How to cite this paper:

Toan Nguyen-Duc, Othmar Othmar Mwambe, Shelena Soosay Nathan, Mohammed Ahmed Taiye, Azham Hussain, Nor Laily Hashim, & Eiji Kamioka. (2017). Visualization of spoken language for deaf people in Zulikha, J. & N. H. Zakaria (Eds.), Proceedings of the 6th International Conference on Computing & Informatics (pp 696-701). Sintok: School of Computing.

VISUALIZATION OF SPOKEN LANGUAGE FOR DEAF PEOPLE

Toan Nguyen-Duc[1], Othmar Othmar Mwambe[3]
Shelena Soosay Nathan[2], Mohammed Ahmed Taiye[2]
Azham Hussain[2], Nor Laily Hashim[2], Eiji Kamioka[1]

[1]Graduate School of Engineering and Science, Shibaura Institute of Technology
[2]School of Computing, Universiti Utara Malaysia
[3]National Institute of Informatics

ABSTRACT. The academic achievement of the deaf is related to the difficulty in communication, not to the cognitive abilities. Without communication, their needs are generally misinterpreted and ignored. Accumulated years of misunderstanding will impair their academic as well as social success. Since the hearing world uses spoken language to communicate, the deaf are commonly isolated because they can use only sign language and lip-reading. The deaf ultimately go to special need schools where they can find a community with common ground and they can use sign language in their daily conversation. The special need schools are either primary or secondary ones, which offer free education to all children. However, special need schools often cannot offer the same broad curriculum as the general schools do. Therefore, in this work, a system that can support the deaf in general education is proposed. The system operates based on the cooperation between smart devices (including mobile devices) and wearable devices. The performance of the system has been validated based on a real test bed. The obtained results show that the system can support the deaf in real-time communication.

Keywords: deaf, sign language, wearable device, local smart network, general education

INTRODUCTION

Over five percent of the world's population is of deaf and hard of hearing people (WHO, 2015). Although some people, who are hard of hearing, still hear some sounds, they may not immediately respond to what they hear. Their hearing impairment delays the development of spoken language. The delays vary and depend on the level of hearing loss, the visual and auditory input they can receive (Blackorby & Knokey, 2006; Nicholas & Greers, 2006). The delays will exclude the deaf from communication and bring a significant impact on their daily life, i.e., causing feelings of loneliness or isolation, as well as poor academic performance. However, the IQ range of the deaf people is as much as of hearing individuals (Nikolaraizi & Makri, 2004/2005). When they have a suitable communication means, they can perform on an equal basis as others.

The deaf students, who study in the environment where verbal language is used, perform severely poor in the class and academic achievement due to the cognitive deprivation. Therefore, sign language was invented to enhance the educational achievements of deaf students. In most cases deaf students have a sign language interpreter. However, the interpreters translate spoken language into sign language based on their intuitive knowledge. The wider the interpreter's knowledge will bring better translation for the deaf students. On the other hand, the development of ICT enables a single computer to have an encyclopedic knowledge. It is reasonable to utilize that knowledge in supporting the translation from the spoken language to sign language. Many speech recognition (SR) engines have already been introduced (Google, Microsoft, and Nexiwave), aiming at translating spoken language into text. Once the speech is being converted in a text form, it can be easily translated from one language to another (Google). The automatic speech recognition (ASR) has many applications such as boosting up the writing skills, supporting reading skills as well as language learning processes (Follensbee, Bob et al, 2000; Nuance Communications, 2009; Gardner, T.J, 2008).Unfortunately, the support for the deaf has not been widely investigated.

There are many attempts to support hard of hearing people such as hearing support devices. However, deaf individuals generally feel like they are being repaired and mostly refuse to use such devices in their daily activities (McCormack, A., & Fortnum, H., 2013). It seems that to support the deaf closely, a friendly device is required. To this end, a smart wearable device is a potential candidate as it can be comfortably worn on the body. However, wearable devices are of battery-based dependency, they have a limited size and capability. As a result, deploying the ASR directly on the wearable device will exceed the limitation in size and consume a lot of energy. This is because the ASR generally requires high performance devices, which are currently not small and are energy-consuming ones. Besides, the devices typically require an Internet connection to refer to services deployed in the cloud. Without the Internet connection, the devices become useless.

