Making Room for Speech Diversity: A 50 Year Retrospective of Speech Science and Technology through a Neurodivergent Lens

Rebecca Lietz [1], *Jingjin Li* [2], *Peiyao Liu* [1], *Jennifer Chien* [3], *Norman Makoto Su* [1],
Shaomei Wu [2]

[1] Computational Media, University of California, Santa Cruz, USA
[2] AImpower.org, USA
[3] Stanford University, USA

`rlietz@ucsc.edu, jingjin@aimpower.org, pliu62@ucsc.edu, drchien@stanford.edu, normsu@ucsc.edu, shaomei@aimpower.org`

Abstract

Interest in diversity and inclusion in speech science and technology has grown in recent years, reflected in an increasing number of Interspeech publications on "atypical" speech from disability communities. We take this momentum as an opportunity to examine how diversity has been framed in Interspeech's research discourse. Using neurodivergent speech as a case study, we conduct a scoping review of relevant Interspeech and ICASSP papers from 1976 to 2025. Our content and discourse analysis identifies three recurring issues: (1) medicalized and interventionist framings position neurodivergent traits as deficits to be corrected; (2) minimal inclusion of neurodivergent individuals in the research process, distancing stakeholders and limiting impact; and (3) ableist language that contributes to the marginalization in speech technology research. We conclude by proposing pathways through which Interspeech can achieve its stated goal of celebrating "***all** voices*" in its research and practice.

Index Terms: neurodivergent speech, disability, ableism, neurodivergence, models of disability, inclusivity, discourse analysis

1. Introduction

Human speech is highly variable. The human brain demonstrates a remarkable adaptability to create a wide range of voices, accents, and speaking styles [1]. Speech technologies, in contrast, cater to a narrow set of "typical" speech patterns, underperforming for individuals with speech diversities such as stuttering [2, 3], deaf speech [4], second language speech [5, 6], as well as regional vernaculars [7]. As speech interfaces are increasingly integrated into everyday communication, including voice assistants, automated customer service, and assistive technologies, improving performance across diverse speech populations has become both a technical and societal priority.

Recognizing this gap, the Interspeech community has shown an increasing interest in speech diversity research, reflected most recently in its 2026 conference theme of "*Speaking Together.*" Given the growing number of publications contributing new models, datasets, and systems aiming to promote this theme, we believe it is an opportune moment to critically examine and reflect on the history of Interspeech to identify future opportunities to better promote inclusivity.

As a step toward this goal, we review existing papers on accessible and inclusive speech technology through the lens of neurodiversity. Neurodiversity frames cognitive and communicative differences as natural variations rather than solely clinical deficits. As an umbrella term encompassing a wide range of conditions (e.g., autism, dyslexia, ADHD, and stuttering), neurodiversity is an apt lens to evaluate inclusivity in Interspeech for three reasons: (1) it affects a significant portion of the population – estimated at 15–20% [8] – and has frequently been the target for inclusion-oriented research efforts; (2) it often affects speech and communication in a non-degenerative and non-reversible way; and (3) as a social and political movement, neurodiversity fosters rich discourse and scholarship on the design and evaluation of assistive and everyday technologies [9, 10, 11]. Leveraging neurodiversity as a case study, we critically examine the existing body of work on speech diversities in the past 50 years. Our systematic review covers Interspeech and IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) – two of the largest speech science and technology conferences – publications from 1976 to 2025 that involve the speech of neurodiverse communities. Instead of cataloging the technical methods and systems described in these publications, our review centers on how they engage with real-world contexts and diverse stakeholders – particularly people with disabilities (PWD) – by examining their research motivations, data practices, evaluation processes, and primary contributions. This approach allows us to identify key trends, common practices, and opportunities for more meaningful inclusion of impacted populations into the research and development of speech science and technologies.

Through our analyses, we find that neurodivergent speech research often takes a pathologizing, deficit-based approach, aiming to "cure" or "correct" neurodivergent traits in order to fit the neurotypical norm. Other stakeholders, like clinicians or PWD, are rarely involved in the research process, normalizing distance and even disconnection between researchers and the populations they aim to support. The use of ableist-style language and misconceptions about neurodiverse conditions further contribute to the othering of PWD. We hope our review inspires reparative reflection within the Interspeech community. We encourage the field to explore new problems grounded in the experiences of people with speech diversities, foster partnerships between researchers and stakeholders, and cultivate an anti-ableist mindset around the research process and its impact.

