SSNCSE@LT-EDI-2025:Speech Recognition for Vulnerable Individuals in Tamil

Sreeja K, Bharathi B
Department of Computer Science and Engineering
Sri Sivasubramania Nadar College of Engineering
sreeja2350625@ssn.edu.in
bharathib@ssn.edu.in

Abstract

Speech recognition is a helpful tool for accessing technology and allowing people to interact with technology naturally. This is especially true for people who wants to access technology but may encounter challenges interacting with technology in traditional formats. Some examples of these people include the elderly or people from the transgender community. This research presents an Automatic Speech Recognition (ASR) system developed for Tamil-speaking elderly and transgender people who are generally underrepresented in mainstream ASR training datasets. The proposed work used the speech data shared by the task organisers of LT-EDI2025. In the proposed work used the fine tuned model of OpenAI's Whisper model with Parameter-Efficient Fine-Tuning (P-EFT) with Low-Rank Adaptation (LoRA) along with SpecAugment, and used the AdamW optimization method. The model's work led to an overall Word Error Rate (WER) of 42.3% on the untranscribed test data. A key feature of our work is that it demonstrates potential equitable and accessible ASR systems addressing the linguistic and acoustic features of vulnerable groups.

1 Introduction

Automatic Speech Recognition (ASR) has made great advancements in the form of multilingual models like OpenAI's Whisper that demonstrate solid performance across a variety of languages and acoustic conditions. Nevertheless, these systems apply in a generic sense and perform poorly when a marginalized population, such as elderly and transgender speakers, is in a low-resource language setting like Tamil. These populations may have unique vocal characteristics, vocabulary, or fluency differences that are not represented in the standard ASR model training data.

To address this disparity, we must improve current ASR systems to better understand the linguistic and phonetic diversity of the group. Recent research describes how ASR systems can introduce significant bias, especially when the target user does not fit the general speaker types used when training a model (Koenecke et al., 2020). There are even greater disparities present in languages such as Tamil because there is rarely an opportunity to directly account for acoustic variability across speaker demographics at scale.

To this end, we explore a Parameter-Efficient Fine-Tuning (PEFT) method with a pretrained version of OpenAI's Whisper model, a state-of-the-art multilingual ASR system (Radford et al., 2022). PEFT approaches like Low-Rank Adaptation (LoRA) allow efficient fine-tuning by only fine-tuning a small subset of model parameters for computational efficiency while also allowing for the performance to be maintained (Hu et al., 2021). This also makes PEFT approaches well-suited for customizing ASR systems for underrepresented user groups in low-resource settings.

We will use this technique to transcribe Tamil speech from old and transgender speakers. We wish to assess inclusive and effective options for Whisper and contribute to equitable and accessible speech technologies.

The rest of the paper is organized as follows: Section 2 analyzes the related works done in the previous research, and Section 3 discusses the speech corpus used in the current work. Section 4 contains a detailed discussion of the proposed model. Section 5 explains the experimental results. Section 6 discusses the limitations. Section 7 concludes the paper.

2 Related work

Automatic Speech Recognition (ASR) for underrepresented and marginalized communities, such as elderly individuals and the transgender community, is an area that is still difficult and underdeveloped,

