Chapter 4

Data collection and Annotation

4. Introduction

Annotation tag sets are essential for organizing and analyzing complex data, giving insightful explanations, and facilitating research across a variety of fields. The creation of precise and thorough annotation schemas is especially important in the field of speech and language processing. Researchers can better understand and analyze various speech phenomena thanks to the systematic labeling of audio data made possible by these schemas.

To address the shortcomings of current tag sets and offer a more in-depth understanding of stammering patterns, this chapter introduces a novel annotation schema that was created specifically for stammering audio data.

The limited nature of the stammering tag sets that have been developed has prevented them from consistently capturing the subtleties and complexity of this speech disorder. These shortcomings have made it more difficult to comprehend the underlying causes of stammering and create targeted treatments. As a result, a more thorough annotation schema is required to give researchers a solid toolkit to investigate stammering patterns at various levels.

A two-level strategy is offered by the proposed annotation schema, which includes tags at both the granular and surface levels. The schema enables the annotation of minute details and nuances of stammering at the granular level, facilitating a deeper comprehension of the speech disorder.

Additionally, the schema provides a thorough view of the speech disruptions at the surface level by capturing wider patterns and manifestations of stammering.

The proposed annotation schema enables researchers to find previously unnoticed patterns by providing a more thorough and nuanced characterization of stammering. With a deeper understanding, targeted therapy techniques can be created, treatment outcomes can be evaluated more easily, and better, more individualized speech recognition and synthesis systems can be created. The schema can also be a useful tool for interdisciplinary research, encouraging collaborations between specialists in linguistics, psychology, and other related fields as well as speech and language processing.

This chapter will review the available annotation tag sets for stammering audio data and point out their shortcomings. The design principles and factors that went into creating the suggested annotation schema will then be discussed. We will outline the schema's detailed organizational structure, describing the granular and surface-level tags and their intended uses. In addition, we will talk about the schema's advantages and potential applications. Readers will have a thorough understanding of the suggested annotation schema for stammering audio data by the end of this chapter, as well as its importance for furthering this field's research.

4.1 Participants

In this study, we engaged a group of fifteen male speakers, all aged between 18 and 25, with an average age of twenty-one. Interestingly, our efforts to involve female participants unfortunately yielded no success, resulting in a male-exclusive sample for this research endeavor. What caught our attention was a distinct trend among the male participants, showcasing a higher predisposition to instances of stammering compared to their female counterparts (Ambrose et al., 1997). This

gender-based variance in participation patterns can be linked to the complex interplay of societal stigmatization and persuasive social pressures, potentially discouraging women from taking part in academic investigations.

Our evaluative framework encompassed a comprehensive analysis of the SSI-4 parameters, delving into aspects such as the frequency, duration, associated physical manifestations, and the innate naturalness of verbal expression (Riley, 2009). Worth noting is the categorical classification of all participants falling within the mild-to-moderate spectrum in terms of the severity of their stammering condition. It is pertinent to highlight that none of the participants exhibited concurrent conditions that might have interfered with the integrity of the research findings. Additionally, it is worth underlining that the entire cohort unanimously identified Hindi as their primary native language (L1), while also reporting exposure to English as a secondary linguistic dimension (L2).

4.2 Data collection

In the pursuit of gathering comprehensive audio data for research purposes, a meticulous and ethically sound approach was undertaken. All participants involved in this study underwent telephonic interviews, during which audio samples were diligently procured. To address the intricate terrain of data collection ethics, a paramount principle of informed consent was scrupulously adhered to. Participants were unequivocally apprised of the research's overarching purpose and were solicitously asked to provide their consent for the recording of speech samples during the telephonic conversations. It is of paramount import to emphasize that participants were intentionally kept unaware of when precisely the audio recording would transpire during the conversation. This precautionary measure was meticulously implemented to mitigate any potential sources of bias that could emerge if participants possessed such foreknowledge.

