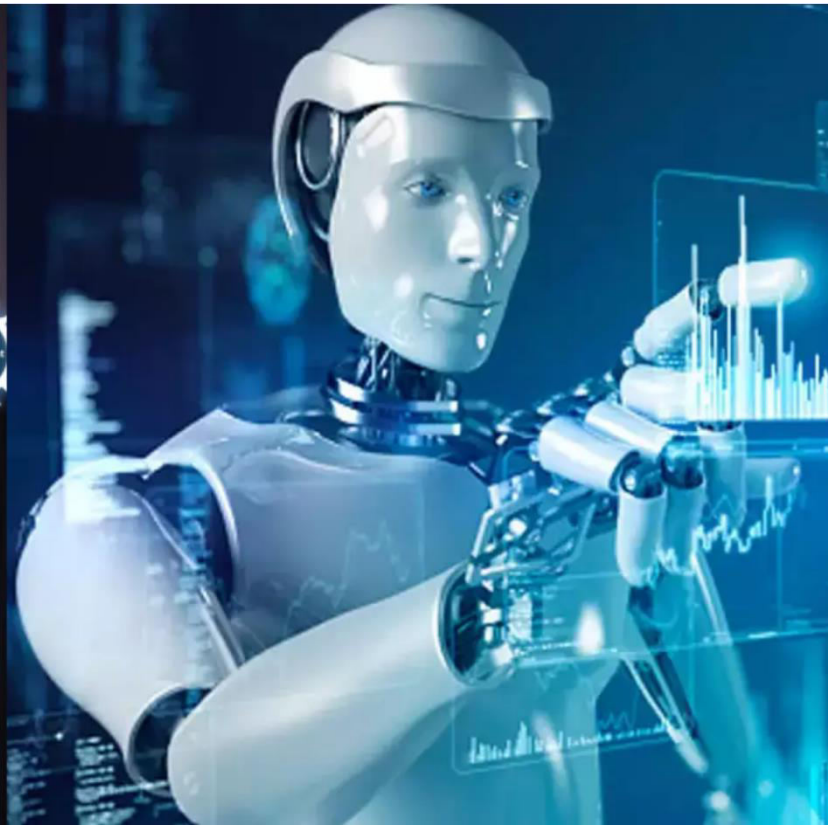
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404029|

Exam Assistive Multi-Modal Voice Transcription for Differently Abled Person

S.Maheswari, R.Vishnukumar, R.Sarvesh, K.Subashri, S.Vinith, Dr.N.Saravanakumar

Department of EEE, Kongu Engineering College, Perundurai, Tamil Nadu, India

Associate Professor, Department of ECE, Mahendra Institute of Technology, Namakkal, India

ABSTRACT: Students with physical limitations frequently struggle with traditional examinations due to a lack of support. Many exam formats do not provide appropriate accommodations for persons who are unable to write, resulting in an unequal academic experience. To address this, the study introduces a real-time voice-to-text transcription system specifically developed for students with disabilities. The device employs a microphone with noise cancellation to clearly record students spoken responses, which are then analyzed by a Raspberry Pi running speech recognition software. The spoken words are quickly transformed to text and shown on a monitor, so students can review and change their responses. Following the exam, the text is saved as a PDF for simple submission. The proposed system supports seven languages of voice to text conversion namely English, Hindi, German, Japanese, Chinese, French, and Portuguese. Also, the proposed system has Wi-Fi enabled notification system to alert support team by triggering an LED and buzzer. This paper offers a comprehensive way to ensure that all students can take exams, guaranteeing them equal chances to show their knowledge and advancing educational equity.

KEYWORDS: Microphone, RaspberryPi, Memory, Sound Card.