In this work, a local smart network is introduced to address above issues. The local smart network operates on the basis of the cooperation among smart devices including wearable devices and mobiles devices. In the local smart network, the mobile device will be in charge of deploying the ASR engine instead of the wearable device. The wearable device is responded to record the speech and stream it to the mobile device for translating into text. The text then being converted into sign language for supporting the deaf in social communication as well as general education. The performance of the system is validated based on a set of experiment on a real testbed. The obtained results show that the system can give the sign language to the deaf with an accuracy of 80%. The response time of the system is shorter than the duration needed to read the script completely.

The rest of the paper is organized as follows: In Section 2, the basics of speech recognition will be described. In Section 3, the proposed local smart network will be explained. In Section 4, the performance of the local smart network will be evaluated. Finally, the paper will be concluded in Section 5.

2. BASIC PRINCIPLES OF SPEECH RECOGNITION

The speech recognition is a process executed by a software called a speech recognition engine (Lawrence Rabiner & Biing-Hwang Juang., 1993).The key function of the speech recognition engine is to process spoken language and translate it into text. Figure 1 shows the main components of a speech recognition engine. As shown in the figure, speech is analyzed to derive the features. The extracted features are used to enrich the training data and is compared with reference patterns given by acoustic model and language model. Acoustic model uses phoneme-like units along with a word lexicon to model the statistics of speech features. The language model takes into account the grammar and semantic for a higher level, i.e., sentence-level, matching. The final result is recognized words or sentences that have the best match to the speech input.

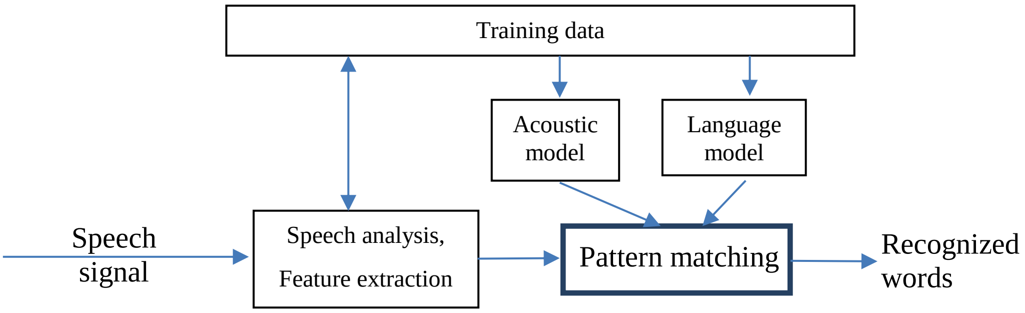

Figure 1. Main components of a speech recognition engine.

3. LOCAL SMART NETWORK DESIGN

3.1 Considered deaf community

In this work, the considered deaf people are all university students and use the sign language as their first language. Currently, the sign language varies in each country, the considered deaf community is assumed to use the same sign language, i.e., American Sign Language (ASL). Also, the sign database is assumed to be large enough to cover the spoken language vocabulary. Although the selected spoken language is English, this work can be applied to other spoken language as well. Besides, it is assumed that a formal language is used. Understanding spontaneous speech is out of scope of this work.

3.2 Design requirements

1. Handle the variation of spoken language

Even in general education environment, the deaf possibly communicate with non-native speakers. The local smart network must be able to support both native speakers as well as non-native ones. In addition, the network must take into account the variation of the spoken language across regions.

2. Support real time conversion

The local smart network must be able to provide sign language to the deaf in real time. However, the continuous connectivity may drain the battery of the wearable device, which is used to display the sign language. Therefore, the recognized words should be sent back to the wearable device in text format. The text then being compared with the stored sign vocabularies in the local database (DB_W) on the wearable device. If there is a matching, the signs will be displayed to the deaf.

