How to cite this paper:

Shelena Soosay Nathan , Ahmed , Azham Hussain, Nor Laily Hashim, Toan Nguyen-Duc, & Eiji Kamioka. (2017). Conceptual framework for enhancing a wearable device that converts sound, text and image into automatic sign language recognizing system (ASLR) in Zulikha, J. & N. H. Zakaria (Eds.), Proceedings of the 6th International Conference on Computing & Informatics (pp 691-695). Sintok: School of Computing.

CONCEPTUAL FRAMEWORK FOR ENHANCING A WEARABLE DEVICE THAT CONVERTS SOUND, TEXT AND IMAGE INTO AUTOMATIC SIGN LANGUAGE RECOGNIZING SYSTEM (ASLR)

Shelena Soosay Nathan[1] , Ahmed[2] , Azham Hussain[3], Nor Laily Hashim[4], Toan Nguyen-Duc[5] and Eiji Kamioka[6]

[1]*Universiti Utara Malaysia, Malaysia, sn.shelena@gmail.com*
[2]Universiti Utara Malaysia, Malaysia, *mohammed_ahmed_taiye@ahsgs.uum.edu.my*
[3]Universiti Utara Malaysia, Malaysia, *azham.h@uum.edu.my*
[4]*Universiti Utara Malaysia, Malaysia, laily@uum.com.my*
[5]*Shibaura Institute of Technology, Japan, nb14501@shibaura-it.ac.jp*
[6] *Shibaura Institute of Technology, Japan, kamioka@shibaura-it.ac.jp*

ABSTRACT. Current production of hearing aids meets lesser than global needs. Hearing aids and other assistive devices can be used to improve Deafness and hearing loss in people especially using sign language. However, expressing sign language is not known or understood by everyone. These possess serious drawbacks in communicating to and from the deaf. Thus, the paper is to propose a framework that will improvised wearable device for the deaf by using machine learning approach. This paper develops a conceptual framework for enhancing a wearable device that converts sound, text and images into automatic sign language recognition system. The proposed study is an initial step towards the full development and deployment of the wearable device. The device is aimed specifically for the physically challenged (deaf).

Keywords: wearable, machine language, deaf, conceptual framework

INTRODUCTION

Communication is an essential part of human life. Without communication, the world today will still be the same as it was in the Stone Age. Business transactions, decision making, development in information technology, advancement in science and technology, political formation, human rights and ethics is impossible without information communication. Communication can be of various forms namely; verbal, non-verbal, written and visual form (graphics, charts and images). In this study we concentrate more on improving verbal communication with the help of information communication technology employing machine learning approach.

Freedom of speech is entitled to everyone, especially the physically challenged (deaf). According to WHO, (2015) deafness and hearing loss fact sheets, the number of deaf people worldwide is about 360 million. The current production of hearing aid meets less than 10% of global needs. Hearing aids and other assistive devices helps to improve hearing loss in people. To enhance communication amongst the physically challenged, sign language is used as a medium of communication for the deaf. As with respect to spoken languages, sign languages varies from different regions. It also represents some basic communication that is not limited in expression. However, expressing sign language is not known or understood by everyone. It is a skill that needs to be properly learnt and understood by people in need of it either (deaf or not). (Dreuw, Rybach, Deselaers, Zahedi, & Ney, 2007).

Sign language employs hand with facial gestures for communication. It is difficult to communicate using the sign language, especially when it is not properly communicated using its known signs. This possesses a serious challenge and has intensified the need for researchers to come up with algorithms, models as well as devices to resolve these issues (Weaver, Starner & Hamilton, 2010; Boulares, & Jemni, 2012). Today many sign language applications are widely available especially on smartphones. Nevertheless, these applications are inadequate in terms of their usability and can still be optimized (Yeratziotis & Greunen, 2013).

To communicate sign language flexibly and reduce the hassle of learning for a long period. Many research has been carried out on courseware applications to ease communications amongst the deaf (Adamo-Villani, 2007; Gennari & Mich, 2008; Mohid & Zin, 2010 Ariffin & Faizah, 2010; Savita & Athirah, 2011; Shaffiei & Aziz, 2012). Moreover, it was noticed that these applications have limitations in terms of usage. Problems relating to usability are more challenging especially when it involves the conversion of text and speech into image using the machine learning approach. It is assumed that an enhanced wearable application for the deaf will foster the level of attentiveness and ease communication process for the deaf (Kim & Gilman, 2008).

