A Low-Cost Speech to Sign Language Converter

Minh Le, Thanh Minh Le, Vu Duc Bui, Son Ngoc Truong
leminh@hcmute.edu.vn thanhlm@hcmute.edu.vn bdv24h@gmail.com sontn@hcmute.edu.vn
HCMC University of Technology and Education, Ho Chi Minh City, Vietnam

Summary
This paper presents a design of a speech to sign language converter for deaf and hard of hearing people. The device is low-cost, low-power consumption, and it can be able to work entirely offline. The speech recognition is implemented using an open-source API, Pocketsphinx library. In this work, we proposed a context-oriented language model, which measures the similarity between the recognized speech and the predefined speech to decide the output. The output speech is selected from the recommended speech stored in the database, which is the best match to the recognized speech. The proposed context-oriented language model can improve the speech recognition rate by 21% for working entirely offline. A decision module based on determining the similarity between the two texts using Levenshtein distance decides the output sign language. The output sign language corresponding to the recognized speech is generated as a set of sequential images. The speech to sign language converter is deployed on a Raspberry Pi Zero board for low-cost deaf assistive devices.
Key words:
Speech recognition, speech-to-text, sign language.

1. Introduction

Speech is a common way that humans use to communicate with each other. However, it is not used for people who suffer from speech and hearing disabilities. There are many people around the globe having disabilities in hearing and speaking. There is an existing language used for such people, which is called sign language. Sign language is a fully visual language with its grammar used for deaf and mute people [1]. In sign language, hand gestures, head, body movements, and facial expressions are used by humans to convey the information [2]. It is difficult for deaf people to understand the information from speech coming from normal people or even from media devices. Therefore, a translator is necessary for minimizing the communication gap between hearing impaired and normal people. The translator owns a speech recognition that translates speech to text and a sign language generator that converts text to appropriate sign language [3]-[5]. Speech recognition is a complicated task. The leading-edge speech recognition model is implemented using machine learning and deep neural network [6]-[9]. However, deep neural networks are based on a massive number of computational tasks which consume a huge amount of power and processing time. For this reason, a simple, portable, and low-cost translator is necessary for helping deaf people to receive information from the real world. In this paper, we experimentally present a design of a low-cost one-way portable translator, which can translate speech to sign language for deaf people. The speech recognition is performed using the open-source library, Pocketsphinx, which can work entirely offline [10]-[13]. Pocketsphinx is a small reconfigurable model which can be deployed on low-cost embedded systems for a mobile device. In this experimental demonstration, we deploy the speech recognition model and the speech to sign language on a low-cost Raspberry Pi Zero [14]-[16]. The optimized Pocketsphinx model has low accuracy because it employs the limited acoustic and language model. To improve the accuracy, a context-oriented language model is proposed. The proposed context-oriented language model is based on the Levenshtein Distance to measure the similarity between the recognized speech and the recommended speech.

Manuscript received March 5, 2021
Manuscript revised March 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.3.5

2. Design a speech to sign language converter

A speech to sign language converter for deaf people must satisfy several requirements such as mobility, low power consumption, low cost, and high reliability. There are three necessary modules inside such the device including speech-to-text module, language understanding module, and text to sign language converter module. For low-cost devices, we utilize the Raspberry Pi Zero for the control unit. The speech to text, language understanding, and text to sign language are deployed on the Raspberry Pi Zero board.

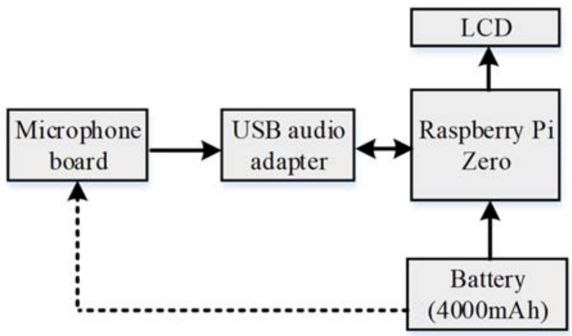

Fig. 1 The block diagram of the speech to sign language converter.

Fig. 1 shows a block diagram of a speech to sign language converter. The Raspberry Pi Zero is a low-cost embedded system being suitable for portable devices. Raspberry Pi Zero has only one channel for output audio. In this design, the input speech signal recorded from the microphone is passed to the Raspberry Pi Zero using a USB Audio Adapter, as shown in Fig. 1. Sign language is the animation composed of a set of sequential images being displayed on an LCD. The system is powered by a 4000 mAh battery. The device can work continuously 4 hours.

