Augmentative Alternative Communication and Text-to-Speech for Dyslexia, Autism, and Parkinson’s using Machine Learning.

Mr. Manupati Jaideep[1], Mr. Embeti Saketh[2], Dr. T. Kishore Kumar[3], and Dr. K. Sunil Kumar[4]
[1]Department of Information Technology, National Institute of Technology, Srinagar, India
[2]Department of Electronics and Communication Engineering, National Institute of Technology, Srinagar, India
[3,4] Department of Electronics and Communication Engineering, National Institute of Technology, Warangal, India

Abstract—In the context of addressing communication challenges faced by individuals with diverse neurodevelopmental and neurological disorders, this project focuses on the development of an advanced speech synthesis system. Leveraging state- of-the-art technologies such as Convolutional Neural Networks (CNN) and Long Short-Term Memory Recurrent Neural Net- works (LSTM-RNN), the system is designed to produce natural and intelligible synthetic speech. This initiative is especially crucial for individuals dealing with visual impairment, Dyslexia, Parkinson’s disease, Autism Spectrum Conditions (ASC), and Amyotrophic Lateral Sclerosis (ALS). The proposed system aims to enhance communication accessibility and effectiveness for those with specific cognitive and motor challenges, contributing to a more inclusive and supportive digital environment.

I. INTRODUCTION

Effective communication is a universal challenge, espe- cially for individuals with speech disorders. This project aims to bridge this gap by developing a sophisticated Text-to- Speech (TTS) system tailored for people with diverse dis- orders. Utilizing advanced deep learning models, the system seeks to make synthesized speech sound more human-like and intelligible.

Dyslexia Definition: Neurological condition impacting reading, writing, and spelling despite average intelligence. Challenges: Difficulty decoding words, affecting reading fluency and comprehension. Autism Spectrum Conditions (ASC) Definition: Spec- trum of neurodevelopmental disorders affecting social inter- action, communication, and behavior. Challenges: Difficulty with social cues, verbal expression, and adapting to changes. Parkinson’s Disease Definition: Progressive neurodegen- erative disorder affecting movement. Challenges: Tremors, stiffness, and changes in speech impacting communication. Amyotrophic Lateral Sclerosis (ALS) Definition: Pro- gressive disease affecting nerve cells, leading to loss of muscle control. Challenges: Muscle weakness, paralysis, and difficulty speaking; often requires alternative communication methods

II. LITERATURE REVIEW

A. *Exploration*

The exploration into existing literature underscores the importance of Augmentative and Alternative Communica- tion (AAC) methods, statistical parametric speech synthesis (SPSS), and deep learning in creating effective solutions for communication barriers related to disorders.
Statistical Parametric Speech Synthesis (SPSS) The lit- erature emphasizes the significance of statistical parametric speech synthesis (SPSS) in creating tailored solutions for individuals with communication disorders. SPSS involves statistical modeling of speech characteristics, allowing for the synthesis of natural-sounding speech. This approach is particularly valuable in crafting personalized and contex- tually relevant synthetic speech, contributing to improved communication experiences for individuals facing speech related challenges.

B. *Augmentative and Alternative Communication (AAC) Methods*

AAC methods play a pivotal role in facilitating commu- nication for individuals facing disorders. These encompass a spectrum of tools and strategies designed to supplement or replace traditional verbal communication. Examples include picture boards, speech-generating devices, and sign language. Research in this area highlights the effectiveness of AAC in addressing the unique needs of individuals with diverse communication disorders, fostering increased accessibility and comprehension

C. *Statistical Parametric Speech Synthesis (SPSS)*

The literature emphasizes the significance of statistical parametric speech synthesis (SPSS) in creating tailored so- lutions for individuals with communication disorders. SPSS involves statistical modeling of speech characteristics, al- lowing for the synthesis of natural-sounding speech. This approach is particularly valuable in crafting personalized and contextually relevant synthetic speech, contributing to improved communication experiences for individuals facing speech-related challenges.

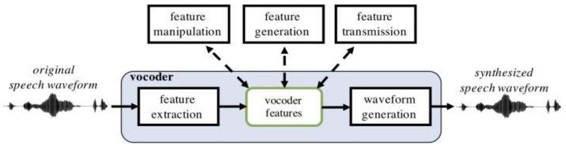


Fig. 1. Statistical Parametric Speech Synthesis Block Diagram.

