden neurons. As one can observe, the input neurons layer consists of 2 input neurons which are represented as 2 dimensional vectors. Following that is the hidden layer which consists of 3 hidden neurons which handle the weights. Finally, the output layer is just a single output neuron. With this kind of system, every single input to the neural network will be absorbed, extracting the components of the input, which in this case are the components of the voice signal, and then will be weighted to adjust the response.

3.2 Text-to-Text

After the Filipino words in speech are converted into Filipino words in text through speech recognition, the Filipino words in text will be translated to Chinese words in text using the text-to-text converter. In the text to text converter, two processes, adding of translation and querying, are undertaken. These processes can be implemented by clicking their respective buttons and considering some conditions.

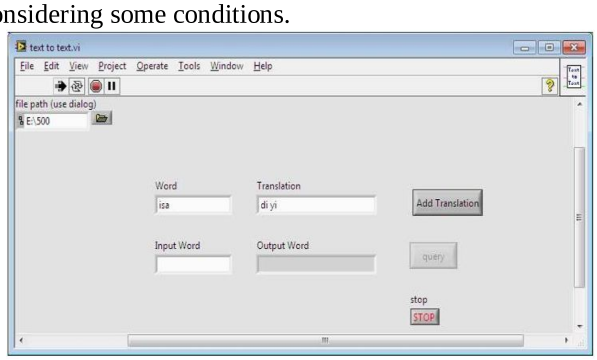

Figure 4: User Interface for Adding of Translation

Before the "add translation" button is enabled, two conditions must be satisfied. First, there must be an existing word in the "word" field of the control panel. The word typed in the "word" box field must be in Filipino Language. Second, the word typed in "translation" box field must be existing and the typed word must be in Chinese Language. If the two conditions are met, the "add translation" button will be enabled and once clicked, relation between the Filipino word and Chinese word will be stored in the database of the system. Figure 4 shows the user interface for adding translation to the system.

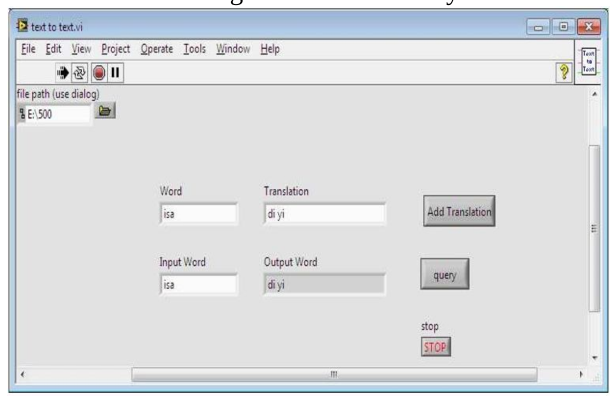

Figure 5: User Interface for Querying

In the querying process, the "query" button will be enable once a word is typed on the "input word" field. When the "query" button is clicked, the system will be displaying the Chinese translation of the word typed in the "input word" field by calling the information in the database of the system. On this process, the displayed word in the "output word" field is just a validation of the displayed word in "translation" field. Once the words in both fields (translation and output word) matched, the text to text converter will output a data in Chinese text. Figure 5 shows the user interface for querying process.

3.3 Text-to-Speech

After the text-to-text system outputs the equivalent Chinese word into the text-to-speech part, it will be recognized and fetched by three possible pathways.

First, when the Chinese word generated into the input is unknown to the database then it will generate a path going directly into a dialog box which will ask for the location of the Chinese speech file. The second part is when the input to the text-to-speech part is empty. When there is nothing going into the system, it will generate a path going into the dialog box again. Both unknown input and empty inputs are asked to input their corresponding Chinese audio files.

There are also some cases where the words inside the system are too close with each other with respect to their pronunciation, that is why the system gets confused on whether which file to make us. This case falls under the unknown input theory because it is not able to generate a specific or accurate file.

Figure 6 shows a Labview subroutine used in the program whenever an unknown input or an empty input is encountered by the program. Also, figure 7 shows a Labview subroutine program used whenever the input signal in detected to be included in the database.

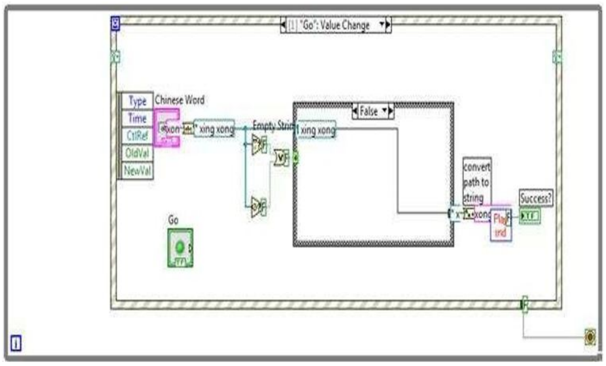

Figure 6: True Mode (Unknown Input and Empty Input)

