UOW MALAYSIA KDU PENANG UNIVERSITY COLLEGE

BACHELOR OF COMPUTER SCIENCE (HONS)

TRANSIGN – [Artificial intelligence-based real-time sign language to speech translation mobile application]

Lam E Shuen (0195136)

CCP3012 PROJECT PROPOSAL

Application Project

SEPTEMBER 2020

Contents

Project Title

Artificial intelligence-based real-time sign language to speech translation mobile application.

Introduction

People with disabilities have always found themselves disadvantaged during common daily life in one way or another. Examples of these disadvantages can be seen in the form of difficulties holding things due to missing limbs or being on the autism spectrum, impairing their learning ability in general education. An age-old disability disadvantage that is still finding ways to improve its issues is being mute(Lc, 2015).

Speech has always been the conventional method of communication between human beings and will be for the foreseeable future. The disability of being deaf or mute makes a person's ability to communicate infinitely harder due to the lack of a proper understanding or medium for speech, that medium being their ability to speak or hear.

Many communication aids have been created across the years to assist these speech impaired difficulties but lag behind the speed and clarity of communication via actual speech. Methods such as writing words, writing texts on screens, or the most common skill found among speech and hearing-impaired people, sign language.

There have been many forms of sign language just as there are many spoken languages. Sign languages were first known to exist in the 18th century but one of the most common forms of sign language, the American Sign Language was first founded in 1814 by Thomas Hopkins Gallaudet, a minister who wanted to teach hearing-impaired nine-year-old how to communicate. (Valli and Lucas, 2000)

As technology has advanced, many improvements have been made to assisting with communication such as text-to-speech applications when reading text during a conversation is not applicable or conversations over the globe through the Internet. With the advent of image processing and artificial intelligence being the global trend when it comes to technology, sign language assistance can once again be improved in ways which will be further explained in in this proposal.

Background of the Project

Even in our current era of the Internet where information and knowledge is easily accessible by common people almost anywhere and anytime in the world, proficiency in understanding sign language is still a skill rarely found among the majority of society which does not have a speech and/or deaf disability(Fenlon and Wilkinson, 2015).

According to the American Sign Language Program held by the University of Iowa, an estimate of 250,000 to 500,000 are users of sign language within America (Mitchell *et al.*, 2006). While in comparison there are, at the time the estimates were made in 2005, 296 million citizens within America(Passel and Cohn, no date). Taking the best estimates from above, only a measly 0.17% of America's population could use and understand the American Sign Language which is a travesty when you think about how a mute and/or deaf person is supposed to communicate within society. (Jay, 2010)

With that, this project will focus on the American Sign Language as the resources and data available open-source on the web is more abundant compared to other sign language data like Malaysian Sign Language or Chinese Sign Language. This will ensure an accurate working product is first created to show the effectiveness of an assistive application such as this. (Ltd, 2017)

Furthermore, even if people with said abilities were to obtain these skills, they would not be able to use it due to the lack of people that can understand them just like how they can't understand another country's spoken language. With translation through technology, the number of sign language users inside and outside of the deaf and mute community will increase due to an intangible increase in usefulness and usability of the language(Oliveira *et al.*, 2019).

Not only that, another caveat that comes with using text as a medium for communication between common people and disabled people is the inability to portray emotion(Marschark *et al.*, 2006). Sign language can differ in speed when signing to represent pause or urgency and the user of the language may also portray emotion through their expressions whereas text does not carry emotional meaning as well as sign language can(Elliott and Jacobs, 2013).

Lastly, the system proposed will be based around object recognition. It is a form of computer vision that recognises objects within images or videos. Object recognition's abilities are based around deep learning and machine learning and are considered a result of the models within them(Bülthoff and Edelman, 1992).

Definition of Key Concepts

At the base of this system is the hardware which will be the mobile handheld devices used daily by almost everyone known as smartphones. A device can be considered a smartphone when it has the processing power and memory to run applications, make phone calls and text messages, access to the internet, sound input and output, and lastly camera capabilities(*What is a smartphone?* | *Digital Unite*, no date).

