Our Collective Voices: The Social and Technical Values of a Grassroots Chinese Stuttered Speech Dataset

Jingjin Li*
jingjin@aimpower.org
AImpower.org
Santa Clara, California, USA

Qisheng Li*
qishengli@aimpower.org
AImpower.org
Seattle, Washington, USA

Rong Gong
rong.gong@stammertalk.net
StammerTalk
Austria

Lezhi Wang
lzwangcn@gmail.com
StammerTalk
USA

Shaomei Wu
shaomei@aimpower.org
AImpower.org
Mountain View, California, USA

Abstract

The lack of authentic stuttered speech data has significantly limited the development of stuttering friendly automatic speech recognition (ASR) models. In previous work, we collaborated with StammerTalk, a grassroots community of Chinese-speaking people who stutter (PWS), to collect the first stuttered speech dataset in Mandarin Chinese, containing 50 hours of conversational and command-recitation speech from 72 PWS. This work examines both the technical and social dimensions of the dataset. Through quantitative and qualitative analysis, as well as benchmarking and fine-tuning ASR models using the dataset, we demonstrate its technical value in capturing stuttered speech at an unprecedented scale and diversity – enabling better understanding and mitigation of fluency bias in ASR – and its social value in promoting self-advocacy and structural change for PWS in China. By foregrounding lived experiences of PWS in their own voices, we also see the potential of this dataset to normalize speech disfluencies and cultivate deeper empathy for stuttering within the AI research community.

CCS Concepts

• **Human-centered computing** → **Accessibility**; **Human computer interaction (HCI)**; • **Computing methodologies** → *Artificial intelligence*.

Keywords

AI FATE, datasets, benchmark, speech technology, accessibility, stuttering, stuttered speech

ACM Reference Format:

Jingjin Li, Qisheng Li, Rong Gong, Lezhi Wang, and Shaomei Wu. 2025. Our Collective Voices: The Social and Technical Values of a Grassroots Chinese Stuttered Speech Dataset. In *The 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25), June 23–26, 2025, Athens, Greece*. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3715275.3732179

*Both authors contributed equally to this research.

FAccT '25, Athens, Greece

ACM ISBN 979-8-4007-1482-5/2025/06
https://doi.org/10.1145/3715275.3732179

1 Introduction

Trained and optimized for "typical" speech, current automatic speech recognition (ASR) systems performs poorly for people with speech diversities, such as older adults [42, 58], people with speech and hearing disabilities [17, 24, 28], second-language speakers [23, 61], and African Americans [25]: these systems often cut them off from speaking and interpret their speech with a multitude higher error rates than average. As ASR becomes a ubiquitous part of today's communication ecosystem – powering smart speakers, in-car navigation, and automated phone menus – its inability to handle diverse speech not only creates access barriers, but also leads to psychological harms [8, 60] and socioeconomic disadvantages [14, 64].

People with stutter are among the groups most profoundly affected by ASR's fairness and accessibility issues [3, 28], yet they remain underrepresented in today's AI fairness, accountability, transparency, and ethics (FATE) discussions. To address ASR's bias against stuttering, we have taken a community-led, participatory approach to create the first and largest Mandarin Chinese stuttered speech dataset, containing nearly 50 hours of stutterer-to-stutterer conversations and voice command dictations from 72 Chinese speaking adults who stutter [32].

Our previous work detailed how grassroots, community-led data collection not only empowered an otherwise marginalized and socially isolated population, but also fostered solidarity and collective agency—laying the groundwork for community-driven technological agenda beyond a single dataset [32].

Extending our previous work, this work examines the technical and social values of the resulting dataset. Through descriptive statistical analysis, as well as the evaluation and fine tuning of state-of-the-art ASR models using this dataset, we show the unique ability of the community-created stuttered speech data to capture the heterogeneity and variability of stuttering, highlighting its efficacy in uncovering and reducing fairness issues in existing ASR models. Our content analysis of the conversations recorded in the data reveals the significant and intersectional challenges PWS face in China, including pervasive social stigma, overt discrimination, mental health impacts, and a lack of access to scientific knowledge of and professional support for stuttering. As such, this dataset also holds profound social and educational value as a platform for self-advocacy and public discourse on stuttering in China, promoting broader understanding and deeper empathy with the stuttering community from AI developers.

Building on existing FAccT research on the values embedded in maching learning datasets [22, 40], our work underscores the significance of self-expression and self-advocacy in the Chinese stuttered speech dataset, beyond its technical utility. In contrast to expert-led speech data collection, where speakers are often asked to respond to generic prompts or read random text passages [2, 35], our community-led approach empowered participants to share not only their speech, but also their values and lived experiences with the AI community.

Presented as a case study of emerging grassroots, community-led efforts to create more fair and inclusive AI models for, with, and by those directly impacted, our work also proposes an alternative model for AI fairness research – one that is non-extractive and capacity-building for the impacted community, and demonstrates its effectiveness to produce high-quality, much-needed technical artifacts, such as the Mandarin stuttered speech dataset and the fine-tuned ASR models. Our findings contribute to the ongoing AI FATE discussions around power, participation [11], and the fair representation of marginalized groups in AI data [40, 43].

Finally, our work advocates for the practice of crip technoscience [20] in AI FATE research. It enables us to examine the often implicit and unseen biases within AI systems through the magnifying lens of disabilities, while also harnessing the creativity and resourcefulness of the disability community, who have long specialized in designing and improvising their environments. Complementing existing disability-centered research within FAccT that has surfaced ableist biases in large language models [15] and identified technical flaws in advanced speech AI models [24], our study not only uncovers the usability and social harms faced by PWS in mainstream speech AI systems, but also showcases the promise of PWS-led solutions to meaningfully address these harms.

2 Related Work

2.1 Stuttering and Speech AI Technologies

Stuttering is a neurodevelopmental condition that affects more than 1% of the world population [39]. While stuttering in adulthood is incurable, it often creates significant behavioral, emotional, and cognitive challenges for people who stutter [4]. These challenges, compounded with social stigma and discrimination towards stuttering [6–8, 12], can lead to a reduced quality of life in many aspects, including mental health, social relationships, education, and employment opportunities [6, 9, 19, 66].

Unlike many other speech disorders, stuttering is characterized by high variability and unpredictability [53, 54]. The pattern and frequency of stuttering vary not only across individuals but also within the same speaker across different contexts [54]. This heterogeneity poses particular challenges for automatic speech recognition (ASR) systems, which often produce word error rates (WER) three to four times higher for stuttered speech than for fluent speech [28]. Even when ASR systems manage to transcribe stuttered speech, the outputs often omit disfluencies such as fillers, repeated words, and long pauses [28, 33], all of which are defining characteristics of stuttering [4]. Such default omissions deny PWS the agency to authentically represent what they actually said in transcriptions and reinforce existing structural marginalization of stuttering.

Drawing from the perspectives from the neurodiversity movement [49], the stuttering community has actively pushed back against ableist expectations of speech fluency, advocating for greater acceptance and respect for stuttered speech in everyday communications [9, 10, 51] and in speech-related technologies [31, 55, 64, 65]. While a previous survey study found PWS in China reported more adverse experiences related to stuttering compared to those in Western or more developed countries [34], the content of the community-created dataset not only captures intimate, personal accounts of the lived experiences of stuttering in China, but also serves as a platform for self-advocacy to drive structural change within the Chinese stuttering community.

2.2 Stuttered Speech Datasets

To address the lack of diverse datasets for inclusive and robust ASR models [36, 38, 69], there have been several industry initiatives to collect diverse speech samples across languages [2], accents [2, 37], and speech disabilities [35, 57]. Despite these efforts, high-quality stuttered speech data remain scarce in the public domain.

Considered highly sensitive personal data, corporate-sponsored stuttered speech datasets are often inaccessible to the broader research community [28, 35, 38, 48, 69]. Datasets collected by academic researchers, such as FluencyBank [45] and UCLASS[21], were limited in sizes and annotation consistency, as they were originally developed for speech therapy [29]. Open and scalable datasets, such as SEP-28k [29] and LibriStutter[26], also have various shortcomings in terms of annotation completeness, representativeness, and authenticity. Collected from public podcasts by people who stutter, SEP-28k consists of 28K 3-second audio clips labeled with only stuttering events (i.e. whether the clip contains prolongation, sound repetition, etc.) but lacks text transcriptions [29]. While LibriStutter does provide transcriptions, it contains no real speech from PWS but synthetic stuttering utterances (e.g. repetitions, prolongations, interjections) injected into audio books read by fluent speakers [26]. By introducing StammerTalk dataset, the first and largest corpus of stuttered speech in Mandarin Chinese, our work advances efforts to represent stuttered speech and its community in AI, and underscores the importance of community-driven data collection to authentically capture stuttering's diversity and amplify the community's voice in AI development.

2.3 Data Justice

Recognizing "existing power asymmetries and inequitable or discriminatory social structures" surrounding personal data [30], there have been increasing scholarly and policy investments on the conceptualization and practice of *data justice*. While most legal (e.g., General Data Protection Regulation) and technical (e.g., the Data Transfer Project) tools focus on individuals' control over their personal data, these tools often require substantial legal and technical capacities that are out of reach for individuals from marginalized groups [63]. In response, a group of collective data models, such as data trusts [41], data cooperatives[41], data commons[46], and data sovereignty[59], have been proposed to shift control form individuals to communities. However, most of these models remain conceptual or demand significant operational overhead beyond the capacities of grassroots communities.

Rather than developing a new data model, we situate our work within a growing body of grassroots, community-led data efforts that mobilize participation around shared values and social goals, often outside traditional legal frameworks. For example, the Quotidian Report [1] used Facebook for citizen-driven crime reporting in Mexico, while 996.ICU [1] leveraged GitHub to protest exploitative labor practices in China's tech industry. These initiatives exemplify bottom-up, issue-driven data collection, often leveraging mainstream platforms for collective expression and advocacy.