D. *Deep Learning*

The exploration of existing literature underscores the growing influence of deep learning techniques in addressing communication barriers. Deep learning, characterized by neural network architectures such as Convolutional Neural Networks (CNN) and Long Short-Term Memory Recurrent Neural Networks (LSTMRNN), demonstrates a capacity to understand and generate human-like speech patterns. The literature highlights how leveraging deep learning method- ologies contributes to the development of advanced com- munication solutions, offering a more nuanced and adaptive approach to address the specific needs of individuals with disorders.

III. METHODOLOGY

Our methodology encompasses a thorough and inclusive approach, addressing key phases, including data collection, model integration, and implementation. To ensure accessi- bility and user friendliness, we leverage open-source tools, specifically the Tesseract Library and Raspberry Pi.

A. *Data Collection*

We initiate the process by collecting diverse datasets relevant to speech patterns and characteristics. This inclusive approach aims to capture a broad spectrum of linguistic nuances and variations.

B. *Model Integration*

Employing advanced deep learning models, particularly Convolutional Neural Networks (CNN) and Long Short- Term Memory Recurrent Neural Networks (LSTM-RNN), we integrate these models into our system. This ensures a robust framework capable of understanding and synthesizing natural speech patterns.

C. *Implementation*

The implementation phase involves translating theoretical models into practical solutions. To enhance accessibility and user-friendliness, we incorporate opensource tools such as the Tesseract Library. This allows for efficient optical character recognition (OCR) and facilitates the conversion of text into synthesized speech.

D. *Utilization of Open-Source Tools*

The Tesseract Library, known for its Optimal Character Recognition capabilities, is harnessed to interpret and process textual information. This aids in seamless conversion from written text to synthesized speech within the system.

E. *Raspberry Pi Integration*

To enhance the practicality and accessibility of our system, we utilize Raspberry Pi, a versatile and affordable single- board computer. This integration allows for the deployment of our Speech Synthesis System in various settings, making it more accessible to a broader user base.

IV. IMPLEMENTATION

Our Text-to-Speech (TTS) system is brought to life through the integration of advanced technologies, specifically Convolutional Neural Networks (CNN), Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN), and attention mechanisms. The implementation is meticulously designed with a primary focus on developing an intuitive interface that caters to individuals with diverse degrees of disorders.

A. *Convolutional Neural Networks (CNN)*

CNNs are utilized to extract hierarchical features from the input data. In our TTS system, CNNs play a crucial role in analyzing and understanding the complex patterns inherent in speech signals. This aids in capturing nuanced charac- teristics, contributing to the naturalness of the synthesized speech.

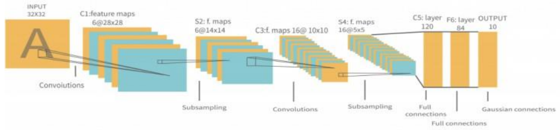


Fig. 2. Convolutional Neural Networks.

B. Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN)

LSTM-RNNs are employed to model temporal dependen- cies within the sequential nature of speech. This technology is well-suited for capturing long-range dependencies, ensur- ing that our system can generate coherent and contextually relevant synthetic speech.

C. Attention Mechanisms

Attention mechanisms enhance the model’s ability to focus on specific parts of the input sequence when generating the corresponding speech output. This ensures that the synthe- sized speech maintains a natural flow and effectively conveys the intended message.

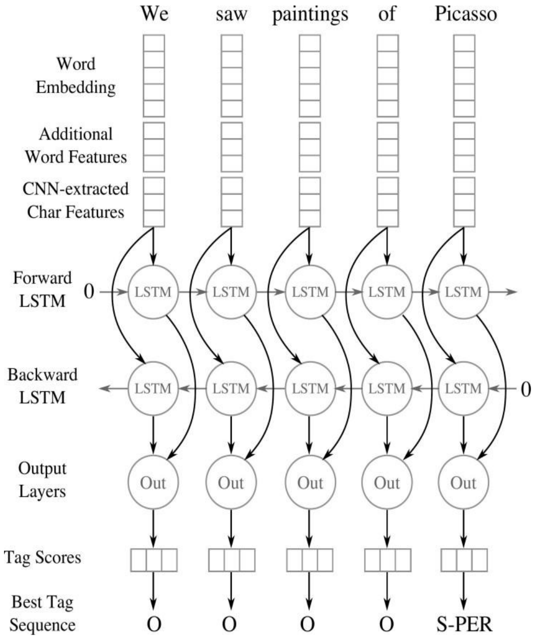


Fig. 3. Long Short-Term Memory.