With that said, the system will be using Android-based smartphones. Android is an operating system for smartphones developed by Google and is the operating system of more than 70% mobile handheld devices worldwide to date(*Mobile Operating System Market Share Worldwide* | *StatCounter Global Stats*, no date).

Moving along, a few keywords must be known that will be used throughout this proposal. The first one being **image processing.** The definition of image processing is to manipulate images by processing it through computers(Young *et al.*, 2004). Images can be broken down into a matrix of precise numbers that can be quantized and manipulated to do tasks such as image enhancement, restoration, and analysis(da Silva and Mendonça, 2005). The system will be mainly focused on the enhancement and analysis aspect of image processing to process then extract useful data from the video feedback given through the smartphone camera.

Besides that is **machine learning,** it is the core of what people of the current era knows as artificial intelligence(Shalev-Shwartz and Ben-David, 2014). It has the power to learn autonomously without any manipulation from a human being through the data fed into it to look for similarities or patterns through computing power and models made by human beings. It uses what it has learned to make decisions when provided input and output a desirable outcome(Decencière *et al.*, 2013).

Moving along is **deep learning**, a subset of machine learning algorithms mainly inspired by the function of a human brain called artificial neural networks(Socher,

Bengio and Manning, 2012). Deep learning models achieve immaculate accuracy and speed, sometimes better than human-levels when it comes to classification tasks directly from images, text, or sound. These models train through enormous sets of labeled data and neural network architectures containing many layers similar to a human brain(Perez and Wang, 2017).

Last of all we have **real-time object recognition,** which has the aim of producing decisions on the spot when given live data on-site like video from a camera feed(Chadalawada, no date). It requires a robust deep learning model that can process information as quickly and as accurately as possible due to the time-sensitive nature of tasks object recognition is used for such as autopilot systems(Gavrila and Philomin, 1999).

Proposed SDLC Methodology

For this system, the Prototyping Model is the methodology the system's development will follow. This methodology is the most suitable for this project as the ideas behind this system are fairly new and never done before so requirements for it may change as functions are developed(Al-Husseini and Obaid, 2018).

Firstly, the system's requirements will be analyzed by doing more research on what deaf and mute people would like to see in an application that will mainly affect them or be used by them. Not only that, the frameworks and models practical usefulness will also be analyzed in this phase.

After that, once all the requirements have been detailed and documented, a feasibility study will be done to see whether the project is feasible to be developed. An example of this is checking what Android devices have adequate resources to run this system once it is developed.

Continuing on, a feasibility study will be conducted when the first design of the system is created. This design is considered the first prototype and once it is created, it will go through an evaluation to see what other requirements or enhancements should be done. This will repeat until a satisfying product is created hence the name prototyping. The first prototype may be functional and able to recognize sign gestures but does not have a friendly UI to be used. It will go through the second prototyping to

redesign the UI with its sign gesture recognition functions and be reviewed again(Carr, 1998).

Lastly, once a satisfying prototype has been created, it will go through final testing and maintenance before becoming a finalized product. This is where all the test code and straggling bugs are cleaned up so that the final product used does not have any problems('SDLC - Software Prototype Model', no date).

Proposed Work

The proposed system is to create a mobile application in Android that will be able to recognize sign language gestures and translate it into speech. This application is mainly targeted to assist deaf and/or mute people with communicating with the general society when text or writing does not suffice.

As its nature states, this system will focus on Android handheld devices. This means that the application will only be able to run on devices that use the Android operating system. The device must have a camera which will not be a problem as the majority of Android devices nowadays come equipped with one.

The system will be designed with Kotlin, a language being pushed by the developers of Android, Google, as the new mainstay for Android application development(*Kotlin and Android | Android Developers*, no date). It will work in tandem with libraries such as OpenCV, TensorFlow, and Keras to recognize objects within the live video feed of the camera and train the data into suitable models.

The system will not have any processes that run in the background and will only be an application that can be run on the main thread. This will comfort the user's privacy is maintained as the system does not have anything to do if it is not currently being used.

As far as cloud connectivity goes, the requirement for it will be decided as development goes during the prototyping. Whether the model will perform better as a live service on the cloud or run locally on the device will be analyzed and evaluated during the developing phase.