Our work takes a different approach: one that centers disability, builds community power, and reimagines data contribution not merely as extraction but as a vehicle for collective agency and advocacy. Specifically, we present the technical and social value of a community-led, grassroots case study that foregrounds a power-sharing model for AI development. Unlike typical data justice efforts where affected communities are consulted but not co-creators, our approach enabled people who stutter to design, curate, and govern a Mandarin stuttered speech dataset and its downstream applications. This model responds to growing FAccT conversations around participatory AI governance [11] and fair representation of minority groups in AI models [43]. Our work also contributes to ongoing debates in data ethics by foregrounding the expressive and communal dimensions of data production. In line with research on values embedded in ML datasets [40], we argue that data from marginalized communities must be collected and used not only for technical gains but also in ways that uphold contributors' dignity, values, and goals.

3 Method

The dataset was created by StammerTalk (口吃说) community[2], an online, grassroots community of Chinese speaking people who stutter. Speech data collection was conducted by two StammerTalk volunteers, who also stutter, with participants over videoconferencing platforms. The recorded speech contains both unscripted conversations between the volunteer and the participant, and the dictation of a list of 200 voice commands by the participant. 70 adults who stutter (AWS) participated in the recording with two StammerTalk volunteers, resulting in a dataset of 48.8 hours speech from 72 AWS. The recorded speech was transcribed semantically and verbatim, with five distinct stuttering event annotations embedded in markups. Obtaining verbatim transcription that includes word repetitions (e.g. "My, my, my name") and interjections (e.g. "hmm") was a deliberate choice made by the StammerTalk community, to allow disfluencies rather than automatically erased by ASR models. The annotation was performed by professional speech data annotators, and reviewed by a StammerTalk volunteer. More details about the data collection and annotation process can be found in previous work [18, 32].

To understand the characteristics and quality of the StammerTalk dataset, we perform the following quantitative and qualitative analysis on its technical and social properties.

[1]https://github.com/996icu/996.ICU
[2]https://www.stammertalk.net/

3.1 Quantitative analysis

We first conduct descriptive analysis of the StammerTalk dataset, comparing its scale and speech diversity with existing stuttered speech datasets. We also benchmark the performance of prominent ASR models with our dataset to assess and diagnose ASR biases towards stuttered speech. Lastly, we fine-tune OpenAI's Whisper model [44] using the StammerTalk dataset and demonstrate the its efficacy in improving ASR performance on stuttered speech.

3.1.1 Descriptive analysis. Stuttering is not a monolith. The frequency and types of stuttering can vary significantly across individuals and situations - a common source of insecurity and frustration for PWS [54]. While existing stuttered speech datasets often fall short in scale and representation of the heterogeneity within stuttering [28], we measure the scale and diversity of the StammerTalk dataset in terms of speakers, speaking tasks, stuttering frequency and severity, and speech variability between and within speakers.

3.1.2 Benchmarking. To understand ASR's ability to transcribe and respect speech disfluencies, we audit two state-of-the-art ASR services – Whisper (v2-large)[3] and wav2vec2.0 (large-chinese-zh-cn)[4] – with two types of ground truth transcriptions: 1) a **semantic** transcription with word repetitions and interjections excluded; 2) a **literal** transcription with the stuttered utterances kept verbatim. We remove all stuttering event markups in both cases.

We calculate the character error rate (CER), a metric commonly used to measure the ASR performance in Mandarin Chinese, using both semantic and literal transcriptions as references. CER measures the errors in model generated transcriptions at the character level, including substitutions (SUB), insertions (INS), and deletions (DEL).

3.1.3 ASR model fine tuning. To understand the effectiveness of StammerTalk dataset in enabling ASR models to preserve disfluencies in their transcriptions, we fine-tuned the LoRA adapter for the Whisper-v2-large model [67] on the StammerTalk Conversation dataset using literal transcriptions.

Data split. We split the data by participants. Given the relatively small number of participants in each stuttering level (20 – mild, 44 – moderate, 6 – severe), we only perform three-fold validation, with each fold containing a roughly 65:10:25 split for train/dev/test and a balanced representation of mild, moderate, and severe stuttering levels in each split. As a result, each fold contains 12 mild, 30 moderate, and 3 severe stuttering participants for train; 2 mild, 4 moderate, and 1 severe participants for dev; and 6 mild, 10 moderate, and 2 severe participants for test. This split strategy ensures robust evaluation of the model's performance across all severity levels.

Setup. We use four NVIDIA A100 80G GPUs with the following configurations: global batch size: 16; warmup steps: 50; learning rate: 0.001; trained epochs: 3; fine-tune method: `AdaLora`; target modules: `transformer`, `k`, `q`, `v`, `output_linear`, `fc1`, `fc2` (fully connected); trainable parameters: 21.6M; use Huggingface `PEFT`.

Fine-tuning is performed with a training objective to minimize the character-level transcription errors as to preserve disfluencies such as word repetitions and interjections. The model is fine-tuned

[3]https://github.com/openai/whisper
[4]A fine-tuned version of wav2vec2.0 optimized for Mandarin speech, see https://huggingface.co/wbbbbb/wav2vec2-large-chinese-zh-cn

using 3 epochs, with early stopping applied based on the validation loss to avoid overfitting. Training hyperparameters include a learning rate of 1e-3, batch size of 16, and the AdamW optimizer.

The fine-tuned model's performance is evaluated on the held-out test set using the same character error rate (CER) metrics as in the benchmarking task. The CERs are also broken down by severity level to assess how well the fine-tuned model handles varying degrees of disfluencies compared to the baseline Whisper model.

3.2 Qualitative analysis

The StammerTalk dataset is unique as it contains of 70 spontaneous conversations between two people who stutter [32]. While the conversations were unscripted, most of them naturally converged on shared experiences and personal stories around stuttering [32], making the StammerTalk dataset the first public archive of lived experiences of PWS in China to the best of our knowledge. To unpack the collective narratives captured in the StammerTalk dataset, we used an inductive open-coding analysis approach [47] to conduct the content analysis of recorded conversations. Our qualitative analysis consists of the following steps:

(1) First, the first two authors and the last author independently reviewed the transcripts of the first five participants and generated initial codes by adding comments directly to the documents. For example, we had comments "Feeling ashamed after stuttering during meetings" to describe emotional feelings after stuttering.
(2) The three researchers then met to read through and discuss the transcripts together. Through this discussion, they refined their initial comments into a set of agreed-upon codes, organized these codes into broader categories, and developed a preliminary coding scheme. For example, codes like "Stuttering is from imitation", "Stuttering can be cured"were grouped under the category "Misconception of stuttering".
(3) The first author then thoroughly reviewed the remaining transcripts multiple times, applying codes as comments and continuously refining the coding scheme through an iterative process. See coding scheme in appendix C.
(4) In subsequent research meetings, the team collaboratively identified key thematic insights emerging from the categorized codes and synthesized these insights for reporting.

4 Findings

4.1 Descriptive Statistics

Prior stuttered speech datasets often skew toward more fluent speech (e.g., reading tasks, isolated speech), leading to biases in ASR models. Our analysis of the StammerTalk dataset is motivated by the need to challenge these biases and reinforce the importance of collecting diverse samples. Our descriptive analysis of the dataset highlights its scale and data diversity, illustrating its unique quality to represent stuttered speech for ASR.

4.1.1 Scale and Scope. We measure the scale and scope of the StammerTalk dataset in terms of speakers, speech duration, stuttering events, and speech content. Key statistics for these aspects are provided in Table 1, along with existing datasets for comparison.

Speakers and Duration. A total duration of 50-hours speech data were included in StammerTalk dataset from 72 speakers. Excluding the two StammerTalk volunteers, most participants (64 out of 70) are from mainland, China. 34% (24) of the participants are female, much higher than the reported 20% or less among adults who stutter. Each participant contributed on average 33.0 minutes of conversational speech (*min*=17.2, *max*=49.9, SD=7.32), with an average of 17.8 minutes (*min*=7.14, *max*=34.93, SD=5.6) and an average of 15.23 minutes (*min*=6.45, *max*=27.6, SD=5.23) of voice command dictation. Many participants found speaking with another PWS both rare and pleasant [13, 32], thus spent more time on the conversations.

Stuttering Events. A total of 28,310 stuttering events were annotated in the StammerTalk dataset. Table 2 compares the frequency and distribution of annotated stuttering events in the StammerTalk dataset with existing stuttered speech datasets with stuttering event annotation, highlighting the quantity and diversity of stuttering events captured in the StammerTalk data. We also compute the *Average Stuttering Rate* by dividing the total count of stuttering events by the duration of the speech, and find that conversational speech in the StammerTalk dataset exhibits approximately 25% more stuttered utterances compared to the stuttering podcast (SEP-28k) and synthetic stuttered speech (LibriStutter).

Table 2 also shows *Event Type Distribution*, the percentage of each stuttering type among all annotated stuttering events. We note that a direct comparison between the Sep-28k and StammerTalk datasets may not provide the full picture, as StammerTalk's event annotation is performed at the character level, which offers greater granularity than the clip-level annotation in Sep-28k. However, we do observe a significant shift towards more word and phrases repetitions and less sound repetitions in the StammerTalk dataset, signaling potential phonological differences between stuttering in Chinese and in English. Meanwhile, we notice that SEP-28k dataset contains 40% more interjections than in StammerTalk Conversations, which could be attributed to different definitions of interjections in these two datasets: while SEP-28k considers any filler words - such as "um," "uh," and "you know" - as stuttering interjections, StammerTalk's annotation excludes natural interjections that blend into the speech flow.

Stuttering Transcription. The StammerTalk dataset contains both voice command dictation and unscripted conversations in Chinese. Excluding stuttering event annotations, the verbatim text transcription of StammerTalk dataset contains 425K Chinese characters (274K for conversations, 171K for voice command dictation).