Remarkably, a substantive volume of audio data, exceeding 6 hours in duration, was judiciously amassed from the participation of fifteen individuals. Within this sonorous corpus, a salient distribution of 70% constituted natural conversational speech, while the remaining 30% was dedicated to the reading of Hindi textual passages. A noteworthy feature of this data collection was the uniformity in the textual material presented to all research participants. The designed Hindi text ingeniously encapsulated the entirety of the Hindi phonemic inventory, thereby ensuring a holistic representation of the phonological facets of the language. This text comprised one hundred sentences, each characterized by an average length of seven words per sentence. The corpus exquisitely balanced the components of approximately 17 minutes of natural conversation and approximately 7 minutes of Hindi text recitation. Such meticulous orchestration of data collection is pivotal in rendering the dataset conducive to rigorous linguistic scrutiny and analysis.

In tandem with the acoustic data acquisition, a concomitant facet of the research methodology involved the elicitation of participants' subjective insights through the employment of a diary study approach. This methodological choice entails participants maintaining meticulous records of their own observations, findings, experiences, or behavioral manifestations over a specified temporal continuum. Through this meticulous documentation, participants concretized their narratives, particularly pertaining to the phenomenon of stammering. This qualitative layer of the study serves as a valuable complement to the quantitative audio data, providing a nuanced and holistic understanding of the linguistic landscape under investigation.

Notably, the research target population comprised native Hindi speakers who exhibited stammering tendencies. Given the relative dearth of extant research pertaining specifically to this demographic, the selection of telephonic conversations as the modus operandi proved to be a wise decision. Furthermore, it is imperative to underscore that participants were accorded an intricate

exposition of the research's nature and objectives prior to data acquisition. Their consent was explicitly secured, and they were duly apprised of the recording of telephonic conversations, a measure adopted to safeguard their privacy and uphold the tenets of research ethics. Additionally, participants received comprehensive elucidations regarding the intricate nuances of data analysis and its concomitant contribution to the uncharted realm of research concerning Hindi-speaking stammerers. This transparency not only elucidates the ethical underpinnings of the study but also empowers participants with a profound understanding of their invaluable role in advancing scholarship in this nascent domain.

The data acquisition process for this research endeavor was orchestrated with meticulous precision, encompassing a multifaceted approach designed to capture an extensive and diversified array of speech samples from the participants. This multifarious approach was categorized into three distinct modes of data collection:

- **Sentence Collection**

  In this methodological facet, participants were presented with a meticulously crafted corpus comprising more than one hundred sentences. Each of these sentences was thoughtfully curated to encompass a rich spectrum of Hindi phonological sounds and distinctive sound combinations. The primary objective underlying this strategy was the systematic evaluation of the participants' aptitude for enunciating these phonological elements within a rigorously controlled milieu. By listening to the recorded renditions of these sentences, we were equipped to discern and analyze the intricate nuances of participants' stammering patterns and the idiosyncratic phonological challenges that are intrinsic to the Hindi language. This approach, characterized by its controlled and structured nature, provided valuable insights into the participants' phonetic proficiency.

- **Informal Conversations**

  To foster an environment of ease and naturalness, researchers engaged in unstructured, informal conversations with the participants. These dialogues were purposefully designed to mitigate any potential inhibitions and induce a more relaxed atmosphere conducive to authentic communication. The underlying rationale was to elicit participants' natural speech patterns and stammering behaviors in the context of everyday discourse. These unscripted interactions sought to capture candid examples of participants' speech during spontaneous conversations, thereby complementing the controlled sentence collection method with a more ecologically valid representation of their linguistic capabilities.

- **Voluntary Recording Sharing**

  In addition to the aforementioned data collection methods, participants were accorded the option to voluntarily contribute recordings of their individual practice sessions. These self-initiated recordings allowed participants to express themselves in environments where they felt most at ease and confident. The objective here was to document the speech patterns and stammering characteristics that participants might manifest during their personal practice routines. By incorporating these self-practice sessions into our data collection repertoire, we strived to attain a holistic understanding of the participants' speech profiles, encompassing both controlled and unscripted scenarios.