particularly in lower-resource languages such as Tamil. (B. Bharathi, 2025) provides an overview of the Fifth shared task on Speech Recognition for Vulnerable Individuals in Tamil. In recent years, there have been increasing efforts aimed at developing ASR systems that are more inclusive for such populations – the LT-EDI shared tasks (B et al., 2022, 2023, 2024) have been very helpful in gathering and sharing annotated datasets of Tamil speech data, with elderly and transgender speakers, and setting benchmarks to evaluate models. The paper (B et al., 2025) provides an overview of this shared task on speech recognition for vulnerable people. Dynamic models such as wav2vec 2.0 and Whisper have been refined in several experiments to consider these distinct speech patterns. A study conducted by (R et al., 2024), Tamil datasets and vulnerable speakers were optimized on Whisper and XLS-R models, and reported a word error rate (WER) of 24. 45%. The wav2vec 2.0 large-xls-r300m-tamil model provided a WER of 29.30% when using speech data that included elderly speakers (Suhasini and Bharathi, 2024). Other works comparing traditional transformer-based models (BERT, RoBERTa) with wav2vec reported WERs ranging from 37.71% to 40.55% (Saranya and Bharathi, 2023). Beyond the Tamil and the LT-EDI shared tasks, international work has been done to develop ASR for underrepresented groups across languages. The authors (Hu et al., 2023) used a domain-adapted self-supervised model (beginning with), wav2vec 2.0 as a feature extractor for the TDNN and Conformer ASR systems, finding lower WERs in dysarthric speech and elderly English-speaking speech corpora. The research (Zheng et al., 2024) pursued a different approach by fine-tuning ASR systems over speech data from people with Parkinson's disease (i.e., subglottic), and enhanced models as a multitask learning architecture to produce more accurate transcriptions, as well as prediction of symptom severity. A new approach to adaptation, spectrotemporal adaptation, was used by (Geng et al., 2022), where the models were specifically fine-tuned with the speech data of dysarthric and elderly speech, which outperformed all baseline adaptation strategies. An improvement in WER for disordered speech recognition was demonstrated by (Wang et al., 2023) by hyperparameter tuning with a set of Conformer models. In low-resource contexts (e.g., elderly Frisian speakers), when the authors used data augmentation and fine-tuned wav2vec 2.0 XLS-R (and trained the models with several learning rate adjustments), the findings indicated a 20% relative WER improvement when each audio sample was 50 seconds or longer. Finally, (Chen and Asgari, 2020) used transfer learning to train ASR on socially isolated seniors and approached the problem with attention mechanisms to improve robustness given the inevitable variability (e.g., cognitive and acoustic) with this population. In general, the studies discussed in this section promoted the importance of language, domain-aware, and inclusive adaptation approaches in the construction of ASR systems for marginalized groups of people.

In this work, our approach proposed an effective Automatic Speech Recognition (ASR) system for elderly and transgender Tamil speakers by applying Parameter-Efficient Fine-Tuning (PEFT) on the Whisper model. Unlike prior work that relies on full model fine-tuning, we proposed that only a small and hopefully manageable subset of parameters is adapted, which provides a cost-effective option while maintaining similar performance to the unwieldy fine-tuning of 176 million parameters. This approach provides a good way forward as ASR can scale to lower resource settings and thus provide an avenue for developers to address inclusive ASR. Overall, this work offers a simple and low-cost pragmatic approach to providing access to communities to the speech landscape.

3 Speech Corpus Description

The dataset is comprised of spontaneous speech data collected to support Automatic Speech Recognition (ASR) for vulnerable Tamil-speaking communities (especially elders and transgender people)(B et al., 2022). People in their older years commonly attend primary points such as banks, hospitals, and administrative offices, which address their needs in their daily lives, where communication is essential. This dataset includes 7 hours and 30 minutes of spontaneous Tamil speech from people whose mother tongue is Tamil. Recordings were done in controlled environments to ensure clarity of audio quality with no background noise or overlap. All files are in .wav format. We have approximately 5.5 hours of the audio transcribed, and as the training set, 2 hours is used for testing, and is un-transcribed. Table 1 describes the data distribution for training and testing.

Data	No. of utterances	Duration in Hours
Training	908	5.5
Testing	451	2
Total	1359	7.5

Table 1: Dataset Distribution

4 Proposed Methodology

This study presents a Tamil Automatic Speech Recognition (ASR) system specifically for elderly and transgender speakers and realizes the differences in linguistic characteristics and phonetic uniqueness that exist in their spontaneous speech. The above speaker groups are noticeably underrepresented in speech data, which can lead to poor or no performance by general ASR systems. To overcome this aspect, the methodology is proposed to adapt a large, pre-trained speech model by way of a parametrically efficient methodology.

A fine-tuned version of OpenAI's Whisper model, known as yaygomii/FYP_Whisper_PEFT_TAMIL (Saranya et al., 2025), is used as the base system. This model is adapted using Low-Rank Adaptation (LoRA), a technique under the Parameter-Efficient Fine-Tuning (PEFT) framework. LoRA allows a model to learn domain-specific features by introducing trainable low-rank matrices into pre-existing attention layers while freezing all but a small number of parameters. This reduces the computational burden and memory use in the training phase, making it a great option for low-resource scenarios.