I. INTRODUCTION

Students with physical limitations, especially those who are unable to write, encounter substantial obstacles during tests. They frequently rely on scribes, causing delays, confusion, and a loss of privacy. Traditional methods, such as voice recording or hand transcription, are time consuming and unreliable. These limits make exams stressful for these pupils and put them at a disadvantage when compared to others. Current solutions are inaccurate or unsuitable for technical tests, exacerbating the issues. To overcome these issues, the Exam Assistive Multi-Modal Voice Transcription System proposes a real-time voice-to-text solution. This method instantaneously converts audible speech into written text, allowing for faster and more accurate transcription. It will also lessen reliance on others by allowing students to finish their exams alone [1-9].

Kurniawati et al. develop a zero-shot voice cloning TTS system tailored for dysphonia disorder speakers, enhancing clarity and adaptation through deep neural networks [1]. Danyang et al. present Neural VC, a real-time any-to-any VC system using neural networks, offering high-quality voice conversion with minimal data requirements [2].Gabrys et al. introduce Voice Filter, a few-shot TTS speaker adaptation method using VC as a post-processing module, achieving high-quality synthesis with minimal data [3].Gong et al. propose a unified pretraining framework for ASR and ST, combining speech and text data to improve performance in multilingual and low-resource scenarios [4].Huang et al. propose a seq2seq Voice Conversion (VC) method leveraging TTS and ASR pretraining, enhancing speech intelligibility and quality in low-resource scenarios for RNN and Transformer architectures [5]. Zhang et al. propose a transfer learning approach for VC using non-parallel data, enabling high-quality any-to-any voice conversion with improved naturalness and speaker similarity [6].Li et al. present FreeVC, a text-free one-shot VC system that employs VITS and WavLM features to achieve versatile, high-quality voice conversions without the need for text data [7].Reddy et al. analyze deep learning advancements in STT and TTS, emphasizing the role of CNNs, RNNs, and Transformers in improving accessibility and accuracy across applications [8].Tang et al. develop a multi-task learning framework that integrates ASR and Speech Translation (ST) with text data, significantly reducing error rates and improving translation accuracy in low-resource settings [9].Yu et al. introduce a non-parallel many-to-many VC framework using knowledge transfer from a multi-speaker TTS model, achieving high-quality conversions for over 200 speakers while retaining emotional tone [10].

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404029|

To address these issues, implementing advanced voice conversion models and transfer learning techniques into multi-modal voice transcription systems has the potential to substantially enhance communication for differently abled people. The use of pre-trained models, the system can improve voice intelligibility and transcription accuracy, resulting in more effective communication. In addition, multi-task learning frameworks can address data availability by enhancing speech translation and voice recognition performance. These advancements aim to improve accessibility by making voice-assisted applications more efficient and inclusive, resulting in better communication options for people with impairments.

II. PROPOSED METHOD

This system assists differently abled individuals during exams by developing a real-time voice-to-text transcription solution. Powered by a Raspberry Pi 4, it uses a USB-connected microphone to capture and process audio input. The system transcribes spoken responses in real-time, displaying the text for immediate review, and allowing manual corrections. Once the exam is completed, the system converts the transcription into a PDF format and sends it directly to the printer, ensuring timely and efficient submission.

The system is designed with accessibility in mind, the system offers an intuitive interface allowing students to easily start and stop transcription as needed. The setup prioritizes accuracy even in moderately noisy environments, providing a reliable tool for students to express their knowledge. With the inclusion of Raspberry Pi, this solution remains cost-effective, scalable and portable making it suitable for diverse educational settings ensuring a fair and inclusive examination process for students with disabilities. The system supports English, Hindi, German, Japanese, Chinese, French, and Portuguese, making it accessible to diverse users. It also integrates with Google Sheets to help examiners monitor students' writing status in real time, enhancing transparency and simplifying data management.