D. User-Friendly Interface

The implementation places a strong emphasis on user experience. We design an interface that is not only tech- nologically sophisticated but also user-friendly, recognizing the diverse needs of individuals with varying degrees of disorders. The goal is to create an environment where users can interact seamlessly with the TTS system, promoting accessibility and ease of use.

V. RESULTS AND DISCUSSION

The outcomes derived from the synthesized speech pro- duced by our Text-to-Speech (TTS) system serve as a testa- ment to its success. In this section, we delve into a compre- hensive discussion, presenting our findings, and making com- parisons with existing literature and transparently addressing the challenges encountered during the development process.

- Synthesized Speech Results: The synthesized speech results showcase the effectiveness of our TTS system in generating natural, clear, and intelligible speech. Through rigorous testing and evaluation, we observe a notable advancement in the quality of synthetic speech, meeting the objectives set forth in the project.
- Comparisons with Existing Literature: Our discus- sion extends to a comparative analysis with existing literature, aiming to contextualize the advancements achieved by our TTS system. By benchmarking against established methodologies and technologies, we validate the innovation and uniqueness of our approach.

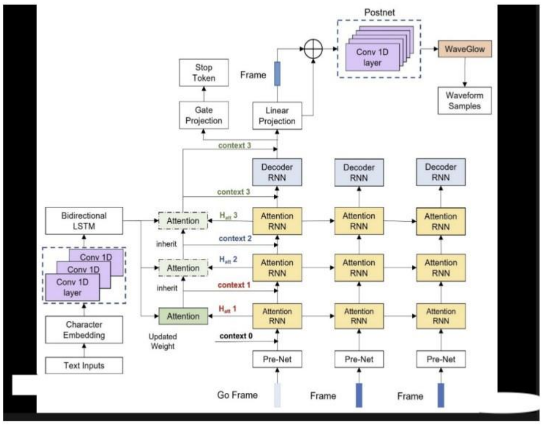


Fig. 4. Attention Mechanism.

- Addressing Challenges: An integral part of our dis- cussion involves a candid acknowledgment and explo- ration of challenges encountered during the develop- ment process. Whether technical complexities, dataset limitations, or unforeseen obstacles, we transparently detail our efforts to overcome these challenges. This transparency contributes to the iterative nature of tech- nological development.
- Implications of Findings: The implications of our findings are discussed in the context of the broader field of assistive technology and communication solutions for individuals with disorders. We explore how the success of our TTS system contributes to enhancing the overall landscape of tools available to

address communication barriers.

- User Feedback: User feedback, if available, is incor- porated into the discussion to provide insights into the practical usability and impact of the TTS system. User experiences and preferences play a crucial role in refining the system for real-world applicability.
- Future Directions: The discussion concludes with con- siderations for future research and development. Oppor- tunities for refinement and expansion of functionalities and potential collaborations are outlined, paving the way for continuous improvement and adaptation to evolving needs.

VI. RESEARCH PAPERS AND RESULTS

Our project draws inspiration and guidance from key research papers, notably those authored by students from IIT Madras working with ESPnet. By incorporating methodolo- gies established in these influential works, our project aligns with proven approaches in the field of speech synthesis and communication technology.

- Influence from IIT Madras Research: We actively engage with and implement methodologies outlined in research papers originating from IIT Madras students collaborating with ESPnet. These papers often bring forth innovative techniques, advanced algorithms, and best practices in the domain of speech synthesis and deep learning.

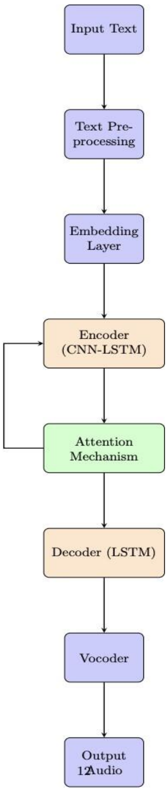
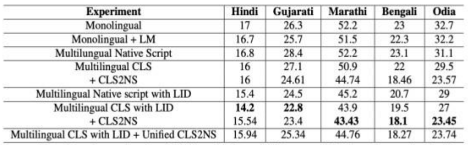


Fig. 5. Main Structure Block Diagram

- Methodological Integration: The integration of methodologies from these research papers is a deliberate and informed choice. We carefully adapt and implement techniques that have demonstrated efficacy in similar contexts. This ensures that our project benefits from the collective knowledge and advancements achieved by researchers in related fields.
- Proven Approaches: By aligning with established methodologies, we leverage proven approaches that have undergone scrutiny and validation in academic and research settings. This strategic alignment contributes to the robustness and reliability of our project.
- Results: The implementation of methodologies inspired by key research papers yields tangible results. Notably, our findings demonstrate a noticeable improvement in the naturalness and clarity of the synthesized speech. This improvement signifies a positive outcome and re- inforces the effectiveness of the selected methodologies.