To summarize, the StammerTalk dataset surpasses existing datasets in its duration, speakers, and stuttering frequency. It contains 20 times more transcribed speech data from people who stutter than what is available today (i.e. FluencyBank), and a multiplied number of speakers who stutter. Additionally, it provides both stuttering event annotations and verbatim transcriptions, enabling versatile applications across a wide range of technical domains. Unlike podcasts or audio books, the StammerTalk dataset contains unscripted conversations and voice command dictations that closely resembles real-world speech product use cases, such as meeting transcriptions and speech-operated devices.

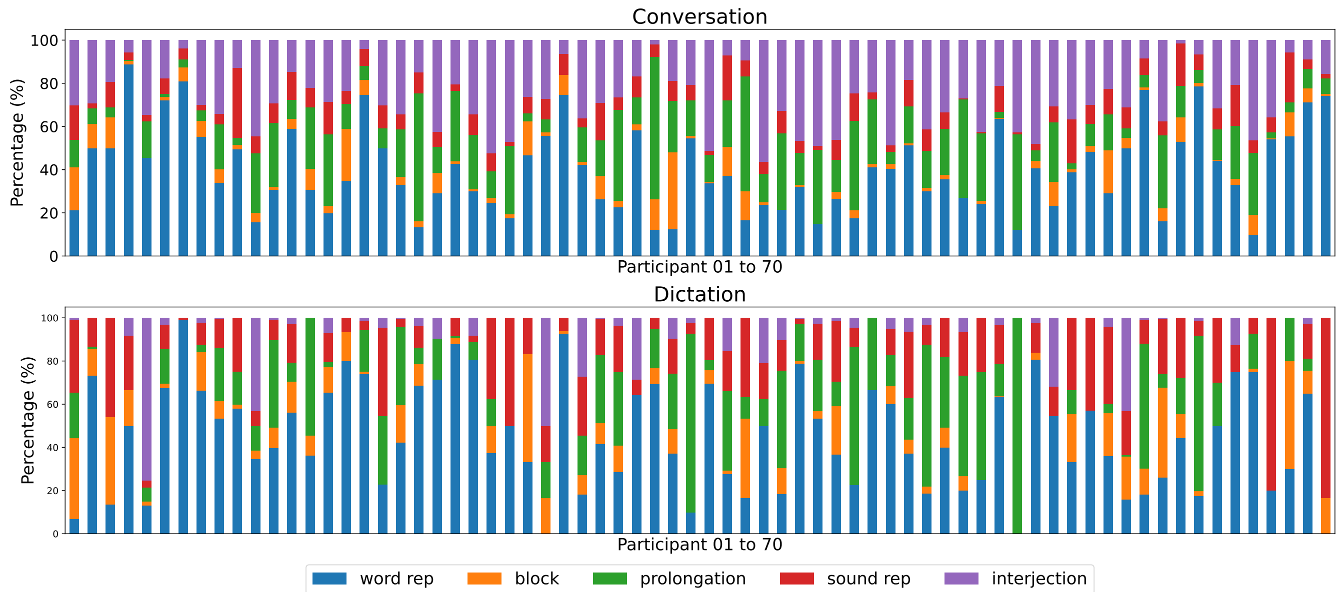


Figure 1: Breakdowns of five annotated stuttering events for 70 participants

Table 1: Dataset scale and scope as characterized by speech duration (Duration), the number and types of speakers (Speakers), whether it provides speech transcription (Transcription), types of speaking tasks (Tasks), and Language.

Dataset	Duration	Speakers	Transcription	Tasks	Language
FluencyBank* [45]	3.5 hrs	32 AWS	Yes	conversation, reading article	English
LibriStutter [27]	20 hrs	50 non-PWS	Yes**	audiobook	English
UCLASS* [21]	53 mins	25 CWS	Yes	conversation	English
SEP-28k [29]	23 hrs***	not reported	No	podcast	English
StammerTalk	**50 hrs**	**72 AWS**	**Yes (verbatim)**	**conversation, voice commands**	**Chinese**

* Limited to the transcribed portion of the dataset.
** Stuttered utterances are masked in the transcription as "STUTTER".
*** Split into 28K 3-second clips.
Abbreviations: AWS - adults who stutter; CWS - children who stutter.

4.1.2 *Speech Diversity.* Contrasting to previous stuttered speech datasets [26, 29], the StammerTalk dataset captures a wide spectrum of stuttering frequency and patterns across PWS in different scenarios, providing a much more comprehensive representation of the variability and heterogeneity of stuttered speech for speech AI.

Stuttering frequency. While all participants self-identified as PWS, their stuttering frequency varied. To quantify individual stuttering frequency, we calculate *disfluency rate*, as defined in [16, 28], by dividing the *total number of stuttering events* over the *total number of transcribed non-stuttering characters* for each speaker.

Table 2: Overall frequency and distribution of annotated stuttering events.

	Avg. Stuttering Rate (per minute)	Total Stuttering Events	Event Type Distribution				
			[]	**/b**	**/p**	**/r**	**/i**
LibriStutter [26]	12.5*	15,000*	20%	20%	20%	20%	20%
SEP-28k [29]	12.26	17,267	16%	19%	16%	14%	35%
StammerTalk: Conversation	15.83	19,674	42%	6%	18%	9%	25%
StammerTalk: Dictation	8.10	8,636	53%	8%	17%	16%	6%

* Stuttered utterances were synthetically generated.

Using the same thresholds as in previous work [16, 28], we categorize speakers into three groups based on their *disfluency rates*, corresponding to mild (0-5%), moderate (6-20%), and severe (over 20%) stuttering. We notice that, while the participants in general stutter more in Conversations (mean=9.2%) than in Command Dictation (mean=7.1%), the stuttering frequency varies more in Command Dictation (std=0.15) than in Conversation (std=0.08) (see Fig. 4 in Appendix A for visualization). As a result, the grouping of speakers varies across two tasks: for Conversation, 20, 44, and 6 speakers are categorized as mild, moderate, and severe stuttering, respectively, whereas for Dictation, the numbers are 46, 18, and 6.

The variation in *disfluency rates* across different tasks and speakers highlights the dynamic and situated nature of stuttering: its severity varies not just across individuals but also within the same individual. For some, reading is much easier than conversations; whereas for others, reading could be extremely challenging (*disfluency rate* as high as 100%).

Stuttering patterns. PWS often stutter differently: some speak with more repetitions, some frequently block, while some stutter covertly [9, 54]. Fig 1 shows the breakdowns of annotated stuttering events, for all 70 participants, highlighting the variation with their stuttering patterns. It also illustrates the change in stuttering patterns for the same speaker with different tasks: participants often have relatively more interjections in conversations, but show increased sound repetitions when dictating commands.

4.2 Benchmarking Results

The benchmarking results with Whisper model are shown in Fig. 2. We notice that it performs reasonably well with semantic transcriptions of mildly stuttered speech, achieving a CER of 11.71% for unscripted conversation (Fig. 2a). However, CER increases with stuttering severity, reaching 13.58% for the moderately and 24.39% for the severely stuttered speech. These CERs are both higher than the reported performance for the general population, namely, 12.8% on the Common Voice 15 dataset and 7.7% on the FLEURS dataset[5].

Comparing Fig. 2a with Fig. 2b, we find a sharp increase in deletion errors (DEL) when referencing on literal transcriptions. Further inspection of the results shows that Whisper has difficulties in generating disfluent literal transcriptions, often "smoothening" its transcriptions by removing repeated words or phrases. We provide examples of this behavior in Table 3 in Appendix B. As presented in Fig. 2c and Fig. 2d, we find that CERs, and in particular substitution errors (SUBs), are higher for Dictation tasks compared to Conversation, potentially due to its reliance on language model to "guess" correct transcription using the semantic context, which is more limited for voice commands. The wav2vec model, in contrast, performs 1.5 to 2 times worse than Whisper, and produces a lot more substitution mistakes even for natural conversations. Manual inspection finds that wav2vec model often substitutes a character with its homophones, undervaluing the semantic context. More detailed results for wav2vec model can be found in Appendix B.

[5]https://github.com/openai/whisper?tab=readme-ov-file

4.3 ASR Model Fine Tuning Results

The results of fine-tuning Whisper with the literal transcriptions of StammerTalk Conversation are presented in Fig 3. Literal transcription fine-tuning was chosen to assess ASR's ability to preserve disfluencies, whereas semantic transcription inherently removes them and is unsuitable for this evaluation. We observe substantial improvements in transcription accuracy across all severity levels of stuttering compared to the baseline model. For mildly stuttered speech, the fine-tuned model achieves a CER reduction from 16.34% (baseline model) to 5.8%, closing Whisper's performance gap between stuttered and fluent speech [44]. Similarly, for moderately and severely stuttered speech, the CER drops from 21.72% to 9.03% and from 49.24% to 20.46%, respectively. Our results illustrate the effectiveness of fine-tuning in improving ASR accuracy for stuttered speech across all severity levels.

We want to call out the significant reduction in deletion errors (DEL) after fine-tuning. As shown in Fig 3, the DEL rate drops from 26.56% to 2.29% for severely stuttered speech, and from 15.77% to 1.27% for moderately stuttered speech. Consistent with our benchmarking results, the baseline Whisper-large-v2 model often smoothens its output by omitting repeated words or phrases. This behavior, while generating more fluent transcript, leads to higher deletion error (DEL) rates when evaluated against verbatim transcriptions. We find the fine-tuned model more inclusive of speech disfluencies: it is more likely to preserve disfluencies rather than erasing them from the generated transcript.