The main aim of this paper is to propose a framework that will improvised wearable device for the deaf by using machine learning approach. The remainder of this paper is structured as follows; section 2 presents related works, section 3 gives detailed description of the proposed framework and Section 4 concludes the paper with further discussion and future recommendations.

RELATED WORKS

Many Electronic Travel Aids (ETAs) have similar principle in operation for example scanning the environment with different technologies as well as displaying the information gathered to other senses such as hearing and touch (Velázquez, 2010). As internet evolves, problems in accessing computer for the deaf arises as well. Some of the proposed solution to help these community are such as voice synthesizers and screen magnifiers. Besides that, sign language output terminals were also deployed for the deaf community (Velázquez, 2010; Angsanto & Lim, 2016).

However, these technologies needs to be well studied to assure that wearable technology will be of utmost assistance to the disabled. Over the years, Artificial Neural Network or Deep neural network has been used by many researchers (Abdel-hamid, Deng, & Yu, 2013). Recently, Convolutional Neural network has proven to outperform the state-of-the-art method. Convolutional Neural network takes audio inputs entropy sample which can be trained and tested for experimentation. Basically this trained audio dataset are processed using its Mel Frequency Cepstrum Coefficient (MFCC) which is a representation of the short term power spectrum of sound data. It is based on a cosine transform of a log power spectrum on a nonlinear mel scale of frequency. MFCC can be applied in the domain of speech recognition (device that can automatically recognize numbers in telephone) and as well in the field of information retrieval.

In addition, CNN has been used by previous studies to optimize the techniques of speech recognition (Abdel-hamid et al., 2013.) and also detecting whales sound using deep learning approach (Nouri, 2014). Another novel research made use of Speech learning to optimize, train and recognize voluminous vocabulary (Dahl et al., 2012). Thus, by proposing this conceptual framework, wearable devices such as smart glasses are able to incorporate speech to text conversion with machine learning techniques.

Existing research used machine language (ML) to perform similar study, but the contribution of this research is to develop a conceptual framework that will allow different input data such as text, image and sound employing the Convolutional Neural Network (CNN) approach to train the inputted datasets and transform datasets into an Automatic sign language recognition system. Display output will not show as an ordinary image but will be displayed in GIF image format. This will allow easy and faster illustration of communicated speech concurrently by the wearable device. Machine Learning is specifically leveraged in the study to supervise inputted dataset for the purpose of training as well as testing device performance implementation and evaluation. The study is not proposing a novel approach in machine learning but however adopting existing techniques with the introduction of Graphic Interchange Format (GIF) image format for ASLR. More so, this paper focuses only on mainly on the framework of the proposed study. Next part of the paper will be on the issues related with machine language and followed by conceptual framework that has been proposed.

