

Automatic Detection of Fluency Disorder in Children

Shrikant S. Joshi, Sachi Teke, Aditya Wayal,
Vivek Tripathi, Rutuja Kothawade

Mentor: Jing Su

Goal of the Project

- The broader goal of this project is to reach out to the remote population via mobile app to assist the people for the automatic detection of stammering or stuttering at early age.
- Considering Indian scenario, the expert medical practitioners and speech therapists are not available at remote locations but only available in big cities.
- A successful completion of this project will make it possible to reach out the needy remote population and medical assistance can be provided via widespread mobile network.
- Final goal is to develop a system which will automatically detect the fluency disorder mainly stuttering.

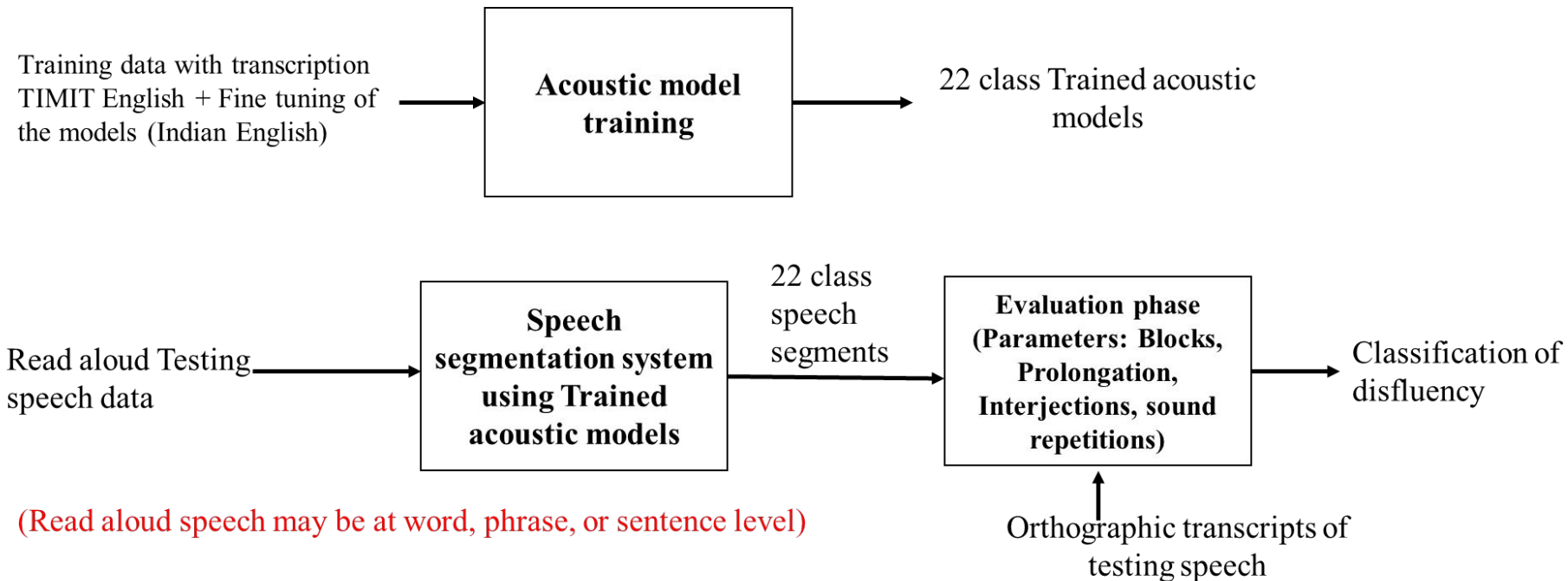
Review of literature

- Shakeel et al. (arXiv July 2022) presented a detailed review of speech disorder and importance of acoustic features, machine learning methods, challenges and future directions.
- Colin lea et al. (arXive Feb 21) elaborated various characteristics of the stutter speech and SEP-28K data set for speech disorder.
- Sebastian et al. (arXive 22), Chopra M et al. (2020), Tedd et al. (arXive 19) presented the work on automatic detection of speech fluency disorder.
- **Relevance of the literature to our approach:** The acoustic characterization and importance is very relevant because of our two stage approach and which would take into account the characteristics such as **speech rate**, **stop gap duration**, **VC transition duration**, **frication duration**, **VOT**, **vowel duration**, **formants**, **signal energy level** etc.

Performance evaluation

Stuttering Labels	Description
Block	Gasps for air or stuttered pauses
Prolongation	Elongated syllable or Sound “[llll]l”, “otherwi[ssss]se”
Sound Repetition	Repeated syllables “[nat-nat-nat-]naturally” or sounds “[t-t-t-]talked to dad”.
Word/Phrase Repetition	“I have [I have] done no such thing”
Interjection	Filler words e.g., “ähm”, “äh”, “naja”, eng: “uhm”, “uh”

Our approach



Dataset and system Implementation

Dataset: The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus for training.

GIE data consists of 20 speakers with 12 speakers used for adaptation and 8 for validation or testing.

Adaptation data consists of around **500 TIMIT sentence** prompts spoken by Indian english speakers while Testing set consists of **160 TIMIT sentence** prompts spoken by Indian english speakers (20 sentences each)

Acoustic modeling: GMM-HMM acoustic models (currently 22 classes including 11 vowels, semivowels, Nasals , 6 diphthongs, obstruents, silence and voice bar.

Toolkit use: Current experiments carried out using HTK 3.5. Plan to use open source KALDI

Results: 22 class phoneme correct recognition of **50%** and with adaptation **73%**

Challenges

- Considering **Indian scenario** with **several regional languages**, it is important to implement the ASR system which should take into account a **particular regional (group of) language (L1)**.
- Majority of the remote population uses **regional languages** and detection of **stuttering** at **early stage** should include L1 only.
- Annotated stutter speakers dataset is a main challenge considering Indian scenario.
- **Dataset:** Annotated adaptation and stutter data of regional languages.
- Acoustic model training with good recognition accuracy: For the proposed two stage system with the importance to extract accurate phoneme segments, the recognition accuracy required, should be significantly high.
- Increasing the number of phonemes/acoustic models with good amount of adaptation data will be useful to balance the tradeoff between accuracy and number of acoustic models.

References

- “Machine Learning for Stuttering Identification: Review, Challenges and Future Directions”, Shakeel A. Sheikh, Md Sahidullah, Fabrice Hirsch, Slim Ouni arXiv July 2022.
- “Detecting Disfluencies in Stuttering Therapy Using wav2vec 2.0”, Sebastian P. Bayerl, Dominik Wagner, Elmar Noth, Korbinian Riedhammer arXiv June 2022.
- “KSoF: The Kassel State of Fluency Dataset – A Therapy Centered Dataset of Stuttering”, Sebastian P. Bayerl, Alexander Wolff von Gudenberg, Florian Honig, Elmar Noth, Korbinian Riedhammer, arXiv June 22
- “SEP-28K: A DATASET FOR STUTTERING EVENT DETECTION FROM PODCASTS WITH PEOPLE WHO STUTTER”, Colin Lea*, Vikramjit Mitra*, Aparna Joshi, Sachin Kajarekar, Jeffrey P. Bigham, arXiv Feb 21
- “DETECTING MULTIPLE SPEECH DISFLUENCIES USING A DEEP RESIDUAL NETWORK WITH BIDIRECTIONAL LONG SHORT-TERM MEMORY”, Tedd Kourkounakis, Amirhossein Hajavi, Ali Etemad arXiv Oct 19