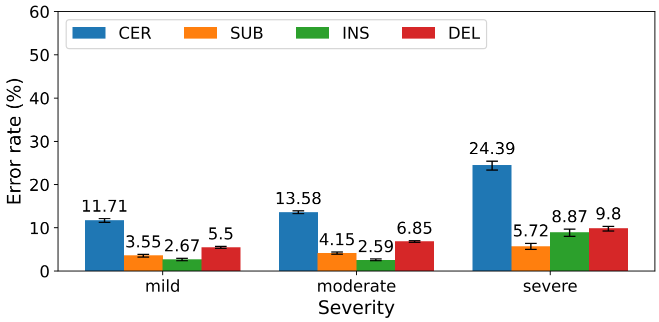
Overall, fine-tuning Whisper with the StammerTalk dataset helps the model better recognize and preserve speech disfluencies and significantly improves its transcription accuracy with stuttered speech. Our results demonstrate both the importance and the effectiveness of model fine-tuning with StammerTalk dataset in addressing ASR's fluency biases.

4.4 Qualitative Findings: Lived Experience of Stuttering in China

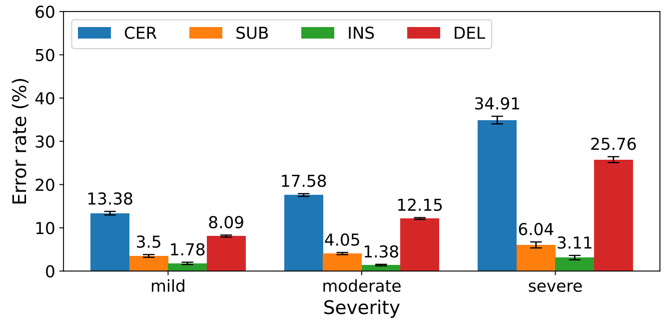
The lived experiences of stuttering in China – as captured in the StammerTalk dataset – provide critical cultural perspectives that are underrepresented in FAccT research. Much of the existing literature and frameworks for AI fairness have been shaped by Western contexts, which may overlook region-specific challenges such as prevalent stigma, systemic workplace discrimination, and limited access to disability accommodations in countries like China.

Our qualitative analysis find that, the conversations recorded in the dataset, although unscripted, are often centered on stuttering and the lived experiences of PWS. Stuttering is a socially isolating experience, [5, 56] and having a deeper conversation with another PWS was reported as among the key motivators for participating in the data collection [32]. We report major themes on the lived experience of stuttering in China. Participants' quotes were translated into English and lightly edited for readability.

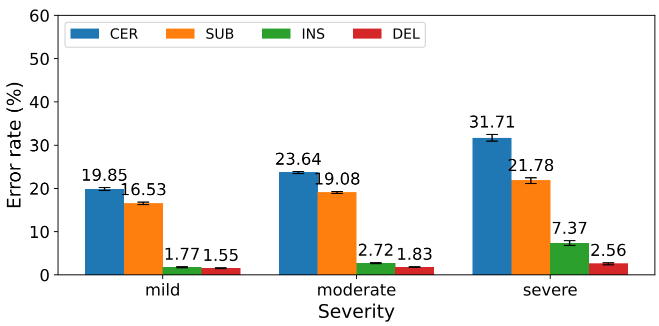
4.4.1 Prevalent social stigma and psychological impact. While stuttering is known to lead to negative emotional and cognitive reactions in PWS, the prevalence of social stigma towards stuttering, as its associated strong psychological impact, stands out in our data. All participants report experiencing some form of systematic



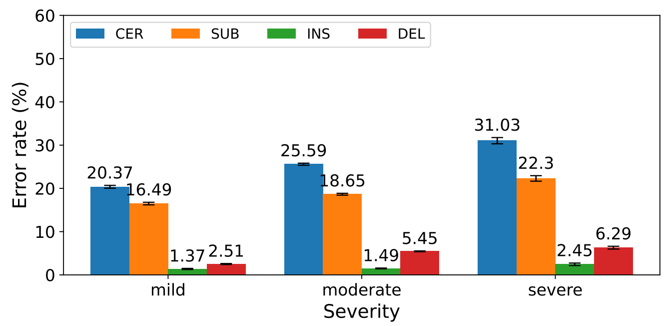
(a) Semantic transcription on Conversation



(b) Literal transcription on Conversation

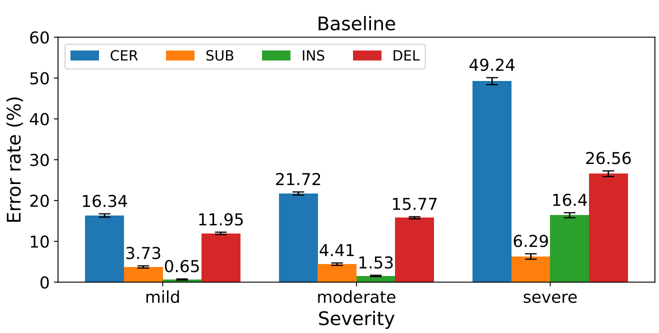


(c) Semantic transcription on Dictation

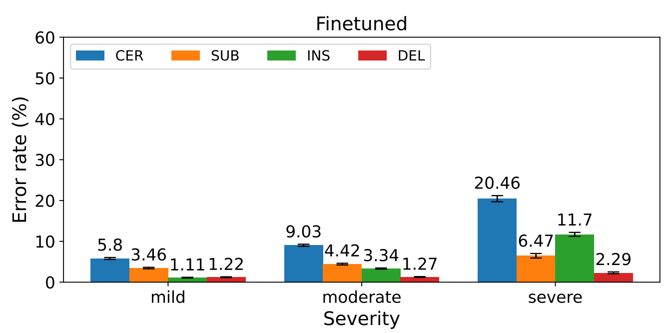


(d) Literal transcription on Dictation

Figure 2: Character error rate (CER), substitution (SUB), insertion (INS) and deletion (DEL) error rates for Whisper evaluated on StammerTalk Conversation and Dictation.



(a) Baseline Whisper



(b) Fine-tuned Whisper

Figure 3: Character error rate (CER), substitution (SUB), insertion (INS) and deletion (DEL) error rates for baseline and fine-tuned Whisper-large-v2 on StammerTalk Conversation.

discrimination and stigma towards stuttering at home, school, or workplace environments, which significantly impacted their social interactions and overall well-being. Nearly a third of the participants bring up the experience of being bullied by peers during childhood due to their stutter, which led to lasting psychological trauma, social anxiety, feelings of inferiority, avoidance behaviors, and depression. For example, P20 shared,

> *"I was afraid of speaking in front of many people. I was also scared of being called on by the teacher to answer questions in class. Sometimes, when I got very nervous,*

> *I couldn't speak at all. My classmates would laugh at me, making my childhood feel quite oppressive."*

The psychological toll of stuttering was so severe in P27's case that she developed self-harm to punish herself for speech difficulties during childhood,

> *"When I was a child, every time I got stuck or stumbled while speaking, or when I repeated words, I would severely punish myself. One way I punished myself as a child was by keeping my fingernails very long. If I couldn't speak properly, I would clench my fists, and my fingernails would dig into my skin."*

Several participants report a lack of understanding and acceptance of their stuttering by family members, which further undermines their psychological well-being. P69 reports her experience of depression and anxiety because of stuttering combined with misunderstandings from her parents.

> *"My parents are really very stubborn and not open-minded...I developed depression and anxiety because of my stuttering, but I had no way to tell them about it...and then I took a leave of absence from school. Gradually, my stuttering caused serious emotional problems, which also affected my physical health. I became extremely anxious and developed some psychosomatic symptoms."*

For a few participants who report neutral or positive experiences with stuttering, they often attribute their experiences to the acceptance attitude. As P26 reflects the acceptance mindset has minimalized the stuttering impact of stuttering on her life, *"What helped me the most was a shift in my mindset. I began to realize that although stuttering does have some impact on my daily life, as long as I handle it properly, its impact on me is actually very small. So now my attitude toward it is to accept it in a healthy way. "*

4.4.2 Workplace and professional discrimination. Although the stigma is prevalent, many participants highlight the tension between stuttering and the competitive employment environment in China today. As noted by P12,

> *"If I weren't working, I would feel that my stuttering isn't particularly severe and that I can adjust it in time. However, in today's society, there are very high demands on a person's overall abilities and competitiveness. If you want to be competent in certain positions, it's important to avoid having any weaknesses. Having a stutter does impact me personally and can also make it difficult to perform well in certain roles."*

Career choices are greatly affected by stuttering, pushing many participants into careers with minimal verbal interactions. P30 shares *"I feel quite anxious (about my stuttering) so ... I don't dare to choose a job that requires a lot of speaking. That's why I'm currently doing research work."* While occupational risks and labor discrimination for PWS were also reported in the US [19], they were much more overt and socially accepted in China, according to the StammerTalk dataset participants. In particular, having a stutter could disqualify someone for professional fields such as teaching and healthcare, as PWS are assumed to not able to meet the verbal communication demands. P27 noted she was discouraged from becoming a medical doctor because *"In the handbook for college applications, it stated that people who stutter are prohibited from applying for clinical majors"*.

While stuttering is highly dynamic and variable across individuals and situations, environmental stressors - such as time pressure and listeners' reaction - are reported to lead to more severe stuttering [54]. Overall, participants report fewer stuttering episodes when communicating with familiar individuals or in intimate settings but greater struggles during presentations, interviews, or interactions with strangers - situations common in the workplace. Stuttering is thus often viewed by employers, and internalized by our participants, as a failure and a sign of incompetence at work. P25 explained, *"in a competitive environment, I don't want to fall far behind my peers. If my supervisor knows that I have a stutter, they might not offer me important opportunities."*

Prevalent workplace discrimination of stuttering, combined with the lack of structural protection from labor unions or employment laws, drives our participants to spend significant efforts to "fix" their stutter or at least "pass" as fluent. For example, despite the discouragement from the college application handbook, P27 applied for medical school – while hiding her stutter – to "help others with similar hardship" and eventually became a physician. Although she was well appreciated by her patients for her skills and patience, she still felt constant pressure to speak fluently and would sometimes took sleeping pills or drank alcohol to reduce stuttering.