Experiment	Hindi	Gujarati	Marathi	Bengali	Odia
Monolingual	17	26.3	52.2	23	32.7
Monolingual + LM	16.7	25.7	51.5	22.3	32.2
Multilungual Native Script	16.8	28.4	52.2	23.1	31.1
Multilingual CLS	16	27.1	50.9	22	29.5
+ CLS2NS	16	24.61	44.74	18.46	23.57
Multilingual Native script with LID	15.4	24.5	45.2	20.7	29
Multilingual CLS with LID	**14.2**	**22.8**	43.9	19.5	27
+ CLS2NS	15.54	23.4	**43.43**	**18.1**	**23.45**
Multilingual CLS with LID + Unified CLS2NS	15.94	25.34	44.76	18.27	23.74

Table 5: *WER comparison of Baseline and various Multilingual ASR model [No external LM is used except for the Monolingual + LM experiment]*

- Contribution to the Field: Beyond individual project success, our adoption of methodologies from influ- ential research papers contributes to the cumulative knowledge base of the broader academic and research community. It reinforces the collaborative nature of scientific progress and the importance of building upon established foundations.

VII. CONCLUSION

The culmination of our efforts results in the introduction of a resilient Endto-End Neural Text-to-Speech (TTS) system, meticulously designed to address the diverse communication needs of individuals grappling with disorders. This section serves as a comprehensive reflection on the project, covering key contributions, potential avenues for future work, and the profound positive impact on the quality of life for those experiencing speech disorders.

- Contributions: Our TTS system represents a substan- tial contribution to the field of assistive technology, particularly in addressing communication challenges associated with various disorders. By leveraging ad- vanced deep learning models, including Convolutional Neural Networks (CNN) and Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN), we have successfully engineered a system capable of producing natural and intelligible synthetic speech.

- Positive Impact on Quality of Life: The core focus of our project is to enhance the quality of life for indi- viduals with speech disorders. The TTS system serves as a tool to empower users, providing them with an effective means of communication. The positive impact extends beyond the technological realm, contributing to increased social engagement, personal expression, and overall well-being.
- Potential Implications for Future Work: While our project marks a significant achievement, there are av- enues for future exploration and enhancement. The discussion of potential implications for future work in- cludes considerations such as refining existing method- ologies, expanding the system's capabilities, and incor- porating user feedback to further tailor the technology to individual needs.
- Holistic Approach to Communication: By adopting an End-to-End Neural approach, our TTS system offers a holistic solution to communication barriers. It seam- lessly integrates data collection, model integration, and user-friendly interfaces to create an inclusive and sup- portive digital environment for individuals with varying degrees of disorders.
- Continued Collaboration and Research: The conclu- sion emphasizes the importance of continued collabo- ration and research in the field of assistive technology. Building upon our project, we advocate for ongoing efforts to explore additional deep learning architectures, expand datasets to cover a broader range of disorders, and collaborate with relevant stakeholders for real-world testing and feedback.

VIII. RECOMMENDATIONS

Our project serves as a foundation for future advancements in the realm of assistive technology for communication disorders. In this section, we propose key recommendations to guide further research and development, aiming to contin- ually improve and expand the impact of our Text-to-Speech (TTS) system.

- Exploration of Additional Deep Learning Architec- tures: Further research is warranted to explore and assess additional deep learning architectures beyond those implemented in the current TTS system, such as novel variations of neural networks or emerging architectures. This exploration can contribute to the re- finement of the technology, potentially uncovering more effective models for speech synthesis in individuals with disorders.
- Expansion of Datasets: To enhance the system's adapt- ability and inclusivity, there is a need to expand the dataset to cover a broader range of disorders. Including a more diverse set of speech patterns and characteristics will enable the TTS system to cater to a wider spectrum of communication challenges, ensuring its effectiveness across various conditions.
- Collaboration with Relevant Stakeholders: Collab- oration with relevant stakeholders, including individu- als with disorders, caregivers, healthcare professionals, and educators, is crucial for refining the TTS system based on real-world needs and experiences. Engaging in partnerships with these stakeholders provides valuable insights, ensuring that the technology aligns closely with the practical requirements and preferences of its intended users.
- User Feedback and Real-World Testing: Actively seeking user feedback and conducting real-world testing are integral components of refining and validating the TTS system. Incorporating the perspectives of end- users in different scenarios and environments allows for a more comprehensive understanding of the system's performance and usability. Continuous testing and re- finement based on feedback contribute to the creation of a more user-centered and effective technology.
- Adaptation to Emerging Technologies: As technology evolves, it is essential to stay abreast of emerging tools, platforms, and methodologies. Integrating the TTS sys- tem with the latest advancements in natural language processing, machine learning, and human-computer in- teraction ensures its relevance and effectiveness in a dynamic technological landscape.
- Ethical Considerations and Accessibility: Further re- search should emphasize ethical considerations related to the use of the TTS system, including privacy, secu- rity, and cultural sensitivity. Additionally, efforts should be directed towards ensuring the accessibility of the technology for individuals with diverse abilities and needs, including those with severe disorders or multiple impairments.