  The amalgamation of these three distinct modes of data collection culminated in a comprehensive repository of speech samples, constituting the foundation upon which the analysis and investigation of stammering patterns among Hindi-speaking individuals are predicated. It is pivotal to underscore that participants were offered the prerogative to remain anonymous throughout the data collection process, should they wish to do so. Those who opted

for anonymity were assured that their identities would be safeguarded with utmost confidentiality, and no research reports or publications would disclose their personal information. Conversely, participants who opted to be identified were approached with the utmost respect, and their explicit consent was sought prior to acknowledging their substantial contributions in the thesis and research papers' acknowledgments sections. This duality in the approach to participant anonymity underscored the ethical underpinnings of the study and ensured that the research was conducted with the highest standards of integrity and sensitivity to the preferences of the participants.

4.2 Tag Set

Within this section, a brief explanation is provided about current annotation schemes used in the domain, a detailed exposition of the novel annotation scheme proposed herein, an intricate delineation of the overarching annotation framework that encapsulates the entire annotation process, and an exhaustive account of the specialized tool that was judiciously employed for the purpose of corpus annotation, thus providing a comprehensive overview of the foundational components underpinning the meticulous annotation endeavors.

- **Annotation schemes**

  Annotation schemes encompass a collection of tags utilized in the annotation process. These tags serve to impart supplementary or interpretive linguistic information alongside speech or text data. The specific nature of the study determines the focus of the information, which can range from phonetic and semantic to syntactic details. For instance, in text and speech corpora, Part-of-Speech (POS) annotation or phonemic boundary mark-up may be employed to annotate and enhance the linguistic content.

- **Existing annotation scheme**

  In previous chapters and in literature review, we explored research endeavors in languages other than Hindi that delve into the linguistic or computational facets of stammering. Table 1 presents a compilation of tags utilized in diverse research papers and articles addressing the intersection of stammering and machine learning (Jouaiti & Dautenhahn, 2022; Kourkounakis et al., 2021; Sheikh et al., 2021).

Tag	Meaning	Examples
I	Interjection	That is ummm my chair.
P	Prolongation	That is maaaaaay chair.
PH	Phrase repetition	That is this is my chair.
PW	Part-word repetition	T t t this is my chair.
R	Revision	That is my seat chair.
W	Word repetition	That this this is my chair.

Table 1: Existing Annotation Scheme

- **Proposed annotation scheme**

  Recognizing deficiencies in the prevailing annotation schemas during corpus collection and analysis, we introduced new tags to address missing features. Moreover, we put forth fine-grained tags as alternatives to ambiguous, coarse-grained tags. The proposed annotation scheme incorporates both coarse-grained and fine-grained tags, facilitating the efficiency of both high-level and low-level AI tasks. The following is a detailed list of tags along with their respective applications in the annotation process.

  - **Beyond word repetition (BWR)**

    This tag is relevant in situations where the speaker repeats the entire phrase to rectify the utterance or to circumvent stammering. Occasionally, subjects may opt to repeat only a few words, even if they do not constitute a complete phrase. Repetitions persist until the subject successfully concludes the sentence or opts for an alternative word or phrase.

  - **Blockage (BL)**

A blockage occurs when individuals who stammer experience tension and breathlessness. In such instances, speech comes to a halt, and the individual grapples to produce any sound. Despite the cessation of speech, the speech organs remain poised as if the subject is on the verge of vocalization. Blockages can endure for several seconds, placing the subject in a state of considerable emotional and psychological discomfort.