3.3 Proposed local smart network

Figure 2 shows the diagram of a complete local smart network for visualizing the spoken language. In the first step of the speech conversion, the speech of the speaker is recorded by the microphone on the wearable device. The device then streams the recorded speech to a mobile device. Although the voice stream does not utilize high bandwidth, the voice is still being encoded for faster transmission. The ASR engine on the mobile device converted the decoded audio into words. The recognized words is then being sent as text messages to the wearable device. When the wearable device receives the messages, it searches its local database to find the appropriate signs. The signs are in GIF format to show the animation of hands. Finally, the matched signs will be displayed to the deaf.

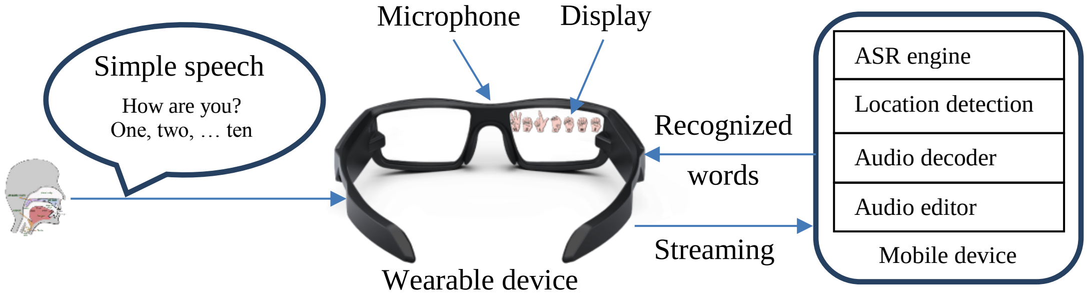

Figure 2. A local smart network for visualizing spoken language

As also shown in the figure 2, the mobile device has four components, namely the ASR engine, the Location detection, the Audio decoder and the Audio editor. The ASR engine is to recognize the speech. It can operate offline using a local database or connect to the cloud service. The engine also works based on location. Once the device get the location of the speakers via GPS, appropriate data will be used. The reference data can be enriched by training or downloading from Internet. The audio decoder is to decode the encoded stream. The last component is the audio editor is used to customize the recorded audio when necessary. Specifically, the audio editor can reduce the tempo of the audio instead of asking the speakers to speak slowly, aiming to increase the accuracy of the recognition process.

4. EVALUATION

To evaluate the performance of the system, the considered parameters are conversion accuracy and response time. The evaluation is performed on a test bed as illustrated in Fig. 2. For the sake of simplicity, the function of the wearable device and the mobile device are implemented using two conventional computers (Core 2 Duo @2.26 GHz processor and 2GB RAM, Ubuntu 14.04 64-bit), called WD and MD, respectively. The WD is to record the speech and streams it to the MD via Bluetooth. On the MD, a program receives the stream and asks the SR engine to convert it into text based on the detected location. The engine work based on the database provided by the cloud service. The recognized text is then being sent to the WD. Based on the received text, the WD searches its database to find the appropriate signs. The database here has 1000 records and is in XML format for quick accesses.

The respond time is defined as the different between when the audio stream is sent and when the sign is displayed. The measurement results show that the travelling duration between WD and MD is as small as 30ms. Also, the duration to get the appropriate sign images and to display the signs is less than 10ms. Therefore, the respond time is considered as the recognition time. On the other hand, the conversion accuracy is defined as the total correct recognized words over the total original words.

The experiments were performed by ten non-native English speakers and two English native speakers, four of them are women. The participants are invited to join two tests. In the first test, each participant reads a simple content (Fig. 2), which includes only short sentences with a few words. In the second test, a complex content was used. The complex content has five long and complex sentences, which include scientific terms. The recorded results given in Fig. 3 shows that spoken words in the first test were recognized correctly. The figure also shows that the native English speakers have higher accuracy than the non-native ones. To improve the recognition accuracy, the recorded speech of non-native English speakers are processed before sent to the SR engine. The audio editor on the MD adjusts the tempo of the speech to slow it down. Figure 4 shows the obtained results when three SR engines, namely Google, Microsoft and Nexiwave, are being used. When Google engine is being used, the recognition accuracy increases up to 72% when the tempo is reduced by 20%. However, when Microsoft and Nexiwave engines are being used the more tempo is reduced, the lower