The model was fine-tuned using 5.5 hours of Tamil speech, manually transcribed from elderly and transgender speakers. Given the relatively small dataset size, this methodology incorporates data augmentation to improve model generalizability. Use of SpecAugment by using time and frequency masking of the input spectrograms is included to simulate variability in speech patterns that may be present in actual recordings from the populations of interest. Background noise is imported into the training dataset to also help improve robustness for the model to handle real environmental conditions.

The goal of these augmentation strategies was to mitigate overfitting and enhance the model's generalization to denoting expected acoustic variation in spontaneous conversation and speech. Once the model was fine-tuned, transcript files underwent post-processing to improve their linguistic and contextual accuracy. Linguistic 'ripe' corrections pertinent to Tamil used domain-specific dictionaries and language model-based rules and corrected graphemic inconsistencies, removed common spelling mistakes while keeping contextual consistency in mind for the sentence; this is very important for a morphologically rich language like Tamil.

In addition, decoder prompts can be forced during the decoding stage, so that the model remains consistent with the Tamil language, particularly in multilingual settings. The proposed strategy focuses not only on improving recognition performance for a minority group but also illustrates a commitment to the ethical and inclusive design of speech technologies.

By adjusting the system to the phonetic, lexical, and syntactic tendencies of marginalized communities, the resulting template is intended to make ASR systems more equitable and representative. The method leaves room for future improvement by proposing the use of speaker embeddings, for greater personalization and context, and the use of semi-supervised learning, to utilize the rest of the untranscribed half of the dataset. These directions continue to align with the larger purpose of developing robust and accessible ASR systems to serve many speaker populations with integrity.

5 Experimental results

The effectiveness of the proposed Tamil ASR system was evaluated, and a 2-hour test set was developed subsequently using the recordings of spontaneous elderly and transgender speech that had not been transcribed. The recordings reflected the same demographic and acoustic characteristics of the data used for training.

Word Error Rate (WER) is the main evaluation metric in this study, which is the percentage of substitutions, insertions, and deletions required to match the system's output to the true output. After decoding the output and performing post-processing with Tamil-specific corrections and dictionaries for the reference, the model was evaluated, and the WER was 42.3% over the test set. Fig. 1 demonstrates the target and predicted sentences. The source code for the proposed approach and found here [1]

[1] https://github.com/SreejaKumaravel/Speech_

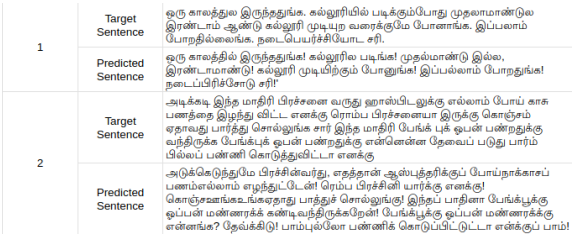

Figure 1: Sample target and predicted sentence

The model has shown evidence of transcribing Tamil speech from older adult and transgender participants while showing respect to dialectal and morphological ambiguities. Overall, the model's syntax and intention worked very well within the speech, but rare forms were difficult and therefore not transcribed well with this original achievement. The model achieved this level of performance in a resource-constrained language environment with LoRA tuning of the pre-trained model, data augmentation, and Tamil-specific post-processes.

6 Limitations

The current work faced several hurdles, including:

- Limited training makes generalization difficult across diverse communicative and acoustic varieties.
- Thus, High WER with pitying 42.3% means using rare word forms and dialectal variation, and complex sentence structures has looked difficult.
- Untranscribed data, which would prevent performance gains in semi-supervised methods.
- No speaker-specific adaptation, meaning the procedure ignores the possibility of speaker embeddings or speaker-specific personalization techniques.

7 Conclusions

This work presents an inclusive Tamil Automatic Speech Recognition (ASR) system tailored for elderly and transgender individuals, developed using a Parameter-Efficient Fine-Tuning (PEFT) approach on OpenAI's Whisper model. The system leverages a range of techniques, including data augmentation, SpecAugment, LoRA, and Tamil-specific post-processing, achieving a word error rate (WER) of 42.3% on spontaneous speech from marginalized communities. Despite limited resources, this work demonstrates the feasibility of building equitable, speech-based technologies in low-resource and underrepresented settings. Future directions include integrating speaker embeddings, employing semi-supervised learning to utilize untranscribed speech, and reducing recognition errors to further enhance accessibility for speakers with diverse speech patterns.