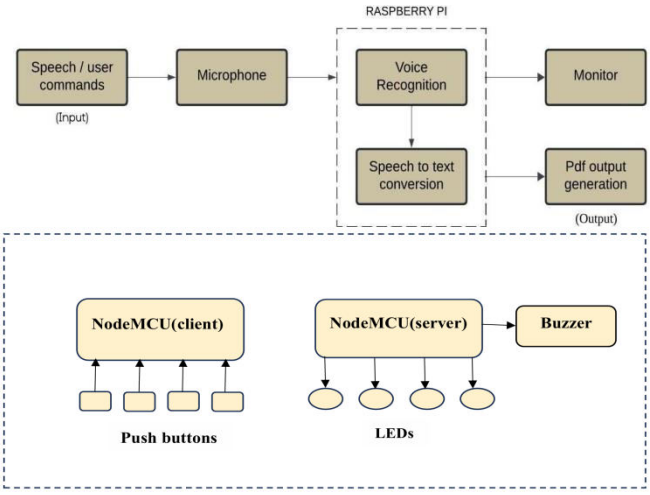


Fig.1 Block diagram of the proposed method

The block diagram in Fig.1 illustrates the use of the Raspberry Pi, coupled with a noise-canceling microphone for clear audio capture and advanced speech recognition software for real-time transcription. The Raspberry Pi processes the captured audio and converts it into text, which is displayed on a monitor for immediate review and editing. Upon completion of the exam, the system generates a PDF for easy submission. This streamlined system enhances accessibility, enabling students to effectively express their knowledge without the limitations of traditional writing methods. It has a feature for alerting the support team. When a button is pressed on client NodeMCU, it sends a signal via Wi-Fi to server NodeMCU, triggering an LED and buzzer to notify the support team promptly.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404029|

III. HARDWARE DESCRIPTION

This chapter discusses the hardware descriptions of the components used in the real-time voice-to-text transcription system designed for examination settings. The components used in this proposed model include: Microphone, Raspberry Pi, SD card, Sound card and Power supply.

A. Microphone

A microphone is an electronic device that captures sound and converts it into an electrical signal, as shown in Fig 4.1. It plays a critical role in accurately recording audio, ensuring clear communication during assessments. The microphone works by converting sound waves into vibrations on a diaphragm, which then generate an electrical signal. This signal corresponds to the sound wave's amplitude and frequency. Key elements of a microphone system include the diaphragm, transducer, housing, output connector, and power supply, all working together to capture sound effectively. Features like frequency response, sensitivity, directionality, and noise cancellation technology improve audio clarity and functionality.

Fig. 3 Microphone

B. Raspberry Pi

The Raspberry Pi 4, shown in Fig. 4 is a low-cost, credit card-sized computer developed by the Raspberry Pi Foundation. It features a Broadcom BCM2711 quad-core Cortex-A72 (ARM v8) 64-bit SoC, clocked at 1.5GHz, and offers 2GB, 4GB, or 8GB of LPDDR4-3200 SDRAM. It includes two USB 3.0 ports, two USB 2.0 ports, and Gigabit Ethernet for connectivity. The Raspberry Pi 4 also has dual micro-HDMI ports for 4K video output and a 40-pin GPIO header for external devices. It supports Linux-based operating systems like Raspbian and programming languages such as Python, C++, and Java. Ideal for learning coding and electronics, it can be used in diverse projects such as media centers, retro gaming consoles, and smart home systems. Its affordability and versatility make it a popular choice for hobbyists and professionals.