- **Deviated/Non-deviated (D/ND)**

  To distinguish instances of stammered speech and normal speech, we employ deviated and non-deviated tags, respectively. These coarse-grained tags play a key role in the binary classification of speech data. While the non-deviated tag stands alone without additional classification, the deviated tag incorporates fine-grained tags that correspond to distinct types of stammering. This hierarchical tagging system allows for nuanced classification, providing detailed insights into the specific nature of stammering occurrences within the speech data.

- **Filler (FI)**

  A filler, also known as embolalia, constitutes a component of formulaic language employed as a linguistic device to signify hesitation or manage turn-taking in verbal discourse, thus preventing confusion. Subjects utilize fillers when seeking alternatives in anticipation of potential stammering situations, contributing to the fluidity and coherence of their speech.

- **Heavy breathing (HB)**

  As an integral component of speech therapy for individuals who stammer, breathing techniques are taught to mitigate stammering occurrences. In instances where these techniques are employed to circumvent stammering, the individual's breath becomes audible. In cases of high-severity stammering, heavy breathing may occur after every individual word. It is considered normal for individuals who stammer to exhibit heavy breathing two or three times within a

single utterance. This tag encompasses both inhaling and exhaling types of breathing, providing a comprehensive annotation for breath-related aspects in the context of stammering.

- **Mixed (MX)**

  This tag is applied in situations where a person who stammers (PWS) exhibits multiple stammering traits simultaneously, making it challenging to establish distinct segmentation. It is employed to annotate instances where various stammering features co-occur without clear separation.

- **Phoneme repetition (PR)**

  A phoneme, explained as a distinct sound in simple terms, represents the minimal meaning-distinctive unit in a language. When a subject repeats the same phoneme multiple times, we designate that occurrence as phoneme repetition.

- **Prolongation (PL)**

  Individuals who stammer often exhibit the tendency to prolong certain sounds, with this phenomenon commonly observed in the production of sibilant sounds. During prolongation, successive sound frames in such instances tend to closely resemble the preceding sound, contributing to the distinctive pattern of repetition associated with stammering.

- **Silence (SL)**

  Individuals who are highly conscious of their stammering tendencies employ this technique. When these individuals anticipate difficulty in pronouncing a specific word, leading to potential stammering, they deliberately fall silent and mentally prepare themselves for the upcoming challenge. This cognitive process results in a sudden and noticeable silence for the listener. The duration of this silence can vary among individuals.

- **Syllable repetition (SR)**

Syllables, fundamental units of a word, may comprise one or more phonemes. Instances where a subject repeats the same syllable more than once are annotated as syllable repetition.

- **Word Dropping (WD)**

  This tag is applied in situations where subjects encounter difficulty with a particular word, making multiple unsuccessful attempts to pronounce it. Faced with this challenge, they opt to omit the problematic word and substitute it with an alternative to complete the sentence. In such cases, the deviated speech, marked by stammering instances, involves a different word compared to the non-deviated speech. This tag is utilized to capture and annotate these instances of word substitution during stammering occurrences.

- **Word repetition (WR)**

  This tag is employed to indicate instances where entire words are repeated within a single utterance. It is important to note that this tagging specifically excludes cases where reduplication is utilized as a linguistic device based on the grammatical properties of a language.

4.3 Knowledge representation in corpus

For organizing and gaining valuable insights from research data, effective knowledge representation is essential. A helpful tool in this regard is the three-tiered representation of stammered speech using the proposed annotation schema within PRAAT software (Boersma, 2001). The data can be organized into tiers that capture linguistic, binary, and fine-grained classifications, allowing researchers to quickly get a thorough overview of all the information that is currently available. This hierarchical representation enables detailed analysis of stammering severity and types, identification of deviated speech segments, and high-level comprehension of

the intended speech. Researchers can quickly gain insight into linguistic trends, classification patterns, and the subtleties of stammered speech by efficiently navigating and exploring the corpus. Such knowledge representation not only enables effective data analysis but also provides a basis for focused research inquiries, allowing researchers to explore particular facets of stammering in greater detail and form well-informed observations and interpretations.