The system will have its model created by TensorFlow and Keras and then trained with data found on the web or data given to it through the OpenCV controlled

camera video input from the Android device(Zebin, 2017). The system will learn to recognize the gestures through the model and link them to the correct translations.

There are a lot of basic American Sign Language datasets that can be found on Kaggle, an open-source dataset website, containing the datasets such as the alphabet and digit gestures which will become the foundation of the data used to train the model here(*Interpret Sign Language with Deep Learning*, no date). As for the image recognition techniques, TensorFlow has APIs with trained models to recognize objects and classify them into objects. For this system, once hands have been recognized then it can go deeper and look for what gestures are being performed to recognize what gestures are being made.

Aim of Project

The project aims to develop software that translates hand gestures viewed by a smartphone's camera into correct translations of the English language and output it as speech.

Objectives

1. To produce a complete proposal that fully explains and shows the charm behind the project idea

2. To produce a Gantt chart illustrating the timeline and process of the project

3. To research ideal frameworks and models to achieve the goal of the project such as OpenCV, Python, Android Kotlin, and Keras.

4. To understand the inner workings of how image processing and object recognition function and how it can recognize sign language gestures.

5. To design a user interface that will be user friendly and eye-pleasing for its intended demographic and a complete UML that helps the relevant people understand the behavior and structure of this system.

6. To develop a mobile application that is robust and efficient, one that can run smoothly on a handheld mobile device while producing translations that are accurate and embodies what the sign language user intends to communicate

7. To extensively test and evaluate the product once it is created.

8. To produce comprehensive documentation at the end of the project detailing the product and its creation process.

Skills

- BSc (Hons) Computer Science (University of Lincoln)
- Bachelor of Computer Science (Hons) specializing in Artificial Intelligence (UOW Malaysia KDU Penang University College)

Skills	Description
System Analysis and Design	Knowing project life cycles and designing UML for the system
Program Design & Development	Designing adequate UIs that are user-friendly
Data Science Tools and Techniques	Knowledge on various languages used to train models like Keras and TensorFlow
Mobile Application Development	Android system lifecycle and frameworks knowledge e.g. Kotlin and Java
Image Processing	Understanding what an image is in digital number matrixes and how to manipulate, enhance, or analyze them.

Sources of Information

Al-Husseini, K. and Obaid, A. (2018) 'USAGE OF PROTOTYPING IN SOFTWARE TESTING', 14, pp. 1–10.

Bülthoff, H. H. and Edelman, S. (1992) 'Psychophysical support for a two-dimensional view interpolation theory of object recognition.', *Proceedings of the National Academy of Sciences of the United States of America*, 89(1), pp. 60–64.

Carr, M. (1998) *Prototyping and Software Development Approaches*, *undefined*. Available at: /paper/Prototyping-and-Software-Development-Approaches-Carr/0b05add730e04843e234937a070f24b19efaadc3 (Accessed: 21 November 2020).

Chadalawada, S. K. (no date) 'Real Time Object Detection and Recognition', p. 66.

Decencière, E. *et al.* (2013) 'TeleOphta: Machine learning and image processing methods for teleophthalmology', *IRBM*, 34(2), pp. 196–203. doi: 10.1016/j.irbm.2013.01.010.

Elliott, E. and Jacobs, A. (2013) 'Facial Expressions, Emotions, and Sign Languages', *Frontiers in psychology*, 4, p. 115. doi: 10.3389/fpsyg.2013.00115.

Fenlon, J. and Wilkinson, E. (2015) 'Sign languages in the world', in, pp. 5–28. doi: 10.1017/CBO9781107280298.002.

Gavrila, D. M. and Philomin, V. (1999) 'Real-time object detection for "smart" vehicles', in *Proceedings of the Seventh IEEE International Conference on Computer Vision*. *Proceedings of the Seventh IEEE International Conference on Computer Vision*, pp. 87–93 vol.1. doi: 10.1109/ICCV.1999.791202.