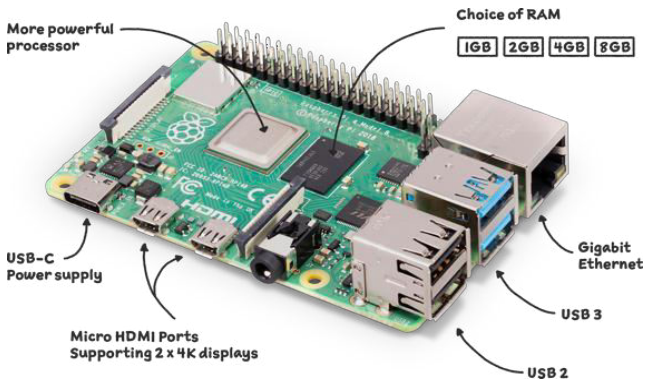


Fig. 4 Raspberry Pi

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404029|

C. SD Card

The 32GB SD card, shown in Fig.5, serves as the primary storage medium for the Raspberry Pi system. It supports the installation of the operating system, essential software, and applications for real-time voice-to-text transcription. With a Class 10 or UHS-I speed rating, it ensures fast data transfer, with write speeds of 10 MB/s and read speeds up to 80 MB/s. This high-speed performance is crucial for seamless operation during examinations. The SD card is durable, shock-resistant, and operates within a wide temperature range, making it suitable for portable setups. Formatted in FAT32, it offers compatibility with the Raspberry Pi and enhances the system's efficiency and reliability.

Fig.5 SD Card

D. Node MCU

The Node MCU shown in Fig.6 is used in Server-Client Setup, enabling seamless communication between multiple Node MCU boards for multiple candidates. In this setup, one Node MCU functions as a client, equipped with a button that, when pressed, transmits a signal to the server Node MCU via Wi-Fi. The server Node MCU, connected to an LED and a buzzer, promptly processes the received signal, triggering both the LED and buzzer to alert.

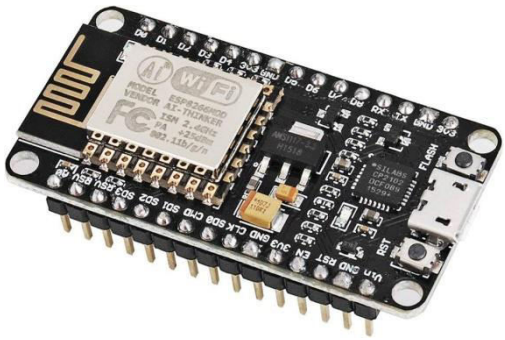


Fig.6 NodeMCU

E. Power Supply

The Official Raspberry Pi 4 Power Supply, shown in Fig.7, is designed to provide stable and reliable power to the Raspberry Pi 4 and its connected devices. It offers a steady 5V 3A output via a USB-C connector, ensuring consistent operation even when multiple peripherals like displays and microphones are in use. Built-in protections against electrical surges and overheating safeguard the Raspberry Pi and its components, making it ideal for applications requiring continuous power, such as voice-to-text transcription systems.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404029|

Fig.7 Power Supply

IV. HARDWARE RESULTS AND DISCUSSION

The proposed real-time voice-to-text transcription system integrates a Raspberry Pi 4 with a USB microphone for capturing high-quality audio from students with disabilities during examinations, as shown in Fig.8. This system converts spoken words into text in real-time using advanced speech recognition software. The transcription appears instantly on a connected monitor, ensuring both students and examiners can view the text immediately. The system also features data storage for future reference, enhancing reliability and providing a record of student responses. Proper microphone placement and efficient processing of audio inputs ensure accurate transcription, offering an inclusive solution for students during exams.