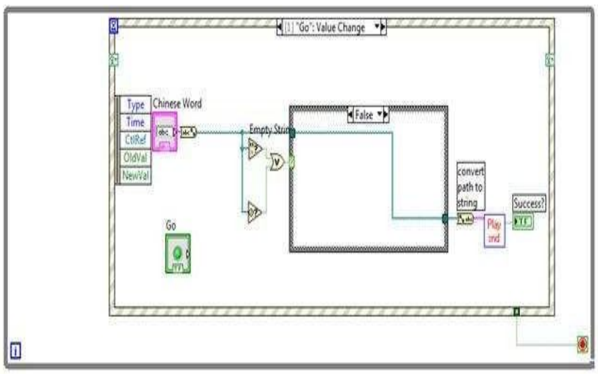

Figure 7: False Mode (Successful state)

For the third case, when the Chinese file detected into the system is recognized by the database, it will produce a short cut going directly to a format converter which will convert the format of the audio file into something that is recognizable by the wav player. In this case, the wav player plays the audio file that is generated from the database of the Chinese audio and are now represented as analog waveforms as output. These output sounds are then classified as the equivalent Chinese speech translation of the Filipino speech at the input. Figure 8 shows the full Labview program used in this research project.

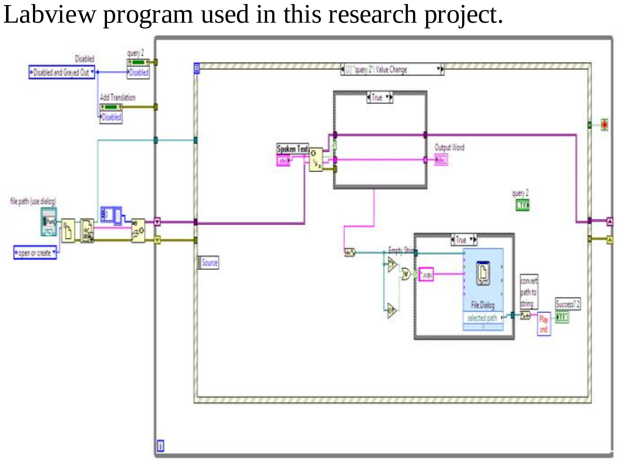

Figure 8: Over-all System Diagram

4. DISCUSSION OF DATA AND RESULTS

4.1 Training and Testing

After some tedious connections that were done in the three stages involved in this study, experiments were done in order to test the validity of the NNS. This is done by sampling the words that are included in the list of 500 Filipino words, and this must be translated into their equivalent Chinese words. Success rate of the overall system in translating the Filipino words must not be less than 90%.

4.2 File Path

Before the translation starts, the user must specify first the file path that is needed in the system.

4.3 Word Translation

After specifying the file path, the Filipino words must be trained in order to have outputs corresponding to their equivalent translation. The system is designed for an accurate translation that only the words included in the lists will be translated. As a result, the Chinese equivalents were produced (expressed in how they are pronounced).

4.4 Voice Training

Just like the previous one, this must be trained to get the respective voice translation from the translated Chinese texts. The system also got accurate translations based on the designed algorithm. As a result, the output is a voice signal recorded by the proponents.

4.5 Training and Testing of Words

This is the most crucial part of the system because of the different parameters and components involved in getting the right words that must serve as inputs. After different types of words that entered the system, it was found out that there are some words that achieved a high percentage in terms of the recognition rate. There are also some who got a low percentage in the testing of the trained words.

4.6 The Overall Training

Like what is observed in the training stage, it was found out that there are some lapses in the other words in the speech-to-text part. These are the words that have almost the same pronunciation with the other words. For example, the words, mataba and matabil, mataba and matabang, and mataas and matalas, they almost have the same sounds. Meaning, the components comprising the words are somehow the same.

Despite these, majority of the words got a low rejection rate (high success rate), and even got perfect score.

4.7 Success Rate in the Training

As one can observe, there are words that amidst the tedious training where a number of iterations were implemented, still these words are not that recognizable by the system. But still, there are instances where in it can be recognized depending on the pronunciation, articulation, diction and other factors concerning speech utterances.

Summarizing the overall performance of the system during the training stage, 467 over 500 words were successfully translated by the system. This is to test if the adjectives that serve as inputs in the system were properly trained after successive iterations. This means that a success rate of 93.4% (a rejection rate of 6.6%) was obtained. This was done after all the 500 Filipino adjectives had been trained into the system. Table 1 below shows the overall summary of the success rate of the system during the training stage. This data is also represented in figure 9 below.