REFERENCES

[1] T. Schultz and A. Waibel, "Fast bootstrapping of lvcsr systems with multilingual phoneme sets," in Fifth European Conference on Speech

Communication and Technology, 1997.

[2] S. Thomas, S. Ganapathy, and H. Hermansky, “Multilingual mlp features for low-resource lvcsr systems,” in 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4269–4272, 2012.

[3] S. Watanabe, T. Hori, and J. R. Hershey, “Language independent end-to-end architecture for joint language identification and speech recognition,” in 2017 IEEE Automatic Speech Recognition and Un- derstanding Workshop (ASRU), pp. 265–271, 2017.

[4] S. Toshniwal, T. N. Sainath, R. J. Weiss, B. Li, P. Moreno, E. Weinstein, and K. Rao, “Multilingual speech recognition with a single end-to-end model,” in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4904–4908, 2018.

[5] A. Kannan, A. Datta, T. N. Sainath, E. Weinstein, B. Ramabhadran, Y. Wu, A. Bapna, Z. Chen, and S. Lee, “Large-scale multilingual speech recognition with a streaming end-to-end model,” arXiv preprint arXiv:1909.05330, 2019.

[6] S. Sivasankaran, B. M. L. Srivastava, S. Sitaram, K. Bali, and M. Choudhury, “Phone merging for code-switched speech recognition,” in Third Workshop on Computational Approaches to Linguistic Code- switching, 2018.

[7] K. Dhawan, G. Sreeram, K. Priyadarshi, and R. Sinha, “Investigating target set reduction for end-to-end speech recognition of Hindi-English code-switching data,” in 2020 National Conference on Communica- tions (NCC), pp. 1–5, 2020.

[8] A. Prakash, A. L. Thomas, S. Umesh, and H. A. Murthy, “Building multilingual end-to-end speech synthesizers for Indian languages,” in Proc. of 10th ISCA Speech Synthesis Workshop (SSW’10), pp. 194– 199, 2019.

[9] V. M. Shetty, M. Sagaya Mary N J, and S. Umesh, “Exploring the use of common label set to improve speech recognition of low resource Indian languages,” in ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021.

[10] V. M. Shetty and M. Sagaya Mary N.J., “Improving the performance of transformer based low resource speech recognition for Indian languages,” in ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8279–8283, 2020.

[11] D.-C. Lyu and R.-Y. Lyu, “Language identification on code-switching utterances using multiple cues,” in Ninth Annual Conference of the International Speech Communication Association, 2008.

[12] K. R. Mabokela and M. J. Manamela, “An integrated language identifi- cation for code-switched speech using decoded-phonemes and support vector machine,” in 2013 7th Conference on Speech Technology and Human-Computer Dialogue (SpeD), pp. 1–6, IEEE, 2013.

[13] J. Gonzalez-Dominguez, D. Eustis, I. Lopez-Moreno, A. Senior, F. Beaufays, and P. J. Moreno, “A real-time end-to-end multilingual speech recognition architecture,” IEEE Journal of Selected Topics in Signal Processing, vol. 9, no. 4, pp. 749–759, 2014.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is All You Need," in Advances in Neural Information Processing Systems, vol. 30, 2017.

[15] S. Watanabe, T. Hori, S. Karita, T. Hayashi, J. Nishitoba, Y. Unno, N. Enrique Yalta Soplin, J. Heymann, M. Wiesner, N. Chen, A. Renduchintala, and T. Ochiai, "ESPnet: End-to-End Speech Processing Toolkit," in Proceedings of Interspeech, pp. 2207–2211, 2018. [On- line]. Available: http://dx.doi.org/10.21437/Interspeech.2018-1456