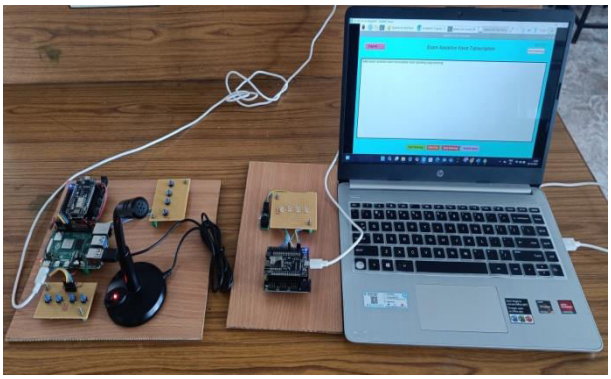


Fig.8 Hardware implementation of the proposed system

The Exam Assistive Voice Transcription system integrates a user-friendly interface, allowing students with disabilities to convert spoken words into text during examinations. At the beginning, the system prompts users to enter their name and ID. The Start Listening button activates the speech recognition process, enabling the system to capture audio from a connected microphone and instantly transcribe it, as shown in Fig. 9. The Stop Listening button allows users to halt transcription, while the Correction button provides manual adjustments for any transcription errors. The Submit Sheet button enables easy submission of the transcribed text for review. The system also includes a noise-canceling microphone for clearer audio capture, even in moderately noisy environments, ensuring accuracy. It supports seven languages. They are English, Hindi, German, Japanese, Chinese, French, and Portuguese ensuring accessibility for a diverse range of users. To enhance monitoring and record-keeping, the system integrates seamlessly with Google Sheets, allowing examiners to track whether students are actively writing or not in real time. The system has a feature for alerting the support team. When a button is pressed, it sends a signal via Wi-Fi to the support team, triggering an LED and buzzer to notify for support.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404029|

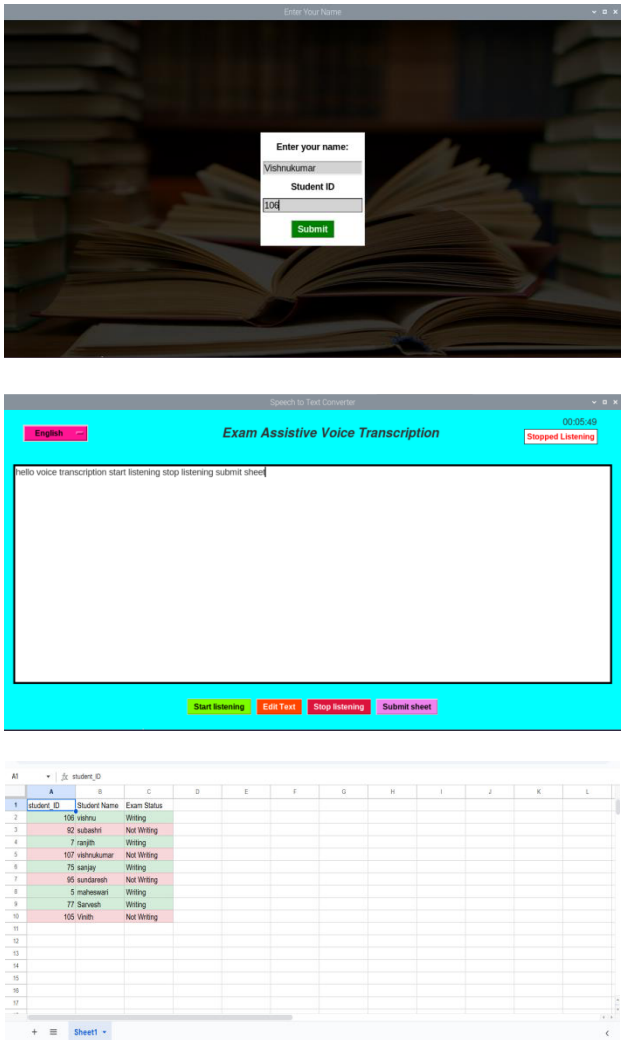


Fig. 9 Output of the proposed method

Figure 10 shows the output results of the proposed method for seven different languages namely English, Hindi, German, Japanese, Chinese, French, and Portuguese.

Fig. 10. (a) (i) Hindi

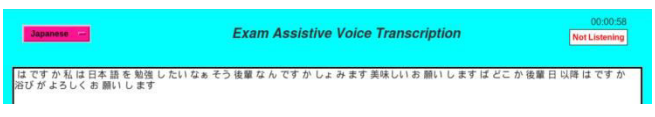
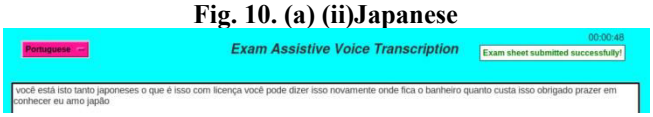