4.4.3 Coping mechanism. In response to the stigma and discrimination towards stuttering, participants develop various coping mechanisms, predominantly focused on concealing their stutter, fluency shaping techniques, and avoidance. Avoidance strategies include avoiding certain words or sounds and substituting challenging vocabulary, avoiding speaking situations and relationships. Some participants report avoiding communication as a strategy to manage the fear or reality of stuttering, which may lead to increased feelings of loneliness and isolation. For example, P51 intentionally refrains from joining conversations with colleagues,

> *"In the office, I rarely initiate conversations with others. Sometimes, even when I'm interested in what they're talking about, I avoid joining in just to prevent stuttering. To some extent, it feels a bit suppressive, but I can accept it because staying silent feels better than stuttering. In a way, it's like closing myself off."*

Although around one-third of the participants have sought "professional" help to manage stuttering, including attending online and in-person stuttering correction programs, seeing pediatrician, stomatologist (mostly during childhood), or other medical professionals, only a few have visited professionally trained speech language pathologists (SLPs). Some participants share that they have attended various speech programs claiming to cure stuttering but gained little or no improvement in speech fluency afterwards. For example, P28 expressed her disappointment,

> *"I have attended a stuttering correction class but the experience was very disappointing. Not only did it fail to make my speech more fluent, but it also increased my frustration with myself. The class promoted the idea that if you don't speak fluently, it's entirely your fault —*

> *you're not using the methods correctly or not practicing breathing properly."*

These programs are often expensive yet not effective, so some participants seek resources online, reading books about stuttering or joining stuttering support groups instead. Participants express that these resources are most helpful in improving their acceptance towards stuttering:*"I started participating in in-person stuttering support groups, and by hearing other people who stutter share their experiences, I was slowly able to accept my stuttering. Even if others outside couldn't accept it, I felt that my mindset had changed." (P16)*

In contrast to the documented benefits of self-disclosure [68], more than half of the participants mention they often avoid disclosing their stuttering. Participants fear that disclosing their stuttering could lead to misunderstandings or negative interpersonal and professional consequences. P4 describes the discomfort and the interpersonal risks of disclosing stuttering, *"I'm afraid that if I disclose my stuttering to my friends, they might leave me or dislike me."*

4.4.4 Misconceptions about stuttering.

We observed a general lack of scientific understanding about stuttering in our data, even within the stuttering community in China. Such deficit of knowledge perpetuates harmful stereotypes and increases social and self stigma. One common misconception is that many people think they acquired their stuttering from imitating stuttered speech during childhood. Similarly, some PWS also worry that their children might develop a stutter by imitating them, which causes them significant psychological stress, P47 expressed,

> *"I might unintentionally influence my child, because young children naturally imitate their parents. I feel that my stuttering not only affects me but could also impact my children's future, including their job interviews, career opportunities, and even their romantic relationships."*

Participants also report a prevalent view that verbal fluency reflects cognitive competence, which is often used to justify social exclusion, discrimination, and limited opportunities for people who stutter. P8 shared, *"Most people have a misunderstanding about stuttering: they assume that people who stutter also have low intelligence."*

Some participants report that their stuttering has been treated as a physical abnormality rather than a complex neurological and psychological condition. As P34 mentioned *"My parents thought that my speech issue was due to a physiological condition. So they took me to have surgery to shorten my tongue frenulum, but it didn't improve my stuttering."*

The lack of scientific understanding of the causes and nature of stuttering could lead to unrealistic expectations for PWS to speak fluently and harsh criticism towards them for not trying hard enough. As P56 shared,

> *"My parents would criticize me harshly about my speech if I didn't speak well. They would say that I must speak properly and that if I couldn't, it would be difficult for me to find a good job in the future...For as long as I can remember, whenever I didn't speak well, they would always criticize me."*

Despite widespread misconceptions, StammerTalk data collectors – both of whom resided outside China and received more comprehensive stuttering therapy and professional support – frequently shared information with participants on various aspects of stuttering, including its cause, techniques to improve fluency, and available resources for management. Thus, the data collection process also served as an educational opportunity for PWS to learn about stuttering and reflect on their personal experiences. For instance, P31 commended the interviewer for sharing the benefits of self-disclosure, *"I gained a lot from your sharing. I might take further steps to actively disclose my stuttering."*

4.4.5 Speech AI adoption and challenges.

Participants report utilizing a range of ASR products for specific use cases in their daily lives. For example, WeChat Voice Messages is commonly used for sending text messages via voice input and converting received voice messages into text. Xiaomi "Xiao Ai" serves purposes such as smart home controls and engaging in casual conversations. iFlytek is primarily used for speech-to-text conversion and daily transcription tasks. Car Navigation Tools enable participants to set destinations and issue navigation commands using voice input. Interestingly, one data collector reflected that some participants reported feeling more comfortable using ASR compared to speaking with real person as they believe ASR would neither judge their speeches nor react differently to their stuttering. Similarly, one participant use ASR to improve fluency and build confidence *"I use that app to practice my speech, such as for the Mandarin proficiency test. On one hand, I do this to desensitize myself, and on the other hand, I feel I need to live up to my job as a teacher." (P3)*

Despite the potential benefits, PWS face unique challenges with ASR products, including recognition errors, time-limited input difficulties, and heightened self-consciousness [28]. Despite these issues, ASR is widely used in China due to its advantages, such as simplifying Chinese typing and improving efficiency. However, usability barriers hinder PWS from leveraging these tools effectively, placing them at a disadvantage in technology use.

5 Discussion

5.1 Technical value of StammerTalk dataset in addressing ASR fluency biases

Representing the disability community adequately and authentically in AI data has been a prominent challenge in AI fairness and accessibility [40, 62]. This challenge is even more pronounced for stuttering, an "invisible" disability that is highly variable and situational. Created by a grassroots stuttering community for AI use, the StammerTalk dataset surpasses existing stuttered speech datasets in its scale, scope, and speech diversity, opening the door for a wide range of technical explorations and interventions for ASR biases.

Although prior research has established stark disparities in ASR model's performance with stuttered versus fluent speech [3, 28, 38], the unprecedented size and diversity of stuttered speech in the StammerTalk dataset will allow deeper understanding of ASR failures across different types of stuttering, stutterers, and speaking contexts. For example, the divergent stuttering patterns captured in unscripted conversation and voice command dictation tasks can inform ASR models about the importance of situational context in understanding stuttered speech, using features exacted from the StammerTalk dataset as a starting point. Also, as previous study on

ASR performance with aphasia speech found increased frequency of model hallucination over utterances containing long pauses [24] - a symptom shared by both aphasia and stuttering, targeted analysis on ASR results for different types of stuttering utterances could lead to new insights on common ASR mistakes as well as potential mitigation strategies. Inspecting the types of mistakes made by different ASR models also shed light on the underlying mechanisms within the otherwise blackboxed models that drive their discriminatory behaviors. For example, our results suggest the reliance on language model and semantic context by Whisper model constrains its ability to recognize and transcribe stuttered utterances, while the over-indexing of acoustic features by wav2vec model could lead to increased homophone errors in its transcript.

Furthermore, the rigorous verbatim transcription, annotated with specific stuttering events, enables ASR systems to recognize and transcribe stuttered utterances as they are, which not only provides a more accurate transcription but also normalizes stuttering in human communications - an attitude clinically proven to benefit people who stutter in the long term [50]. While conventional approach for ASR evaluation routinely remove disfluencies - such as the filler words - from both ground truth and model generated transcript to make it easier to align and compare the reference with the inference [25], the verbatim transcriptions provided in the StammerTalk dataset allows us to better measure and address fluency biases in ASR models. For example, our audit of the Whisper model using the literal transcript reveals its tendency to artificially "smooth out" stuttered speech in the transcriptions and exposes its embedded ableist biases against speech disfluencies. We also show that such biases can be partially addressed by fine tuning ASR models using the StammerTalk dataset. Compared to the off-the-shelf Whisper model, the fine-tuned model produces more accurate transcriptions of stuttered speech transcriptions with consistent reductions in general and all sub-types of mistakes.

5.2 Social and educational values of collective stuttered voices in Chinese

The StammerTalk dataset also offers unique social and educational values. While speech interfaces and ASR-mediated interactions have been increasingly adopted for convenience, accessibility, and cost-efficiency, the lack of inclusion of users with diverse speech patterns in the research and development of these systems could lead to new accessibility barriers and psychological harms [3, 28, 60]. The StammerTalk dataset can inform HCI researchers about the diversity and variability of speech input, contributing new user personas and design considerations for inclusive speech technologies.

As the only stuttered speech corpus in a non-Western language to our knowledge, the StammerTalk dataset also fills in an important language gap for stuttered speech and opens doors to quantify the linguistic and cultural differences in stuttering between Chinese and other, mostly Eurocentric, languages. Besides, compared to plain text transcripts, the audio format of the personal experiences told by PWS in China creates an intimate channel for self advocacy and empathy building. Listening to personal stories—particularly those highlighting systemic discrimination and psychological struggles-told in stuttered voices —can provide speech AI researchers and designers with a deeper understanding of the goals and needs of PWS, as well as greater awareness of their own fluency biases. On the other hand, for many participants, it was the first time they were able to speak about their stutter and have their stuttered voices heard by the public. As one of the first public discourses about stuttering experiences in China, the dataset provides a platform for collective actions, claiming the much needed space for stuttering in Chinese society.

The dataset also enhances understanding of the social context around stuttering in China, which is essential for creating socially aware and inclusive products. For instance, while products aimed at masking of stuttering have been increasingly rejected by the stuttering community in the US [31, 55, 64], such solutions may appeal to participants in China, where stuttering carries significant personal and professional risks but lacks support infrastructure. Building socially aware products could introduce an ethical dilemma between addressing pressing harms and maintaining fundamental values such as justice and authenticity [52].