CONCEPTUAL FRAMEWORK FOR THE PROPOSED STUDY

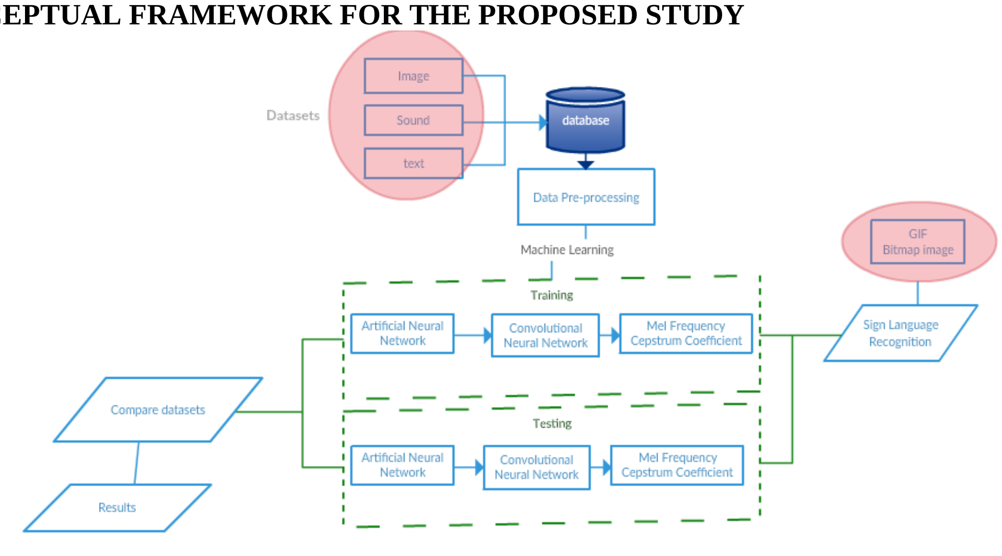

Figure 1. Proposed Framework

Figure 1 is a model of the proposed framework. However, the proposed framework will be able to pre-process three input namely; Text /Image recognition and Speech recognition. Images such as sign, text, and symbols will be translated using a Machine learning supervised methods. The Knowledge Based System will be incorporated in the proposed wearable device to allow easy conversion of text/ images into sign language (Caridakis, Asteriadis, & Karpouzis, 2014; Dreuw et al., 2007).

IMPLEMENTATION

The proposed study tends to improvise a wearable device using machine learning approach to convert text, image and speech data into sign language (image) (IR/ASR – to - ASLR) for easy communication. Sign language in our context will be in a Graphics interchangeable format (GIF). This is a bitmap image format that was developed in the United States in the late 80's. Animated images in form of frames will be added to the proposed GIF spec leveraging Graphics Control Extensions (GCE) (Patel, 2016). GCE allows various image frames in the device to be painted as inputted data are being processed without delay to form a video-like display of inputted information. These frame images will be processed in succession as a series of painted animation that is introduced by its own GCE. Communicated speech can be fast a times, for device to conveniently capture spoken words, it is necessary for the proposed device to enhance an improved method of displaying information to the user. GIF is proposed instead of 3D- video animation because Bitmap images consumes less storage than the 3D video animation and most times it is mostly used for graphics (Patel, 2016).

On the other hand, speech recognition mode will be basically focus on the grammatical and prosodic information (Dahl, Yu, Deng, & Acero, 2012). This information needs to be modelled with (speech/facial expression) to extract facial gestures with spoken words for direct communication and effective translation of spoken words by the device (Harashima, 1991). Different spoken dialects entails a close set of attribute language phonology. Thus, it is essential for morphological notations to be rich and sufficient for image recognition technologies. To enhance these technologies, some issues needs to be well addressed namely vocabulary recognition / size supported on a real time basis and the design and implementation of IR/ASR to ASLR feature extraction scheme application. The application of this device needs to be well ensured (Dahl et al., 2012).

CONCLUSION

This paper develops a conceptual framework for enhancing a wearable device that converts sound, text and images into automatic sign language recognition system. The proposed study is an initial step towards the full development and deployment of the wearable device. The device is aimed specifically for the physically challenged (deaf). It is an enhancement of already existing wearable device with other functional specification that improves communication amongst the deaf. This function includes the concurrent display of multiple framed images in form of GIF to meet up with the conversation speed when dialoguing with the deaf. GIF image is selected due to its flexibility and size.

The research aims at looking into other aspects apart from the Machine Learning. These aspects includes, sensors and control settings of the device, Energy and Data storage. However we only concentrate on a narrow aspect of the conceptual framework of the machine learning functions of the proposed study. We working on the other functions and how they can be incorporated together to achieve the aforementioned goals. This research requires a long term implementation of the device so as to evaluate its performance measures.

ACKNOWLEDGMENTS

We are grateful to UUM and SIT for giving us this opportunity and motivation for this study.

REFERENCES

Abdel-hamid, O., Deng, L., & Yu, D. (n.d.). Exploring Convolutional Neural Network Structures and Optimization Techniques for Speech Recognition.

Adamo-Villani, N., & Wright, K. (2007, August). SMILE: an immersive learning game for deaf and hearing children. *In ACM SIGGRAPH 2007* Educators Program (p. 17). ACM.