Fig. 10. (a) (ii)Japanese



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404029|

Fig. 10. (a) (iii) Portuguese

Fig. 10. (a) (iv) Chinese

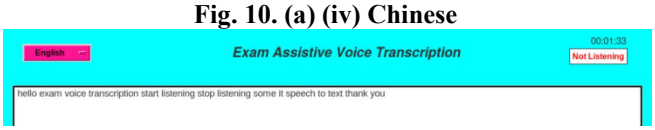


Fig. 10. (a) (v) English

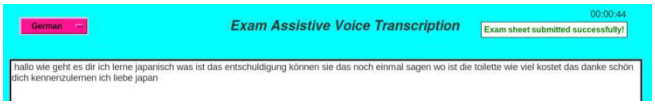


Fig. 10. (a) (vi) German

Fig. 10. (a) (vii) French

Fig.10(b) Text converted as PDF file

V. CONCLUSION

In conclusion, the proposed real-time voice-to-text transcription system successfully enhanced accessibility for students with disabilities during examinations. By using advanced speech recognition technology and a high-quality microphone, the system provided accurate and real-time transcription of student responses. The interactive features, such as the ability to review and edit transcriptions, ensured greater control over the process, improving accuracy and minimizing errors. The automatic PDF submission feature streamlined the evaluation process for examiners. Overall, the system fostered inclusivity and equity in education, providing students with disabilities the opportunity to demonstrate their knowledge effectively. Future improvements can focus on further enhancing transcription accuracy and expanding accessibility features to continue breaking down barriers in education.

REFERENCES

[1] Azizah, Kurniawati."Zero-Shot Voice Cloning Text-to-Speech for Dysphonia Disorder Speakers." IEEE Access Vol.99, no. 23, pp .5656-5249, 2024.

[2] Bargum, Anders R., Stefania Serafin and Cumhur Erkut. "Reimagining speech: a scoping review of deep learning-based methods for non-parallel voice conversion." Frontiers in Signal Processing, Vol.30, no. 8, pp. 13-19, 2024,

[3] Cao, Danyang, Zeyi Zhang and Jinyuan Zhang, "NeuralVC: Any-to-Any Voice Conversion Using Neural Networks Decoder for Real-Time Voice Conversion",IEEE Signal Processing Letters, Vol 79, pp. 65565- 69449, 2024.

[4] Choi, Jeong-Eun,Karla Schafer and Sascha Zmudzinski."Introduction to Audio Deepfake Generation: Academic Insights for Non-Experts." In Proceedings of the 3rd ACM InternationalWorkshop on Multimedia AI against Disinformation,Vol. 49, no. 20,pp. 3-12, 2024.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404029|

[5] Gabrys, Adam, Goeric Huybrechts, Manuel Sam Ribeiro, Chung-Ming Chien. "Voice filter: Few-shot text-to-speech speaker adaptation using voice conversion as a post- processing module."In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vol.25, no. 46, pp. 7902-7906. IEEE, 2022.

[6] Hongyu Gong, Ning Dong, Changhan Wang, Wei-Ning Hsu, Jiatao Gu,et.al."Unified speech-text pre-training for speech translation and recognition." arXiv preprint arXiv Vol.22, no. 2, pp. 400-409, 2022.

[7] Huang, Wen-Chin, Tomoki Hayashi, Yi-Chiao Wu, Hirokazu Kameoka, and Tomoki Toda. "Pretraining techniques for sequence-to-sequence voice conversion." IEEE/ACM Transactions on Audio, Speech, and Language Processing Vol.29, no. 3, pp. 745- 755,2021.

[8] Le, Phuong-Hang,Hongyu Gong, Changhan Wang, Juan Pino, Benjamin Lecouteux, and Didier Schwab. "Pre-training for speech translation: Ctc meets optimal transport." In International Conference on Machine Learning, Vol.37, no. 4, pp. 18667-18685. PMLR, 2023.

[9] Li, Haizhou."Transfer Learning from Speech Synthesis to Voice Conversion with Non- Parallel Training Data." Vol.95, no. 78, pp. 5746-3541, 2021.