Speech to text is an important task that decides the reliability of the device. Recently, deep learning-based speech-to-text modules have been demonstrated as the leading edge model for automatic speech recognition systems. However, the deep neural network is a resource-hungry platform since it is based on huge computational tasks. For being suitable to be deployed on a low-cost computer such as Raspberry Pi Zero, we used an offline open-source speech-to-text module, Pocketsphinx instead. We propose a context-oriented language model to improve speech recognition accuracy. The Pocketsphinx is followed by the proposed context-oriented language model is shown in Fig. 2. Natural language processing is commonly used for language understanding tasks. However, the natural language proposed is a complicated task that requires a huge resource. In this design, we use a simple method that compares the recognized speech and the one stored in the database to decide the output. Levenshtein Distance method is suitable for such a task. The last task is the text to sign language converter. Having decided the output, a sign language is generated as a set of consecutive images playing on the output device. Fig. 2 shows the low-cost speech to sign language converter with proposed context-oriented language model.

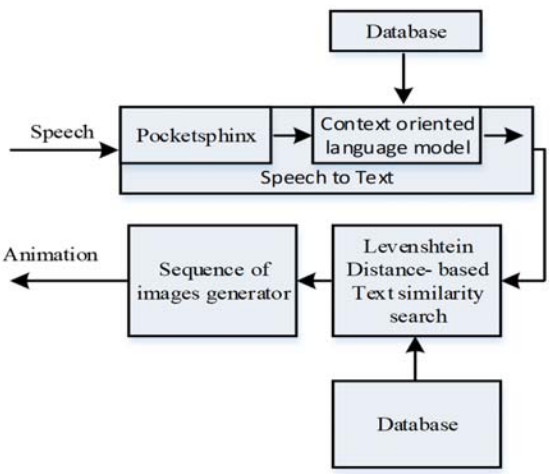

Fig. 2. The proposed speech to sign language converter with context-oriented language model.

In Fig. 2, the speech is recorded from the microphone and then enters the Raspberry Pi Zero where speech is converted to text by Pocketsphinx module. Pocketsphinx is an optimized Sphinx for low-cost computers. The proposed context-oriented language model is based on the Levenshtein Distance to measure the similarity of recognized speech and desired speech stored in the database. The output of the context-oriented language model is the speech obtained from the recommended speech which is the best match to the recognized speech. By using the proposed context-oriented language model, the recognized speech is corrected according to the expected speech stored in the database. The corrected text is then entered into the language understanding module where the output sign language is decided. Here the text is compared with the predefined text to determine which sign language output will be.

3. Experimental Results

The proposed architecture with three modules of speech to text, language understanding, and text to sign language is deployed on a Raspberry Pi Zeros board for a low-cost speech to sign language converter. The speech recognition accuracy is improved by using the proposed context-oriented language model which corrects the recognized speech. In table 1, we demonstrate the operation of the proposed context-oriented language model

Table I. The proposed context-oriented language model

Pocketsphinx output	Database	Similarity score	Context-oriented language model output
What do you doing	What are you doing	0.86	What are you doing
What are you do	What are you doing	0.9	What are you doing
why do you go	What are you doing	0.5	Where do you go
why do you go	Where do you go	0.8	Where do you go
I go working	I am working	0.83	I am working
I have to walk now	I have to work now	0.89	I have to work now

In table I, we evaluate the performance of the proposed context-oriented language model in enhancing the accuracy of offline speech recognition. The recognized speech is compared with the recommended speech stored in the database to decide the output speech if the similarity score is higher than 0.8. The Levenshtein Distance is utilized to measure the similarity of two strings. By doing this, the text converted from speech resulted from Pocketsphinx is corrected. In table I, the recognized from Pocketsphinx is "What do you doing", it is similarly the recommended speech of “What are you doing”, then the corrected output is “What are you doing”. Similarly, the recognized speech of “why do you go” better matches with “where do you go” rather than “What are you doing”, the output is “where do you go”. Using the proposed context-oriented model, the speech recognition rate is improved significantly. The database is composed of possible sentences. To evaluate the speech recognition module, we measure the accuracy for 500 sentences from 5 speakers. The speech is recorded from the microphone in realtime. The Pocketsphinx without the proposed context-oriented model can recognize the speech and convert it to text with an accuracy as high as 71%. The Pocketsphins followed by the proposed context-oriented language model has accuracy as high as 92%. The proposed context-oriented language model can improve the accuracy by 21%. Having received the speech, the sign language is generated at the output. Sign language is based on a set of sequential images as shown in Fig. 3

Fig. 3. The sign language of “what are you doing” is composed of a set of consecutive images.