5.3 Limitation and future work

Despite the unprecedented scale of the StammerTalk dataset, our work still has several limitations. First, our focus on Chinese stuttered speech restricts its applicability to other languages. Future work could replicate the StammerTalk data collection model across additional languages and dialects, further expanding the diversity and scale of stuttered speech datasets. Additionally, while prior research demonstrates the promise of fine-tuning general ASR models with small amounts of stuttered speech [28, 38], future work could explore how the StammerTalk dataset can advance this direction further. Second, our process focuses on the curation of the dataset but managing the StammerTalk dataset requires significant effort. Although the community intends to open-source their data for scientific and technological advancements, future work should help the community navigate complex legal and technical systems to identify suitable infrastructure for collective ownership, personal data protection, and cross-border data regulations. Finally, while the StammerTalk dataset highlights the trade-offs between technical and social values, future work should explore strategies to achieve balance. For instance, there may be tensions between selecting topics that resonate deeply with the community and ensuring the diversity of vocabulary required for technical advancements in data collection.

6 Conclusion

This work examines the technical and social value of a community-led stuttered speech dataset in Mandarin Chinese. Through quantitative and qualitative analysis, as well as benchmarking and fine-tuning ASR models using the dataset, we demonstrate how grassroots, disability-centered data efforts can both expose and begin to rectify the fluency bias in speech AI systems. Beyond improving ASR performance for people who stutter, our approach fosters self-advocacy, community-building, and a more inclusive vision of AI development—one grounded in the lived experiences and leadership of those most affected.

7 Ethics Statement

Ethical considerations statement: This research was conducted with a strong commitment to ethical standards and participant safety, privacy, and autonomy. Recruitment messaging clearly communicated the purpose of the study, the nature of participation, potential risks, and data usage details. Participants were explicitly informed that their speech data would be collected, analyzed, and shared exclusively for non-commercial research purposes aimed at improving ASR technologies for people who stutter. Each participant provided informed consent through a comprehensive consent form, clearly specifying permitted uses of their data, including limitations on data sharing, storage duration, and security protocols. The consent documents and data management protocols were reviewed and approved by pro bono legal experts specializing in data privacy and protection. Data security measures included anonymization of participants, encryption of digital records, and secure storage with restricted access limited strictly to authorized researchers.

Positionality statement: We acknowledge that our personal backgrounds and identities shape how we engage with communities and interpret our findings. All authors are Mandarin-speaking and currently reside outside of mainland China. Our research team includes both PWS and non-stuttering allies. The third, fourth, and fifth authors identify as PWS and contributed essential insider perspectives to the research. The third and fourth authors are core members of the StammerTalk community and led the data collection effort. While this team composition enabled us to build close personal and professional relationships with the stuttering community, we also recognize that our socioeconomic and educational backgrounds have afforded us certain privileges relative to many community members we engaged with—privileges that may have influenced participants' trust and willingness to share their experiences. We remain mindful of these dynamics throughout our research process and strive to amplify community voices with care.

Acknowledgments

We extend our heartfelt gratitude to the StammerTalk community for entrusting us with their data. We also thank the Authentic User Experience Lab at University of California, Santa Cruz, especially professor Norman Makoto Su, for stimulating discussions and collaboration on the future development of this work. This work is supported by NSF Award #2427710 and the Patrick J. McGovern Foundation.

References

[1] Adriana Alvarado Garcia and Christopher A. Le Dantec. 2018. Quotidian Report: Grassroots Data Practices to Address Public Safety. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 17 (nov 2018), 18 pages. https://doi.org/10.1145/3274286

[2] Rosana Ardila, Megan Branson, Kelly Davis, Michael Kohler, Josh Meyer, Michael Henretty, Reuben Morais, Lindsay Saunders, Francis Tyers, and Gregor Weber. 2020. Common Voice: A Massively-Multilingual Speech Corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (Eds.). European Language Resources Association, Marseille, France, 4218–4222. https://aclanthology.org/2020.lrec-1.520

[3] Anna Bleakley, Daniel Rough, Abi Roper, Stephen Lindsay, Martin Porcheron, Minha Lee, Stuart Alan Nicholson, Benjamin R Cowan, and Leigh Clark. 2022. Exploring Smart Speaker User Experience for People Who Stammer. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–10.

[4] O. Bloodstein, N.B. Ratner, and S.B. Brundage. 2021. *A Handbook on Stuttering, Seventh Edition*. Plural Publishing, Incorporated. https://books.google.com/books?id=Abw0EAAAQBAJ

[5] Elaine Blumgart, Yvonne Tran, and Ashley Craig. 2010. Social anxiety disorder in adults who stutter. *Depress. Anxiety* 27, 7 (July 2010), 687–692. https://doi.org/10.1002/da.20657

[6] John Van Borsel, Marie Brepoels, and Janne De Coene. 2011. Stuttering, attractiveness, and romantic relationships: The perception of adolescents and young adults. *Journal of Fluency Disorders* 36, 1 (2011), 41–50.

[7] Michael P. Boyle. 2018. Enacted stigma and felt stigma experienced by adults who stutter. *Journal of Communication Disorders* 73 (2018), 50–61. https://doi.org/10.1016/j.jcomdis.2018.03.004

[8] Geoffrey A. Coalson, Alexus Crawford, Shanley B. Treleaven, Courtney T. Byrd, Lauren Davis, Lillian Dang, Jillian Edgerly, and Alison Turk. 2022. Microaggression and the adult stuttering experience. *Journal of Communication Disorders* 95 (2022), 106180. https://doi.org/10.1016/j.jcomdis.2021.106180

[9] Christopher Constantino, Patrick Campbell, and Sam Simpson. 2022. Stuttering and the social model. *Journal of Communication Disorders* 96 (2022), 106200. https://doi.org/10.1016/j.jcomdis.2022.106200

[10] Christopher Dominick Constantino. 2023. Fostering Positive Stuttering Identities Using Stutter-Affirming Therapy. *Language, Speech, and Hearing Services in Schools* 54, 1 (2023), 42–62. https://doi.org/10.1044/2022_LSHSS-22-00038 arXiv:https://pubs.asha.org/doi/pdf/10.1044/2022_LSHSS-22-00038

[11] Eric Corbett, Emily Denton, and Sheena Erete. 2023. Power and Public Participation in AI. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23)*. Association for Computing Machinery, New York, NY, USA, Article 8, 13 pages. https://doi.org/10.1145/3617694.3623228

[12] S. Davis, P. Howell, and F. Cooke. 2002. Sociodynamic relationships between children who stutter and their non-stuttering classmates. *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 43, 7 (2002), 939–947.

[13] Xianghua Ding, Patrick C Shih, and Ning Gu. 2017. Socially embedded work: A study of wheelchair users performing online crowd work in china. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 642–654.

[14] Kelechi Ezema, Chelsea Chandler, Rosy Southwell, Niranjan Cholendiran, and Sidney D'Mello. 2025. "It feels like we're not meeting the criteria": Examining and Mitigating the Cascading Effects of Bias in Automatic Speech Recognition in Spoken Language Interfaces.. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 109, 13 pages. https://doi.org/10.1145/3706598.3714059

[15] Vinitha Gadiraju, Shaun Kane, Sunipa Dev, Alex Taylor, Ding Wang, Emily Denton, and Robin Brewer. 2023. "I Wouldn't Say Offensive but...": Disability-Centered Perspectives on Large Language Models. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 205–216. https://doi.org/10.1145/3593013.3593989

[16] Andrews Gavin and MARY HARRIS. 1964. The Syndrome of Stuttering'. *Clinics in developmental Medicine* 17 (1964).

[17] Abraham Glasser, Kesavan Kushalnagar, and Raja Kushalnagar. 2017. Deaf, Hard of Hearing, and Hearing Perspectives on Using Automatic Speech Recognition in Conversation. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) *(ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 427–432. https://doi.org/10.1145/3132525.3134781

[18] Rong Gong, Hongfei Xue, Lezhi Wang, Xin Xu, Qisheng Li, Lei Xie, Hui Bu, Shaomei Wu, Jiaming Zhou, Yong Qin, et al. 2024. As-70: A mandarin stuttered speech dataset for automatic speech recognition and stuttering event detection. *arXiv preprint arXiv:2406.07256* (2024).

[19] Gerlach H, Totty E, Subramanian A, and Zebrowski P. 2018. Stuttering and Labor Market Outcomes in the United States. *J Speech Lang Hear Res.* 61, 7 (2018), 1649–1663. https://doi.org/10.1044/2018_JSLHR-S-17-0353 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6195060/.

[20] Aimi Hamraie and Kelly Fritsch. 2019. Crip technoscience manifesto. *Catalyst: Feminism, theory, technoscience* 5, 1 (2019), 1–33.

[21] Peter Howell, Stephen Davis, and Jon Bartrip. 2009. The university college london archive of stuttered speech (uclass). (2009).