Table 1: Success Rate of the system after different iterations

# of Iterations	Success Rate %
25	86.8
37	88.2
49	89.6
61	91.8
73	92.0
85	93.4

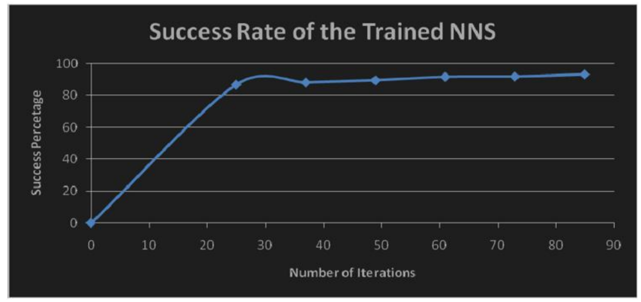

Figure 9: Graph of the success rate of the trained system

4.8 Overall Testing and Success Rate

Just like in the training stage, there are still some lapses in the overall testing stage. 25 testers were asked to utter the 500 Filipino adjectives, and different results were found.

Summarizing the overall performance of the system during the testing stage, an average success rate of 91.59% (a rejection rate of 8.41%) was obtained.

As one can observe, there are some testers who really got a low percentage. This was because of their way of pronouncing the words, sometimes unclear and sometimes not that correctly pronounced. But due to the tedious training done with the system, still the delays, wrong pronunciations and other significant factors that can further affect the performance of the system were somehow neglected. Table 2 shows some of the 15 words included in the 500 Filipino adjectives where 20 testers were considered.

Table 2: Example of 17 Filipino words with ambiguity

Filipino word	Ambiguous pronounciation	English word	Chinese word Pronunciation
ambisyo so	am-bi-syo- syo	Ambitious	xiong xin bo bo de
Anim	ha-nim	Six	liu
maepal	ma-he-pal	Annoyin g	tao yan de ren
Apat	ap-hat	Four	si
masayah in	mas-ya-hn	Joyful	le guan de ren
artistahin	ar-tith-tahin	Artist Looking	yi shu xing biao
bago	ba-gow	New	xin
nakakain is	na-ka-ka-ins	Irritating	fan ren de
baluktot	ba-lug-tot	Crooked	wan qu de
bastos	bs-tos	Pervert	duo luo zhe
nakakalit o	na-ka-li-li-to	confusin g	han hun bu qing de
bente	beyn-te	Twenty	er shi
bihira	bi-i-ra	Sometim es	you shi
bilugan	bi-lu-gin	circular	yuan xing de
palamuni n	pa-la-mu-nn	indolent	lan duo
ordinary o	hor-di-nar- yo	Ordinary	pu tong
normal	nr-mal	Normal	zheng chang de

5. CONCLUSIONS, RECOMMENDATIONS, AND FUTURE DIRECTIVES

5.1 Conclusions

This study is mainly about implementing neural network and database system aided by LabVIEW with the association of the different algorithms and applications like the Windows Speech Recognition. The system used a Feed Forward Type of Neural Network System with the help of the Speech recognizer, which was mainly used in this study. This was used in order to translate 500 Filipino words of 25 voice variations into their equivalent Chinese translation, having a delay of less than 2 seconds.

Three stages were made: the speech-to-text, text-to-text, and the text-to-speech part which have different LabVIEW functional algorithms, block diagrams and front panels. During the training stage, a success rate during the training stage of 93.4% (rejection rate of 6.6%) was obtained in the study upon translation from its Filipino speech input up to the equivalent Chinese translated speech output. On the other hand, an average success rate of 91.59% was obtained during the testing stage of the 25 testers uttering the 500 given words.

5.2 Recommendations and Future Directives

Throughout the course of the study, the researchers were able to accomplish the study's objectives as well as encounter various challenges. To further improve the study, the researchers recommend the following to overcome the study's challenges and to result into a better implementation of Neural Network System in speech to speech translation.

First, the researchers recommend more voice variations per word. It is proven in the study that voice variations are used as training iterations per word. An increase in the number of trainings also increases the efficiency of recognizing the input word. In connection to this, the researchers suggest to increases the efficiency of recognizing the input word- make it around 97 to 99 percent. The 1 percent will be left for the external variables which are uncontrollable.

Next, esearchers are recommended to increase the number of words, most likely around 2000 to 3000 words. This means that the kind of words used does not only involve adjectives but also nouns, pronouns, adverbs and other commonly used words in different parts of speech. The definition of common words also includes modern words such as slangs, expressions and the like. The reason for this is for the program to encompass all kinds of words spoken by modern people as much as possible.

In the long run, the researchers suggest an expanded application of neural network system in recognizing phrases and sentences. NNS will be used in arriving at a translated phrase or sentence with a correct grammar. The reason for this is to further improve the goal of having a totally computer-driven translator.