[10] Li, Jingyi, Weiping Tu, and Li Xiao. "Freevc: Towards high-quality text-free one-shot voice conversion." In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vol.22, no. 2 pp. 1-5. IEEE, 2023.

[11] Ling, Shaoshi, Yuxuan Hu, Shuangbei Qian, Guoli Ye, Yao Qian, Yifan Gong, Ed Lin, and Michael Zeng. "Adapting large language model with speech for fully formatted end-to-end speech recognition." In ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vol.69, no. 9, pp. 11046-11050. IEEE, 2024.

[12] Reddy, B. Raghavendhar, and Erukala Mahender. "Speech to text conversion using android platform." International Journal of Engineering Research and Applications (IJERA) Vol.3, no.1, pp. 253-258, 2021.

[13] Reddy, V. Madhusudhana, T. Vaishnavi, and K. Pavan Kumar. "Speech-to-Text and Text-to- Speech Recognition Using Deep Learning." In 2023 2nd International Conference on Edge Computing and Applications (ICECAA), Vol.30, no. 64, pp. 657- 666. IEEE, 2023.

[14] Siddique, Zeba Qamaruddin. "Child Speech Synthesis using Deep Learning." PhD diss., Dublin, National College of Ireland, Vol.73, no. 9, pp .6453-7569, 2022.

[15] Tang, Yun, Juan Pino, Changhan Wang, Xutai Ma, and Dmitriy Genzel. "A general multi-task learning framework to leverage text data for speech to text tasks." In ICASSP 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vol.55, no.9, pp. 6209-6213, 2021.

[16] Wang, Qing, Hongmei Guo, Jian Kang, Mengjie Du, Jie Li, Xiao-Lei Zhang, and Lei Xie. "Speaker Contrastive Learning for Source Speaker Tracing." arXiv preprint arXiv: Vol.57, no. 43, pp. 2409.10072 (2024).

[17] Wang, Wupeng, Zexu Pan, Xinke Li, Shuai Wang, and Haizhou Li. "Speech Separation with Pretrained Frontend to Minimize Domain Mismatch." IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol.40, no. 93, pp. 08736-84375, 2024.

[18] Xinyuan, Y. U., and Brian Mak. "Non-parallel many-to-many voice conversion by knowledge transfer from a text-to-speech model." In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vol. 32, no. 8, pp. 5924-5928, 2021.

[19] Yang, Zeyu, Minchuan Chen, Yanping Li, Wei Hu, Shaojun Wang, Jing Xiao, and Zijian Li. "ESVC: Combining Adaptive Style Fusion and Multi-Level Feature Disentanglement for Expressive Singing Voice Conversion." In ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vol.98, no. 15, pp. 12161-12165. IEEE, 2024.

[20] Yang, Zijiang, Xin Jing, Andreas Triantafyllopoulos, Meishu Song, Ilhan Aslan, and Björn W. Schuller. An overview & analysis of sequence-to-sequence emotional voice conversion. arXiv preprint, arXiv: Vol.27, no. 4, pp. 252203.15873, 2022.

[21] Yu, Xinyuan. "Non-parallel Many-to-many Voice Conversion by Knowledge Transfer from a Pre-trained Text-to-speech Model." PhD diss., Hong Kong University of Science and Technology, Vol. 1, pp. 128-135, 2020.

[22] Zhang, Mingyang, Yi Zhou, Li Zhao, and Haizhou Li. "Transfer learning from speech synthesis to voice conversion with non-parallel training data." IEEE/ACM Transactions on Audio, Speech, and Language Processing Vol.29, no. 65, pp. 1290-1302, 2021.

[23] Zhang, Yuhao, Chen Xu, Bei Li, Hao Chen, Tong Xiao, Chunliang Zhang, and Jingbo Zhu. "Rethinking and improving multi-task learning for end-to-end speech translation."arXiv preprint arXiv: Vol.89, no. 6, pp. 2311-03810, 2023.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details