Recognition/blob/main/Automatic_speech_detection.ipynb

References

Bharathi B, Bharathi Raja Chakravarthi, Subalalitha Cn, Sripriya N, Arunaggiri Pandian, and Swetha Valli. 2022. Findings of the shared task on speech recognition for vulnerable individuals in Tamil. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 339–345, Dublin, Ireland. Association for Computational Linguistics.

Bharathi B, Bharathi Raja Chakravarthi, Subalalitha Cn, Sripriya Natarajan, Rajeswari Natarajan, S Suhasini, and Swetha Valli. 2023. Overview of the second shared task on speech recognition for vulnerable individuals in Tamil. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 31–37, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Bharathi B, Bharathi Raja Chakravarthi, Sripirya N, Rajeswari Natarajan, Rajalakshmi R, Suhasini S, and Swetha Valli. 2025. Overview of the fourth shared task on speech recognition for vulnerable individuals in tamil. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity, Inclusion*, Naples. Association for Computational Linguistics.

Bharathi B, Bharathi Raja Chakravarthi, Sripriya N, Rajeswari Natarajan, and Suhasini S. 2024. Overview of the third shared task on speech recognition for vulnerable individuals in Tamil. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion*, pages 133–138, St. Julian's, Malta. Association for Computational Linguistics.

N. Sripriya Rajeswari Natarajan Rajalakshmi R S. Suhasini B. Bharathi, Bharathi Raja Chakravarthi. 2025. Overview of the Fifth Shared Task on Speech Recognition for Vulnerable Individuals in Tamil. In *Proceedings of the Fifth Workshop on Language Technology for Equality, Diversity and Inclusion*, Italy. Fifth Conference on Language, Data and Knowledge (LDK2025).

Liu Chen and Meysam Asgari. 2020. Refining automatic speech recognition system for older adults. *Preprint*, arXiv:2011.08346.

Mengzhe Geng, Xurong Xie, Zi Ye, Tianzi Wang, Guinan Li, Shujie Hu, Xunying Liu, and Helen Meng. 2022. Speaker adaptation using spectro-temporal deep features for dysarthric and elderly speech recognition. *Preprint*, arXiv:2202.10290.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *Preprint*, arXiv:2106.09685.

Shujie Hu, Xurong Xie, Zengrui Jin, Mengzhe Geng, Yi Wang, Mingyu Cui, Jiajun Deng, Xunying Liu, and Helen Meng. 2023. Exploring self-supervised pre-trained asr models for dysarthric and elderly speech recognition. *Preprint*, arXiv:2302.14564.

Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, Dan Jurafsky, and Sharad Goel. 2020. Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences*, 117(14):7684–7689.

Jairam R, Jyothish G, Premjith B, and Viswa M. 2024. CEN_Amrita@LT-EDI 2024: A transformer based speech recognition system for vulnerable individuals in Tamil. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion*, pages 190–195, St. Julian's, Malta. Association for Computational Linguistics.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. Robust speech recognition via large-scale weak supervision. *Preprint*, arXiv:2212.04356.

S Saranya and B Bharathi. 2023. Sanbar@ lt-edi-2023: Automatic speech recognition: vulnerable old-aged and transgender people in tamil. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 155–160.

S Saranya, B Bharathi, S Gomathy Dhanya, and Aishwarya Krishnakumar. 2025. Real-time continuous tamil dialect speech recognition and summarization. *Circuits, Systems, and Signal Processing*, 44(4):2855–2881.

S Suhasini and B Bharathi. 2024. Asr tamil ssn@ lt-edi-2024: Automatic speech recognition system for elderly people. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion*, pages 294–298.

Tianzi Wang, Shoukang Hu, Jiajun Deng, Zengrui Jin, Mengzhe Geng, Yi Wang, Helen Meng, and Xunying Liu. 2023. Hyper-parameter adaptation of conformer asr systems for elderly and dysarthric speech recognition. *Preprint*, arXiv:2306.15265.

Xiuwen Zheng, Bornali Phukon, and Mark Hasegawa-Johnson. 2024. Fine-tuning automatic speech recognition for people with parkinson's: An effective strategy for enhancing speech technology accessibility. In *Interspeech 2024*, interspeech$_2$024, $page$2485˘2489.$ISCA$.