[22] Ben Hutchinson, Andrew Smart, Alex Hanna, Remi Denton, Christina Greer, Oddur Kjartansson, Parker Barnes, and Margaret Mitchell. 2021. Towards Accountability for Machine Learning Datasets: Practices from Software Engineering and Infrastructure. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 560–575. https://doi.org/10.1145/3442188.3445918

[23] Seung-Eun Kim, Bronya R. Chernyak, Olga Seleznova, Joseph Keshet, Matthew Goldrick, and Ann R. Bradlow. 2024. Automatic recognition of second language speech-in-noise. *JASA Express Letters* 4, 2 (02 2024), 025204.

https://doi.org/10.1121/10.0024877 arXiv:https://pubs.aip.org/asa/jel/article-pdf/doi/10.1121/10.0024877/19669029/025204_1_10.0024877.pdf
- [24] Allison Koenecke, Anna Seo Gyeong Choi, Katelyn X. Mei, Hilke Schellmann, and Mona Sloane. 2024. Careless Whisper: Speech-to-Text Hallucination Harms. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Rio de Janeiro, Brazil) *(FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 1672–1681. https://doi.org/10.1145/3630106.3658996
- [25] 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 (2020), 7684–7689.
- [26] Tedd Kourkounakis, Amirhossein Hajavi, and Ali Etemad. 2021. FluentNet: End-to-End Detection of Stuttered Speech Disfluencies With Deep Learning. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.* 29 (sep 2021), 2986–2999. https://doi.org/10.1109/TASLP.2021.3110146
- [27] Tedd Kourkounakis, Amirhossein Hajavi, and Ali Etemad. 2021. Fluentnet: End-to-end detection of stuttered speech disfluencies with deep learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29 (2021), 2986–2999.
- [28] Colin Lea, Zifang Huang, Jaya Narain, Lauren Tooley, Dianna Yee, Dung Tien Tran, Panayiotis Georgiou, Jeffrey P Bigham, and Leah Findlater. 2023. From User Perceptions to Technical Improvement: Enabling People Who Stutter to Better Use Speech Recognition. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 361, 16 pages. https://doi.org/10.1145/3544548.3581224
- [29] Colin Lea, Vikramjit Mitra, Aparna Joshi, Sachin Kajarekar, and Jeffrey Bigham. 2021. Sep-28k: A Dataset for Stuttering Event Detection from Podcasts with People Who Stutter. https://arxiv.org/pdf/2102.12394.pdf
- [30] David Leslie, Michael Katell, Mhairi Aitken, Jatinder Singh, Morgan Briggs, Rosamund Powell, Cami Rincon, Antonella Perini, and Smera Jayadeva. 2022. Data justice in practice: A guide for impacted communities. *SSRN Electron. J.* (April 2022). https://doi.org/10.2139/ssrn.4080046
- [31] Jingjin Li, Shaomei Wu, and Gilly Leshed. 2024. Re-envisioning Remote Meetings: Co-designing Inclusive and Empowering Videoconferencing with People Who Stutter. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) *(DIS '24)*. Association for Computing Machinery, New York, NY, USA, 1926–1941. https://doi.org/10.1145/3643834.3661533
- [32] Qisheng Li and Shaomei Wu. 2024. "I Want to Publicize My Stutter": Community-led Collection and Curation of Chinese Stuttered Speech Data. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2, Article 475 (Nov. 2024), 27 pages. https://doi.org/10.1145/3687014
- [33] Qisheng Li and Shaomei Wu. 2024. Towards Fair and Inclusive Speech Recognition for Stuttering: Community-led Chinese Stuttered Speech Dataset Creation and Benchmarking. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 331, 9 pages. https://doi.org/10.1145/3613905.3650950
- [34] Yan Ma, Judith D. Oxley, J. Scott Yaruss, and John A. Tetnowski. 2023. Stuttering experience of people in China: A cross-cultural perspective. *Journal of Fluency Disorders* 77 (2023), 105994. https://doi.org/10.1016/j.jfludis.2023.105994
- [35] Bob MacDonald, Pan-Pan Jiang, Julie Cattiau, Rus Heywood, Richard Cave, Katie Seaver, Marilyn Ladewig, Jimmy Tobin, Michael Brenner, Philip Q Nelson, Jordan R. Green, and Katrin Tomanek. 2021. Disordered Speech Data Collection: Lessons Learned at 1 Million Utterances from Project Euphonia.
- [36] Valentin Mendelev, Tina Raissi, Guglielmo Camporese, and Manuel Giollo. 2021. Improved robustness to disfluencies in rnn-transducer based speech recognition. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6878–6882.
- [37] Meta. 2023. Speech Fairness Dataset. https://ai.meta.com/datasets/speech-fairness-dataset/.
- [38] Vikramjit Mitra, Zifang Huang, Colin Lea, Lauren Tooley, Panayiotis Georgiou, Sachin Kajarekar, and Jefferey Bigham. 2021. Analysis and Tuning of a Voice Assistant System for Dysfluent Speech. https://arxiv.org/pdf/2106.11759.pdf
- [39] National Institute on Deafness and Other Communication Disorders. 2016. NIDCD Factsheet: Stuttering. https://www.nidcd.nih.gov/health/stuttering. Retrieved on April 19th, 2022..
- [40] Joon Sung Park, Danielle Bragg, Ece Kamar, and Meredith Ringel Morris. 2021. Designing an Online Infrastructure for Collecting AI Data From People With Disabilities. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 52–63. https://doi.org/10.1145/3442188.3445870
- [41] Reema Patel. 2021. Participatory data stewardship. (Sept. 2021). https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/
- [42] Alisha Pradhan, Amanda Lazar, and Leah Findlater. 2020. Use of intelligent voice assistants by older adults with low technology use. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 4 (2020), 1–27.
- [43] Rida Qadri, Renee Shelby, Cynthia L. Bennett, and Remi Denton. 2023. AI's Regimes of Representation: A Community-centered Study of Text-to-Image Models in South Asia. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 506–517. https://doi.org/10.1145/3593013.3594016
- [44] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *Proceedings of the 40th International Conference on Machine Learning* (Honolulu, Hawaii, USA) *(ICML'23)*. JMLR.org, Article 1182, 27 pages.
- [45] Nan Bernstein Ratner and Brian MacWhinney. 2018. Fluency Bank: A new resource for fluency research and practice. *Journal of fluency disorders* 56 (2018), 69–80.
- [46] Anouk Ruhaak, Greg Bloom, Angie Raymond, Willa Tavernier, Divya Siddarth, Gary Motz, and Melanie Dulong de Rosnay. 2021. A Practical Framework for Applying Ostrom's Principles to Data Commons Governance. https://www.mozillafoundation.org/en/blog/a-practical-framework-for-applying-ostroms-principles-to-data-commons-governance/. Accessed: 2025-5-12.
- [47] Johnny Saldaña. 2021. *The coding manual for qualitative researchers.* sage.
- [48] Olabanji Shonibare, Xiaosu Tong, and Venkatesh Ravichandran. 2022. Enhancing ASR for Stuttered Speech with Limited Data Using Detect and Pass. arXiv:2202.05396 [eess.AS]
- [49] Judy Singer. 1999. Why can't you be normal for once in your life? From a problem with no name to the emergence of a new category of difference. *Disability discourse* (1999), 59–67.
- [50] Vivian Sisskin. 2023. Disfluency-Affirming Therapy for Young People Who Stutter: Unpacking Ableism in the Therapy Room. *Language, Speech, and Hearing Services in Schools* 54, 1 (2023), 114–119.
- [51] Joshua St. Pierre. 2012. The Construction of the Disabled Speaker: Locating Stuttering in Disability Studies. *Canadian Journal of Disability Studies* 1, 3 (Aug. 2012), 1–21. https://doi.org/10.15353/cjds.v1i3.54
- [52] Sharifa Sultana, François Guimbretière, Phoebe Sengers, and Nicola Dell. 2018. Design Within a Patriarchal Society: Opportunities and Challenges in Designing for Rural Women in Bangladesh. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174110
- [53] Seth E. Tichenor and J. Scott Yaruss. 2019. Stuttering as Defined by Adults Who Stutter. *Journal of Speech, Language, and Hearing Research* 62 (2019), 4356–4369. https://doi.org/10.1044/2019_JSLHR-19-00137
- [54] Seth E Tichenor and J Scott Yaruss. 2021. Variability of stuttering: Behavior and impact. *American Journal of Speech-Language Pathology* 30, 1 (2021), 75–88.
- [55] Space to Stutter. 2024. Open Letter to Samsung. https://www.spacetostutter.org/impulse. Accessed on 1/23/2025..
- [56] Seda Türkili, Serkan Türkili, and Zeynep Feryal Aydın. 2022. Mental well-being and related factors in individuals with stuttering. *Heliyon* 8, 9 (Sept. 2022), e10446. https://doi.org/10.1016/j.heliyon.2022.e10446
- [57] UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN. 2023. Speech Accessibility Project. https://speechaccessibilityproject.beckman.illinois.edu/. Accessed on 1/21/2024..
- [58] Ravichander Vipperla, Steve Renals, and Joe Frankel. 2010. Ageing Voices: The Effect of Changes in Voice Parameters on ASR Performance. *EURASIP Journal on Audio, Speech, and Music Processing* (2010). https://asmp-eurasipjournals.springeropen.com/articles/10.1155/2010/525783
- [59] Maggie Walter and Michele Suina. 2019. Indigenous data, indigenous methodologies and indigenous data sovereignty. *Int. J. Soc. Res. Methodol.* 22, 3 (May 2019), 233–243. https://doi.org/10.1080/13645579.2018.1531228
- [60] Kimi Wenzel, Nitya Devireddy, Cam Davison, and Geoff Kaufman. 2023. Can Voice Assistants Be Microaggressors? Cross-Race Psychological Responses to Failures of Automatic Speech Recognition. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 109, 14 pages. https://doi.org/10.1145/3544548.3581357
- [61] Kimi Wenzel and Geoff Kaufman. 2024. Designing for Harm Reduction: Communication Repair for Multicultural Users' Voice Interactions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.
- [62] Meredith Whittaker, Cynthia L. Bennett Meryl Alper, Sara Hendren, Elizabeth Kaziunas, Mara Mills, Meredith Ringel Morris, Joy Lisi Rankin, Emily Rogers, Marcel Salas, and Sarah Myers West. 2019. Disability, Bias & AI Report. *AI Now Institute* (20 11 2019).
- [63] Janis Wong. 2023. Data Practices and Data Stewardship. *Interactions* 30, 3 (may 2023), 60–63. https://doi.org/10.1145/3589133
- [64] Shaomei Wu. 2023. "The World is Designed for Fluent People": Benefits and Challenges of Videoconferencing Technologies for People Who Stutter. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [65] Shaomei Wu, Jingjin Li, and Gilly Leshed. 2024. Finding My Voice over Zoom: An Autoethnography of Videoconferencing Experience for a Person Who Stutters.