We use a robust annotation schema that includes linguistic and para-linguistic tags to ensure that stammered speech is fully represented in the corpus. With the help of this schema, the gathered data can be arranged and analyzed in a way that allows for a multidimensional examination of stammering patterns. The data is organized into three tiers within the PRAAT software (Boersma, 2001), each of which is intended to offer more in-depth understandings of the characteristics of stammered speech.

- **Tier 1 Intended speech**

  Instead of concentrating only on the actual speech produced, the first tier, also referred to as the "Intended Speech" tier, is essential in capturing the essence of what people who stammer intended to communicate. Researchers can examine language trends and patterns present in stammered speech using this transcript of intended speech because it gives them a textual representation of the underlying linguistic content. By distinguishing between intended speech and deviated speech, we create a vital connection between the person's desired message and the difficulties they encounter in expressing it.

- **Tier 2 Binary classification**

  Our annotation schema includes two distinct tags: "Deviated" and "Non-deviated" for the "Binary Classification" tier. This level of classification allows for a binary distinction between speech segments that stammer and those that do not. For training machine learning algorithms

and creating models specifically for identifying stammered speech, this tier's binary classification data is an invaluable resource. We open the way for automated classification systems that successfully identify instances of stammering by differentiating between deviated and non-deviated speech.

- **Tier 3 Low-level classification**

  The "Low-level Classification" tier, which is the third tier, focuses on offering fine-grained tags to capture the complex nuances of stammered speech. This level includes a comprehensive set of annotations that allow for a fine-grained analysis of stammering severity and the identification of particular stammering subtypes. The data gathered in this tier has enormous potential for the creation of specialized tools and algorithms, such as the classification of various stammering types and the assessment of stammering severity. This degree of specificity enables more in-depth comprehension of the various stammering manifestations, facilitating targeted therapeutic interventions.

