

## **Preface**

In the vast tapestry of linguistic diversity, Hindi emerges as one of the most widely spoken languages, resonating not only within the borders of India but also reverberating through the Indian diaspora scattered across the globe. This linguistic prominence underscores the importance of delving into the intricacies of speech-related challenges faced by Hindi speakers, with a particular focus on stammering. This research endeavors to navigate the uncharted waters of computational linguistic analysis of stammering in Hindi speakers, paving the path towards a nuanced understanding of linguistic characteristics unique to this community.

The research problem at the heart of this thesis centers on the paucity of existing computational linguistic works dedicated to complexities of stammering in Hindi. Despite the widespread prevalence of Hindi as a language, there exists a noticeable gap in the literature concerning stammering, hindering the development of targeted interventions and advancements in speech recognition technology tailored to the needs of this linguistic community. The absence of a comprehensive exploration of linguistic nuances in stammering among Hindi speakers underscores the urgency and significance of this research.

The research objectives outlined in this dissertation encapsulate the multifaceted approach adopted to address the identified research problem. Beginning with the formulation of an annotation schema designed to capture the granular classification of stammering-specific traits, the research unfolds into a linguistic analysis of a stammering audio corpus. This analysis seeks to unveil

linguistic cues indicative of instances where a subject is prone to stammer, culminating in the development of a binary classifier capable of distinguishing between stammering and non-stammering audio streams. These objectives collectively serve as pillars supporting the overarching goal of shedding light on the intricate relationship between language and stammering in the Hindi-speaking population.

The chosen methodology, a blend of linguistic analysis and computational experiments, lays the groundwork for a comprehensive investigation. With a primary focus on speakers of Hindi, the research involves the collection of speech samples from diverse stammering speakers, forming the backbone of a comprehensive dataset. A refined annotation schema, developed through expert consultation and multiple iterations, facilitates the capture of intricate stammering linguistic features. The subsequent steps involve transcription, cleaning, and annotation of the data, paving the way for the training of a binary classifier using advanced machine learning techniques. This classifier, integrated into Automatic Speech Recognition (ASR) pipelines, facilitates real-time decision-making and specialized processing for stammered speech.

The key contributions of this research are the creation of a novel annotation schema which not only enhances the granularity of stammering pattern documentation but also provides a solid foundation for future linguistic studies in this domain. Moreover, this research contributes to the growing body of knowledge on stammering, fostering empathy, awareness, and a binary classifier to distinguish between stammering and non-stammering audio signal. Recognizing that this work represents just the initial stride in a protracted journey, the aim is to promote inclusiveness, understanding, and effective communication for those affected by stammering in the Hindi-speaking community and beyond. In a creative leap, our research introduces the innovative concept

of a "Sleep Word" parallel to the well-known "Wake Word." This concept will help in making voice activated technologies inclusive for people with speech conditions.

As we embark on this academic odyssey, the hope is that this research serves as a catalyst for further exploration, fostering a deeper understanding of stammering in Hindi speakers and paving the way for transformative advancements in linguistic research and technology.