Interpret Sign Language with Deep Learning (no date). Available at: https://kaggle.com/paultimothymooney/interpret-sign-language-with-deep-learning (Accessed: 21 November 2020).

Jay, M. (2010) *American Sign Language*, *Start ASL*. Available at: https://www.startasl.com/american-sign-language/ (Accessed: 21 November 2020).

Kotlin and Android | Android Developers (no date). Available at: https://developer.android.com/kotlin (Accessed: 21 November 2020).

Lc, F. (2015) *Witnesses to mute suffering: quality of life, intellectual disability, and the harm standard*, *The Journal of clinical ethics*. J Clin Ethics. Available at: https://pubmed.ncbi.nlm.nih.gov/25794290/ (Accessed: 2 November 2020).

Ltd, N. I. (2017) 'The Different Types of Sign Language', *Niki's Int'l Ltd.*, 12 September. Available at: https://nilservices.com/different-types-sign-language/ (Accessed: 21 November 2020).

Marschark, M. *et al.* (2006) 'Benefits of Sign Language Interpreting and Text Alternatives for Deaf Students' Classroom Learning', *The Journal of Deaf Studies and Deaf Education*, 11(4), pp. 421–437. doi: 10.1093/deafed/enl013.

Mitchell, R. E. *et al.* (2006) 'How Many People Use ASL in the United States? Why Estimates Need Updating', *Sign Language Studies*, 6(3), pp. 306–335. doi: 10.1353/sls.2006.0019.

Mobile Operating System Market Share Worldwide | StatCounter Global Stats (no date). Available at: https://gs.statcounter.com/os-market-share/mobile/worldwide (Accessed: 3 November 2020).

Oliveira, T. *et al.* (2019) 'Automatic Sign Language Translation to Improve Communication', in, pp. 937–942. doi: 10.1109/EDUCON.2019.8725244.

Passel, J. S. and Cohn, D. (no date) *U.S. Population Projections: 2005-2050 | Pew Research Center*. Available at: https://www.pewresearch.org/hispanic/2008/02/11/us-population-projections-2005-2050/ (Accessed: 2 November 2020).

Perez, L. and Wang, J. (2017) 'The Effectiveness of Data Augmentation in Image Classification using Deep Learning', *arXiv:1712.04621 [cs]*. Available at: http://arxiv.org/abs/1712.04621 (Accessed: 3 November 2020).

'SDLC - Software Prototype Model' (no date), p. 3.

Shalev-Shwartz, S. and Ben-David, S. (2014) *Understanding Machine Learning: From Theory to Algorithms*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107298019.

da Silva, E. A. B. and Mendonça, G. V. (2005) '4 - Digital Image Processing', in Chen, W.-K. (ed.) *The Electrical Engineering Handbook*. Burlington: Academic Press, pp. 891–910. doi: 10.1016/B978-012170960-0/50064-5.

Socher, R., Bengio, Y. and Manning, C. D. (2012) 'Deep learning for NLP (without magic)', in *Tutorial Abstracts of ACL 2012*. USA: Association for Computational Linguistics (ACL '12), p. 5.

Valli, C. and Lucas, C. (2000) *Linguistics of American Sign Language: An Introduction*. Gallaudet University Press.

What is a smartphone? | Digital Unite (no date). Available at: https://www.digitalunite.com/technology-guides/smartphones-tablets/smartphones/what-smartphone (Accessed: 3 November 2020).

Young, I. *et al.* (2004) 'Fundamentals Of Image Processing'.

Resources

Software:

1. Microsoft Word – Documentation and report writing medium

2. Google Chrome – Research and resource download medium

3. Android Studio – Software development medium

4. OpenCV – Image recognition library

5. TensorFlow – Model training library

6. Keras – Model library

Hardware:

1. Desktop for programming

2. Android phone for testing – Android 8 OS is the minimum requirement due to library requirements

Report structure

a) Title Page

b) Authorship Declaration

c) Acknowledgements

d) Abstract

e) List of Contents

f) Introduction

g) Analysis

h) Synthesis

i) Evalution and Conclusions

j) References

k) Bibliography

l) Appendices

- Terms of Reference
- Test Case
- User Manual (Basic)