Angsanto, S. R., & Lim, W. (2016). Device-Independent Input-Assistive Wearable Keyglove for Visually Impaired Individuals. *International Journal of Applied Engineering Research*, *11*(15), 8392-8396.

Ariffin A. M. & Faizah M. (2010). Guidelines of assistive courseware (AC) for hearing impaired students. *In Proceedings of Knowledge Management International Conference*. 43, 49-340.

Barto, A. G. & Sutton, R. (1997). Introduction to Reinforcement Learning. MIT Press.

Boulares, M., & Jemni, M. (2012, April). Mobile sign language translation system for deaf community. *In Proceedings of the International Cross-Disciplinary Conference on Web Accessibility* (p. 37). ACM.

Caridakis, G., Asteriadis, S., & Karpouzis, K. (2014). Non-manual cues in automatic sign language recognition. *Personal and Ubiquitous Computing*, *18*(1), 37–46. http://doi.org/10.1007/s00779-012-0615-1

Dahl, G. E., Yu, D., Deng, L., & Acero, A. (2012). Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on Audio, Speech and Language Processing*, *20*(1), 30–42. http://doi.org/10.1109/TASL.2011.2134090

Dreuw, P., Rybach, D., Deselaers, T., Zahedi, M., & Ney, H. (2007). Speech Recognition Techniques for a Sign Language Recognition System. *Interspeech 2007: 8th Annual Conference of the International Speech Communication Association, Vols 1-4*, 705–708.

Gennari, R., & Mich, O. (2008, January). Designing and assessing an intelligent e-tool for deaf children. In Proceedings of the 13th International Conference on Intelligent user interfaces (pp. 325-328). ACM.

Harashima, H. (1991). A Media Conversion from Speech to Facial Image for Intelligent Man-Machine Interface. *IEEE Journal on Selected Areas in Communications*, *9*(4), 594–600. http://doi.org/10.1109/49.81953

Jain, A.K., Murty, M. N., and Flynn, P. (1999), Data clustering: A review, *ACM Computing Surveys*, 31(3): 264–323.

Kim, D., & Gilman, D. A. (2008). Effects of text, audio, and graphic aids in multimedia instruction for vocabulary learning. *Educational Technology & Society*, 11(3), 114-126.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques.

Mohid, S. Z., & Zin, N. A. M. (2010, June). Courseware accessibility for hearing-impaired. *In Information Technology (ITSim), 2010 International Symposium in (Vol. 1, pp. 1-5). IEEE.*

Nouri, D. (2014). Using deep learning to listen for whales. Retrieved December 1, 2016, from http://danielnouri.org/notes/2014/01/10/using-deep-learning-to-listen-for-whales/

Patel, D. V. (2016). Compare and Review of Various Steganalysis Security Techniques. *International Journal of Scientific Research*, (2277), 442–443.

Savita, K. S., & Athirah, A. N. (2011). Malay sign language courseware for hearing-impaired children in Malaysia. W*orld Applied Science Journal*, 12, p59-64.

Shaffiei, Z. A., & Aziz, N. (2012). Assistive courseware for hearing-impaired learners in Malaysia based on theory of multiple intelligences (MI). *International Journal of Computer Science & Emerging Technologies*, 2(6).

Velázquez, R. (2010). Wearable assistive devices for the blind. In *Wearable and autonomous biomedical devices and systems for smart environment* (pp. 331-349). Springer Berlin Heidelberg.

Weaver, K. A., & Starner, T. (2011, October). We need to communicate!: helping hearing parents of deaf children learn American Sign Language. *In The proceedings of the 13th international ACM SIGACCESS Conference on Computers and Accessibility* (pp. 91-98). ACM.

WHO. (2015). WHO Deafness and hearing loss. Retrieved December 10, 2016, from http://www.who.int/mediacentre/factsheets/fs300/en/

Yeratziotis, G., & van Greunen, D. (2013, May). Making ICT accessible for the deaf. *In IST-Africa Conference and Exhibition (IST-Africa*), 2013 (pp. 1-9). IEEE.