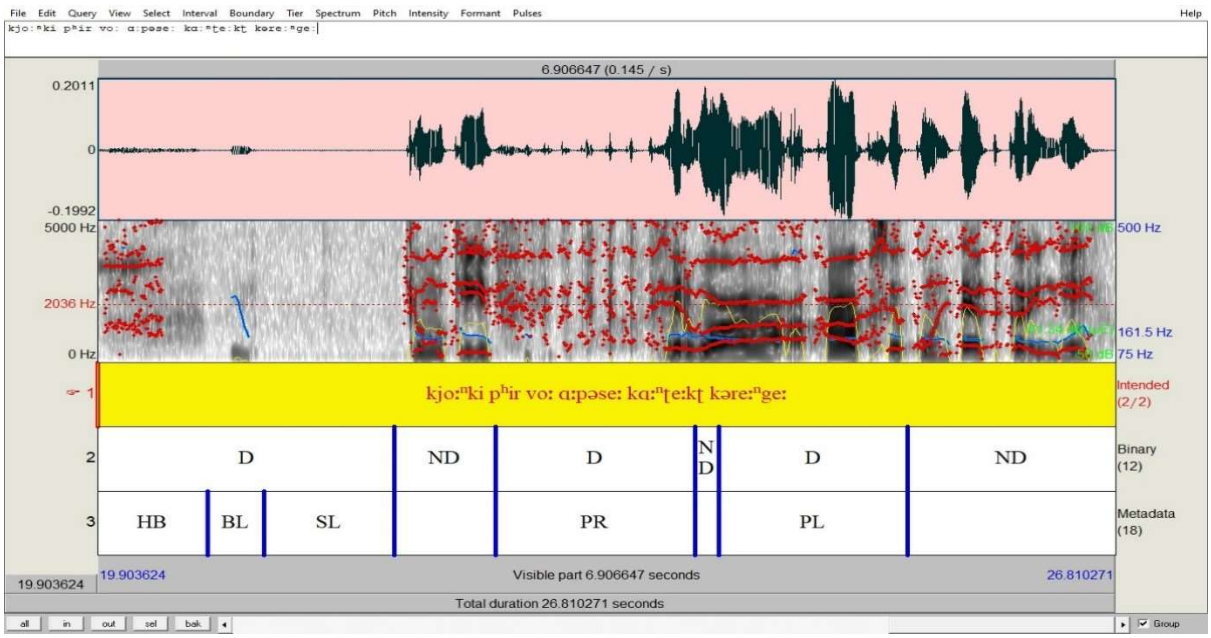

Figure 1: A sentence with three-tier analysis in PRAAT

We thoroughly investigated a variety of open-source tools during the annotation and corpus building process. We carefully considered our options before deciding that PRAAT would be the best tool for annotation. Due to its feature-rich interface, widespread acceptance in the research community, and user-friendliness, PRAAT stood out among the alternatives (Boersma & Weenink, 2007). We were able to efficiently annotate and organize the stammered speech data thanks to its wide range of functionalities and user-friendly design. The selection of PRAAT as our annotation tool turned out to be crucial in facilitating the efficient operation of the annotation process and the ensuing corpus construction.

We create a reliable framework for representing stammered speech within the corpus by utilizing the capabilities of PRAAT software and putting the suggested tag set into practice. Researchers can gain a thorough understanding of stammering patterns thanks to this meticulous and multi-tiered approach, which includes linguistic analysis of intended speech, binary classification of speech segments that deviate from the norm, and in-depth characterization of stammering severity and types. An effective tool for advancing stammering research is the use of the annotation schema within PRAAT software, which enables a thorough examination of the complexities present in this communication disorder.

4.4 Conclusion

As a result, creating a thorough and detailed annotation schema for stammering audio data has a lot of potential to help us learn more about this difficult speech disorder. The proposed schema offers researchers a potent toolkit to investigate the intricate nature of stammering patterns by addressing the shortcomings of existing tag sets and incorporating tags at both the granular and surface levels.

The annotation of minute stammering characteristics and subtleties is made possible by the granular-level tags, allowing for a thorough analysis of particular speech disruptions like repetitions, prolongations, and blocks. This degree of specificity gives researchers a better understanding of the temporal and spectral aspects of stammering, illuminating underlying mechanisms and assisting in the development of more focused intervention methods.

The surface-level tags also identify more general stammering manifestations and patterns, such as speech rate, severity, and corresponding actions or reactions. These labels offer a comprehensive understanding of stammering, facilitating research into how the disorder affects overall speech production and communication. The annotation schema's combination of granular and surface-level tags ensures a thorough and in-depth investigation of stammering phenomena.

Surface-level tags simultaneously record more extensive trends, and the proposed annotation schema has numerous advantages and uses. It is a helpful tool that enables stammering researchers and clinicians to precisely analyze and compare stammering patterns across individuals, contexts, and interventions. The schema also shows promise for the creation of individualized therapeutic strategies because it enables the accurate characterization of stammering profiles and the discovery of targeted tactics for those who stammer.

The annotation schema advances the field of automatic speech synthesis and recognition. It supports the training and improvement of algorithms that can precisely recognize and process stammered speech by providing a mechanism to annotate stammering data. This has the potential to improve assistive tools and programs that help people who stammer communicate effectively.

A crucial step toward better understanding and analysis of this challenging speech disorder is the suggested annotation schema for stammering audio data. Researchers can gain new knowledge,

create focused interventions, and advance the study of stammering by capturing the subtleties of stammering at various levels. The potential impact of the schema goes beyond research and therapy and includes the fields of technology and assistive devices. This annotation schema will support interdisciplinary partnerships and advance the study of stammering as more research and development are conducted.

With a refined dataset in hand, the subsequent chapter delves into the linguistic analysis. This analysis aims to uncover patterns, correlations, and linguistic nuances associated with stammering in Hindi speakers.