In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 916, 16 pages. https://doi.org/10.1145/3613904.3642746

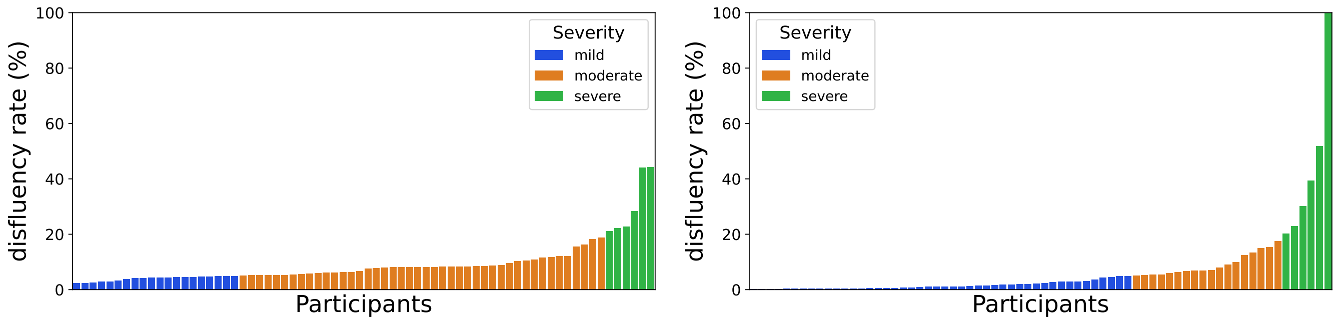
[66] J Scott Yaruss and Robert W Quesal. 2006. Overall Assessment of the Speaker's Experience of Stuttering (OASES): Documenting multiple outcomes in stuttering treatment. *Journal of fluency disorders* 31, 2 (2006), 90–115.

[67] yeyupiaoling. [n. d.]. Whisper-Finetune. https://github.com/yeyupiaoling/Whisper-Finetune. Accessed: 2025-01-22.

[68] Megan M Young, Courtney T Byrd, Rodney Gabel, and Andrew Z White. 2022. Self-disclosure experiences of adults who stutter: An interpretative phenomenological analysis. *Am. J. Speech. Lang. Pathol.* 31, 5 (Sept. 2022), 2045–2060. https://doi.org/10.1044/2022_AJSLP-22-00048

[69] Xin Zhang, Iván Vallés-Pérez, Andreas Stolcke, Chengzhu Yu, Jasha Droppo, Olabanji Shonibare, Roberto Barra-Chicote, and Venkatesh Ravichandran. 2022. Stutter-TTS: Controlled synthesis and improved recognition of stuttered speech. *arXiv preprint arXiv:2211.09731* (2022).

A Additional Descriptive Statistics of StammerTalk Dataset

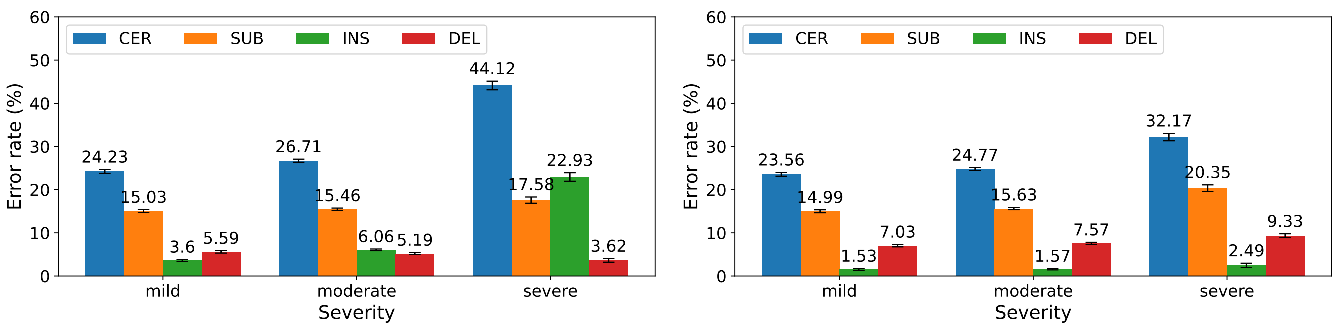


(a) Individual disfluency rates in StammerTalk Conversation

(b) Individual disfluency rates in StammerTalk Dictation

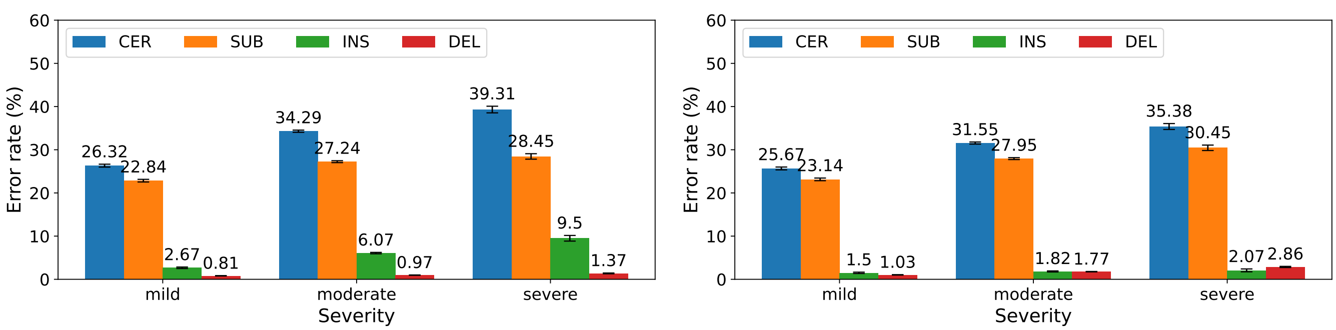
Figure 4: Disfluency rates of all 70 participants in Conversation and Dictation tasks, sorted from low to high, and categorized into mild (0-5%), moderate (5-20%), and severe (20%+) stuttering groups.

B Additional Benchmarking Results



(a) Semantic transcription on Conversation

(b) Literal transcription on Conversation



(c) Semantic transcription on Dictation

(d) Literal transcription on Dictation

Figure 5: Character error rate (CER), substitution (SUB), insertion (INS) and deletion (DEL) error rates for fine-tuned wav2vec 2.0 evaluated on StammerTalk Conversation and Dictation. Note that wav2vec demonstrates comparatively better performance when evaluated against literal transcriptions, as opposed to semantic transcriptions used as ground truth.

Table 3: Whisper model "smooths" the transcriptions by removing words with low semantic value, as indicated by the underlined characters.

Reference	Whisper model output
就在那个继续深造的也有	就在继续深造的也有
嗯我觉得深圳他到处他都是花钱的地方就是吃喝玩乐他肯定是	我觉得深圳到处都是花田的地方吃喝玩乐肯定是

Table 4: Three examples of utterances from a severe PWS, characterized by frequent word repetitions. The wav2vec model produced homophone substitutions, as indicated by the underlined characters.

Annotation*	wav2vec model output
当[当当当]时我上/b[上]去的时候	当当档单舍瓦上上去的时候
我/b现[现现现]就[就]挺自/r卑[卑]的	我先线先千就就点自杯给的
呃/i进[进]行那个自[自]我介[介]绍，呃/i	而仅仅进行了个自自我界介绍和儿

* Stuttering events markups: **[ ]** - word repetition, **/r** - sound repetition, **/b** - blocks, **/p** - prolongation, **/i** - interjection.

C Coding Scheme

Category	Subcategory	Codes
1. Stuttering Experiences	1.1 Personal Experiences	Fear, anxiety; Tension, feeling out of control; Emotional responses after stuttering (e.g., embarrassment, frustration); Mental health problems (e.g., depression, anxiety disorder)
	1.2 Social Experiences	Negative reactions from others (e.g., mockery, making fun of, distrust); Discrimination or prejudice; Positive experience: Supportive social interactions
	1.3 Misconceptions of Stuttering	Stuttering is from imitation; Stuttering can be cured; Stuttering is because of nervousness; "If you speak slowly, you wouldn't stutter"
	1.4 Attitudes Toward Stuttering	Acceptance; Not accepting and wanting to cure
2. Coping Mechanisms	2.1 Strategies for Managing Stuttering	Avoidance behaviors (e.g., avoiding certain words, speaking situations, professions, relationships); Speech therapy techniques (e.g., fluency shaping, deep breath); Speech therapy program; Existing resources (e.g., books, online); Seeking help from SLP
	2.2 Emotional Coping	Internal dialogues (e.g., self-reassurance, self-acceptance); Seeking support from family and friends; Seeking support from online communities; Seeking support from mental health professionals
3. Self-Disclosure	3.1 Levels of Disclosure	Public disclosure; Disclosure to family and close friends; Disclosure to co-workers or in professional settings; No explicit self-disclosure (assume others are aware)
	3.2 Barriers to Self-Disclosure	Fear of stigma; Previous negative experiences (e.g., dismissal, denial of stuttering)
	3.3 Challenges After Disclosure	Misguided advice; Negative reactions (e.g., mockery, denial); Social and professional consequences
4. Occupational Distribution and Challenges	4.1 Occupation	Coding by type (e.g., teacher, doctor, customer service)
	4.2 Work-Related Challenges	Communication-intensive roles (e.g., teaching, public speaking); Impact on career progression or job opportunities
	4.3 Urban vs. Cosmopolitan vs. Overseas Experiences	Coding by area if any
5. ASR Products and Usage	5.1 Product Usage	Frequency of use (e.g., daily, occasionally); Frequent use at home or in solo settings; Avoidance in public or group settings; Purpose: Speech-to-text, Control smart home devices, Accent reduction, Fluency shaping
	5.2 Challenges	Recognition issues: inaccurate recognition of the stuttered speech; Premature cancellation of input if there are delays; Error-prone results requiring manual corrections
	5.3 Social Dynamics	Hesitation or avoidance of ASR in front of others; Preference for using ASR in private
6. Dynamic Nature of Stuttering	6.1 Variability Over Time	Changes in stuttering severity across life stages; Impact of specific situations (e.g., stress, public speaking)
	6.2 Contextual Factors	Variations in stuttering based on the audience or setting; External triggers or mitigators (e.g., pressure, comfort levels)

Table 5: Coding Scheme