Sign language corresponding to the text output from the speech to text module is a set of sequential images which specific meaning as shown in Fig. 3. The sign language is displayed on an LCD device.

4. Conclusion

This paper presented a design of a speech to sign language converted for deaf people. The device is mobility, low power consumption, and can work without an internet connection. The speech recognition is implemented by using an open-source library, Pocketsphinx module. To enhance the accuracy, we proposed a context-oriented language model, which measures the similarity between the recognized speech and the predefined speech to decide the output. The proposed model can improve speech recognition accuracy by 21%. A decision module is based on a similarity between the two texts using Levenshtein distance decides the output sign language.

Acknowledgments

This work belongs to the project grant No: T2020-39TĐ, funded by Ho Chi Minh City University of Technology and Education, Vietnam.

References

[1] U. Bellugi and S. Fischer, “A comparison of sign language and spoken language” Cognition, vol. 1, no. 2–3, pp. 173-200, 1972.

[2] O. Aran and L. Akarun, “Sign Language Processing and Interactive Tools for Sign Language Education,” 2007 IEEE 15th Signal Processing and Communications Applications, Eskisehir, 2007, pp. 1-4.

[3] L. Boppana, R. Ahamed, H. Rane and R. K. Kodali, “Assistive Sign Language Converter for Deaf and Dumb,” 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Atlanta, GA, USA, 2019, pp. 302-307.

[4] N. C. Camgoz, S. Hadfield, O. Koller, H. Ney and R. Bowden, “Neural Sign Language Translation,” 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 7784-7793.

[5] L. Kau, W. Su, P. Yu and S. Wei, “A real-time portable sign language translation system,” 2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS), Fort Collins, CO, 2015, pp. 1-4.

[6] P. Lakkhanawannakun and C. Noyunsan, "Speech Recognition using Deep Learning," 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), JeJu, Korea (South), 2019, pp. 1-4

[7] I. Gavat and D. Militaru, “Deep learning in acoustic modeling for Automatic Speech Recognition and Understanding - an overview,” 2015 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), Bucharest, Romania, 2015, pp. 1-8

[8] A. Kumar, S. Verma and H. Mangla, “A Survey of Deep Learning Techniques in Speech Recognition,” 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, 2018, pp. 179-185

[9] N. K. Mudaliar, K. Hegde, A. Ramesh and V. Patil, “Visual Speech Recognition: A Deep Learning Approach,” 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 1218-1221.

[10] K. Lee, H. Hon, M. Hwang, S. Mahajan and R. Reddy, “The SPHINX speech recognition system,” International Conference on Acoustics, Speech, and Signal Processing,, Glasgow, UK, 1989, pp. 445-448 vol.1, doi: 10.1109/ICASSP.1989.266459.

[11] K. Lee, H. Hon and R. Reddy, “An overview of the SPHINX speech recognition system,” in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 38, no. 1, pp. 35-45, Jan. 1990.

[12] D. Huggins-Daines, M. Kumar, A. Chan, A. W. Black, M. Ravishankar and A. I. Rudnicky, “Pocketsphinx: A Free, Real-Time Continuous Speech Recognition System for Hand-Held Devices,” 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Toulouse, 2006, pp. I-I.

[13] B. Lakdawala, F. Khan, A. Khan, Y. Tomar, R. Gupta and A. Shaikh, “Voice to Text transcription using CMU Sphinx

A mobile application for healthcare organization," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 749-753.
[14] D. B. C. Lima, R. M. B. da Silva Lima, D. de Farias Medeiros, R. I. S. Pereira, C. P. de Souza and O. Baiocchi, "A Performance Evaluation of Raspberry Pi Zero W Based Gateway Running MQTT Broker for IoT," 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2019, pp. 0076-0081
[15] N. S. Yamanoor and S. Yamanoor, "High quality, low cost education with the Raspberry Pi," 2017 IEEE Global Humanitarian Technology Conference (GHTC), San Jose, CA, USA, 2017, pp. 1-5
[16] A. P. Jadhav and V. B. Malode, "Raspberry PI Based OFFLINE MEDIA SERVER," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 531-533