HICAS: Hearing Impairment Communication Assistive System with the Extraction Search Network

Pei-Xin Ye, Geng-Kai Wong, Samantha Shih, Pei-Hsin Chang, Yi-Hsiang Cheng, Chiun-Li Chin*

***Abstract*—Hearing-impaired individuals frequently encounter substantial communication difficulties when seeking healthcare services for their hearing impairment. To address this issue, we present the Hearing Impairment Communication Assistive System (HICAS). This pioneering system integrates Speech-to-Image (S2I) functionality to revolutionize communication for the hearing-impaired. This research introduces the Extraction Search Network, comprising three essential components: Speech-to-Text (S2T), Text-to-Keywords Extraction (T2KE), and Keyword-based Image Retrieval (KBIR). The S2T component employs ESPnet for instantaneous and precise speech recognition. Subsequently, the T2KE component utilizes the RoBERTa model and Saliency Maps to extract keywords from the recognized text. Finally, the KBIR method searches for image locations, with the confidence scores determining the optimal image output. Experimental results demonstrate HICAS achieving an impressive real-time image search accuracy rate of 94.62% during conversations, displayed through AR glasses, effectively alleviating communication challenges for users.**

Keywords —Hearing-impaired,* communication difficulties, *HICAS, S2I, Extraction Search Network, ESPnet, RoBERTa, Saliency Maps, AR glasses.

I. INTRODUCTION

Hearing impairment (HI) is a chronic condition that often leads to communication barriers, especially when seeking healthcare services for their hearing impairment. Research by Stevens et al. [1], highlights these challenges, prompting the development of innovative solutions. Furthermore, we find that speech-to-image (S2I) retrieval can be effectively aiding language learning [2], allowing hearing-impaired individuals to comprehend location images in conversations by visualizing the information. Notably, Lee et al. [3] embedded S2T and Multichannel Acoustic Beamformer in portable devices, improving communication between hearing-impaired and normal-hearing individuals. Nonetheless, the difficulty in capturing pivotal details within discussions persists due to hearing loss. This paper introduces the concept of the Hearing Impairment Communication Assistive System (HICAS), an innovative solution aimed at surmounting communication barriers confronted by the hearing-impaired. This system instantaneously records conversations between hearing-impaired individuals and others, annotating keywords within the text content. Subsequently, HICAS employs advanced retrieval techniques to source pertinent location images from the internet, presented via AR glasses. This innovative interface strives to effectively address the aforementioned challenges.

P. X. Ye, is with the Chung Shan Medical University, Taichung, Taiwan (e-mail: 1158032@live.csmu.edu.tw).

G. K. Wong is with the Feng Chia University, Taichung, Taiwan (e-mail: D1019632@o365.fcu.edu.tw).

S. Shih, is with the Morrison Academy Taichung, Taichung, Taiwan (e-mail: shihs24@ma.org.tw).

P. H. Chang, is with the Taipei Municipal XiSong Senior High School, Taipei, Taiwan (e-mail: 11030162@ms2.hssh.tp.edu.tw).

Y. H. Cheng, is with the National HsinChu Senior High School, Hsinchu, Taiwan (e-mail: g0110141@hchs.hc.edu.tw).

C. L. Chin is with the Chung Shan Medical University, Taichung, Taiwan. (corresponding author to provide phone: +886-911-865102; e-mail: ernestli@csmu.edu.tw).

II. METHODS AND RESULTS

To realize the three components of the Extraction Search Network method, namely S2T, T2KE, and KBIR. Firstly, we leveraged the ESPnet framework as the foundation [4] to train an end-to-end S2T model, streamlining the system construction process and achieving S2T functionality. Subsequently, for the T2KE component, we employed the pre-trained weights of a RoBERTa model as initial weights and fine-tuned them. We then utilized Saliency Maps to obtain gradients for each input token, categorizing sentences with location-related content. The Softmax activation function was applied at the output layer to label and extract keywords with high probabilities. In the KBIR part, we treated keywords as input for image queries, performing image searches through image retrieval. We used a CRF probability model and MaxEnt classifier to calculate the optimal confidence [5], selecting the option with the highest confidence as the output image result. The results were presented to users through AR glasses, combining text and image outcomes.

To validate whether the HICAS system could assist hearing-impaired individuals in improving communication difficulties, we conducted tests on both hearing-impaired individuals and the elderly. Through experimental results, we found that the accuracy of S2T conversion reached up to 90.57%. Regarding T2KE, we involved 26 participants to annotate keywords for the same text content, resulting in a keyword similarity of approximately 81.34%. Lastly, in the KBIR component, real-time image retrieval from conversations took approximately 0.8 to 1.5 seconds, achieving an accuracy of about 94.62%.

III. CONCLUSION

This paper proposes a HICAS with S2I functionality to address communication difficulties faced by individuals with hearing impairment and their challenges in capturing key points of paragraphs or sentences. Based on the experimental results, the system demonstrates promising data outcomes, affirming its effective assistance for users. Through HICAS, communication content with others, conversation highlights, and images related to Hearing Care Center locations can be conveniently and visually accessed.

REFERENCES

[1] Stevens, M. N., Dubno, J. R. and Wallhagen, M. I., et al. "Communication and healthcare: self-reports of people with hearing loss in primary care settings." Clinical gerontologist 42.5 pp. 485-494, 2019.

[2] Merkx, D., Frank, S. L., and Ernestus, M. "Language learning using speech to image retrieval. " arXiv preprint arXiv:1909.03795, 2019.

[3] Lee, S., Kang, S. and Ko, H., et al. "Dialogue enabling speech-to-text user assistive agent with auditory perceptual beamforming for hearing-impaired." 2013 IEEE International Conference on Consumer Electronics (ICCE). IEEE, pp. 360-361, 2013.

[4] Watanabe, S., Hori, T. and Karita, S. et al., "Espnet: End-to-end speech processing toolkit." arXiv preprint arXiv:1804.00015, 2018.

[5] Culotta, A., and McCallum, A., "Confidence estimation for information extraction." In Proceedings of HLT-NAACL 2004: Short Papers pp. 109-112, 2004.