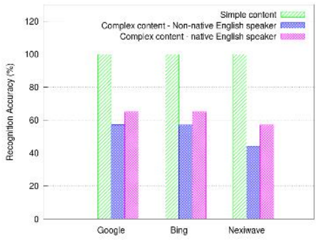

Fig. 3: Recognition accuracy measurement

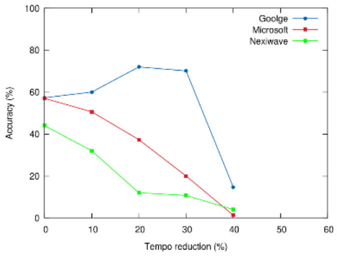

Fig. 4: Tempo adjustment

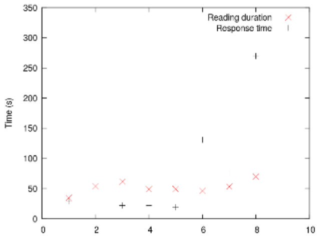

Fig. 5: Response time measurement

accuracy is. The accuracy also decreases when the tempo is reduced more than 30% even with Google engine. At the same time, the response time is also being measured. As shown in Fig. 5, the response time sometimes is larger than the reading time because the SR engine has to ask the cloud service to suggest for unknown input. In other cases, the response time is smaller than the duration the participants need to read the message. This means the signs language is displayed to the deaf immediately.

CONCLUSION

In this work, a local smart network for a visualization of spoken language has been introduced. The performance of the proposed system has been confirmed by a set of experiment on a real test bed. The captured results have confirmed that the system correctly recognizes speech of non-native English speakers when the content is simple. When the content is complex, the accuracy is low, however, it can be increased by adjusting the tempo of the speech. The system can also give the signs to the deaf immediately, however, in some cases, it takes time to ask the service on the cloud about the unknown terms. The future work will apply machine learning techniques to train data to increase the accuracy as well as reduce the response time.

REFERENCES

American Sign Language. http://www.lifeprint.com/

Blackorby, J., &Knokey, A. (2006). A national profile of students with hearing impairments in elementary and middle school: A special topic report of the special education elementary longitudinal study. Accessed December12, 2016, http://www.seels.net/grindex.html.

Follensbee, Bob; McCloskey-Dale, Susan (2000). "Speech recognition in schools: An update from the field". Technology and Persons with Disabilities Conference 2000. Retrieved 26 March 2014.

Gardner, T.J. (2008). Speech recognition for students with disabilities in writing. Physical Disabilities: Education and Related Services, 26(2), 43-53.

Google cloud speech recognition engine. https://cloud.google.com/speech/

Lawrence Rabiner and Biing-Hwang Juang. (1993). Fundamentals of Speech Recognition. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

McCormack, A., & Fortnum, H. (2013). Why do people fitted with hearing aids not wear them? International Journal of Audiology, 52(5), 360–368. http://doi.org/10.3109/14992027.2013.769066

Microsoft Cognitive Services. https://www.microsoft.com/cognitive-services/en-us/speech-api

Nexiwave Voice to Text services. https://nexiwave.com/

Nicholas, J., &Greers, A. (2006). Effects of early auditory experience on the spoken language of deaf children at 3 years of age. Ear and Hearing, 27 (3), 286–298

Nikolaraizi, M., &Makri, M. (2004/2005). Deaf and hearing individuals' beliefs about the capabilities of deaf people. American Annals of the Deaf, 149, 404–414

Nuance Communications. (2009). Dragon NaturallySpeaking: Helping all students reach their full potential. March 2009 White Paper, Nuance Communications.

World Health Organization. (2015). Deafness and hearing loss [online]. Availablehttp://www.who.int/mediacentre/factsheets/fs300/en/http://www.acpha-cahm.org/forms/acpha/acphahandbook04.pdf