REFERENCES

1. Libatique, N.J.C et al, **Design of a tropical rain - Disaster alarm system: A new approach based on wireless sensor networks and acoustic rain rate measurements**, *2009 IEEE Intrumentation and Measurement Technology Conference*, I2MTC 2009 https://doi.org/10.1109/IMTC.2009.5168663
2. J. A. C. Jose, M. K. Cabatuan, E. P. Dadios, and L. A. G. Lim, **Depth estimation in monocular Breast Self-Examination image sequence using optical flow**, *2014 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 2014.
3. S. L. Rabano, M. K. Cabatuan, E. Sybingco, E. P. Dadios, and E. J. Calilung, **Common Garbage Classification Using MobileNet,** *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 2018. https://doi.org/10.1109/HNICEM.2018.8666300
4. ` J. A. C. Jose, M. K. Cabatuan, R. K. Billones, E. P. Dadios, and L. A. G. Lim, **Monocular depth level estimation for breast self-examination (BSE) using RGBD BSE dataset**, *TENCON 2015 - 2015 IEEE Region 10 Conference*, 2015.
5. Dulay, A., Sze, R., Tan, A., Yap, R., Materum, L., **Development of a wideband PLC channel emulator with random noise scenarios**, *Journal of Telecommunication, Electronic and Computer Engineering*, 2018
6. P. Loresco and A. Africa, **ECG Print-out Features Extraction Using Spatial-Oriented Image Processing Techniques.,** *Journal of Telecommunication, Electronic and Computer Engineering*. Vol. 10, Nos. 1-5, pp. 15-20, 2018.
7. Ankush Chunn and Alok Naugarhiya, **Use of Open Source CAD Tools in VLSI Design Curriculum for Developing Countries**, *International Journal of Emerging Trends in Engineering Research*, Volume 6, No. 4, 2018
8. Virtudez, K.J.D.A., Gustilo, R.C., **FPGA implementation of a one-way hash function utilizing HL11-1111 nonlinear digital to analog converter**, *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, 2012 https://doi.org/10.1109/TENCON.2012.6412237
9. Dulay, A.E., Sze, R., Tan, A., Huang, Y.-H., Yap, R., Materum, L., **FPGA implementation of an indoor broadband power line channel emulator**, *Proceedings of KICS-IEEE International Conference on Information and Communications with Samsung LTE and 5G Special Workshop*, ICIC 2017, . https://doi.org/10.1109/INFOC.2017.8001678
10. Materum, L., **Stochastic tapped delay line based one-sided beamformed channel impulse response models**

of LoS and reflected waves at 62.5 GHz in a conference room environment, *Journal of Telecommunication, Electronic and Computer Engineering,* 2017

11. Dulay, A.E., Yap, R., Materum, L., **Hardware Modelling of a PLC Multipath Channel Transfer Function, Journal of Telecommunication**, *Electronic and Computer Engineering*, 2017
12. A. Africa, **A Rough Set Based Solar Powered Flood Water Purification System with a Fuzzy Logic Model.** *ARPN Journal of Engineering and Applied Sciences*. Vol. 12, No. 3, pp.638-647, 2017
13. A. Africa, **A Rough Set-Based Expert System for diagnosing information system communication networks.** *International Journal of Information and Communication Technology*. Vol. 11, No. 4, pp. 496-512, 2017. https://doi.org/10.1504/IJICT.2017.10008315
14. J. A. C. Jose, M. K. Cabatuan, E. P. Dadios, and L. A. G. Lim, **Stroke position classification in breast self-examination using parallel neural network and wavelet transform**, *TENCON 2014 - 2014 IEEE Region 10 Conference*, 2014.
15. Hanpinitsak, P., Saito, K., Takada, J.-I., Kim, M., Materum, L., **Multipath clustering and cluster tracking for geometry-based stochastic channel modeling**, *IEEE Transactions on Antennas and Propagation*, 2017 https://doi.org/10.1109/TAP.2017.2754417
16. S. Brucal, A. Africa, and E. Dadios, **Female Voice Recognition using Artificial Neural Networks and MATLAB Voicebox Toolbox.** *Journal of Telecommunication, Electronic and Computer Engineering*. Vol. 10, Nos. 1-4, pp. 133-138, 2018.
17. A. Africa, **A Mathematical Fuzzy Logic Control Systems Model Using Rough Set Theory for Robot Applications.** *Journal of Telecommunication, Electronic and Computer Engineering*. Vol. 9, No. 2-8, pp. 7-11, 2017.
18. WooLim Kim and Sang Boem Lim, **Smart Chair Cover for Posture Correction**, *International Journal of Emerging Trends in Engineering Research*, Volume 7, No. 8, 2019 https://doi.org/10.30534/ijeter/2019/14782019
19. Yeap, Travlex. **Neural Network Applications**. *Singapore: Japan Singapore Artificial Intelligence Centre*, 2006