2. Background & Related Work

2.1. Models of Disability

Different models of disability have been developed to conceptualize disability and shape research priorities. Historically, the **medical model of disability** has been the dominant framework in research and society, positioning disability as an individual pathology or deficit in need of diagnosis, treatment, prevention, or cure [12, 13]. While a useful framework for designing assistive technologies, the medical model of disability has increasingly received criticism from disability scholars and activists for framing disability as deterministic and biologically inferior, reinforcing social stigma around disability-related traits [11, 14]. This model reproduces asymmetric power relations by framing the individual as deficient and positioning medical and technical professionals as the authorities that define needs, set goals, and determine success [15, 16]. PWD are then treated as passive recipients of care instead of as agents entitled to access and participation [12]. This dynamic can elevate professional gatekeeping, amplifying the social consequences of labeling and stigma.

In technology and engineering contexts, the medical model often manifests through research and design that prioritize diagnosis, detection, and symptom reduction. Such framings treat disability as a technical anomaly requiring correction rather than as a sociotechnical mismatch that should be addressed through accessibility, accommodation, and environmental changes [17, 18]. Adopting such medicalized perspectives risks pathologizing ordinary variation in communication and social interaction. It also shifts attention away from structural causes (e.g., inaccessible environments, stigma, discrimination) and toward individual correction [19]. Overall, critiques of the medical model do not deny the reality of impairment or the value of clinical care. Rather, they question the assumption that care should aim to make PWD behave more "normally" and that professional and clinical institutions should define goals and standards of success on their behalf [17, 16].

In contrast, the **social model of disability** argues that disability arises from the way society is structured: an individual is not disabled by their physical condition but by environmental, social, and systemic factors [20]. Developed by PWD, the social model of disability has been embraced by the disability community to recognize the value of disability experiences and advocate for societal changes that ensure full inclusion of PWD [11]. This perspective reframes disability as a matter of civic rights and social justice, highlighting how barriers in architecture, education, and communication norms restrict PWD's participation in society [12]. PWD communities have leveraged the social model to advocate for more inclusive practices and designs. Members of the stuttering community, for example, have argued that both listener and speaker are involved in "constructing a disabled speaker" [21], claiming that communicative breakdown is not produced solely by disfluent speech patterns, but also by listeners' expectations of fluency, speed, and uninterrupted turn-taking. Within the stuttering community, more people are embracing stuttering pride [22], reframing their experiences to find value and strength in their stuttering while advocating for recognition and reduced stigma [13]. Overall, the social model of disability places responsibility for inclusion onto society and institutions rather than individuals alone [17, 16, 18].

2.2. Neurodiversity Movement

The neurodiversity movement originated in the late 1990s as an alternative to medicalized views of cognitive and communicative differences [23]. Initial efforts aimed to reframe conditions such as autism, ADHD, and dyslexia as natural variations in human cognition rather than pathologies. Today, neurodiversity has sparked social and political movements for the rights and liberation of those living with developmental neurological conditions. For example, these efforts now include promoting inclusive education and employment, reducing stigma, challenging deficit-based language, and advocating for structural and environmental changes that support diverse ways of thinking and communicating [8, 24]. Neurodiversity advocates emphasize the unique skills and advantages of different neurological traits and demand the acceptance, recognition, and the right to live authentically with these traits [25].

In recent years, the neurodiversity framework has expanded to include speech diversity such as stuttering [14]. This addition empowers people who stutter to find meaning and joy in their stuttering experience while still acknowledging their vulnerability to social and structural exclusion that necessitates accommodations from listeners.

2.3. Technoableism and Disfluent Speech

Despite the social model of disability and the neurodiversity movement challenging deficit-based notions of disability, ableism – the assumption that able-bodied and neurotypical experiences are superior – remains deeply embedded in contemporary society. This is reflected in the tools aimed to help such individuals. Technoableism is enmeshed in assistive technologies that are designed to make PWD *appear* or *act* "normal" rather than augment or remove the structural barriers that pathologize alternative means of expression and living [10]. While these technologies may benefit some users, they often function as what activists disparagingly call "disability dongles" – tools developed without meaningful input from disabled communities that fail to address their real-world needs [26].

Speech technologies often exemplify technoableism, treating fluent speech as the default and the ideal design goal. Products specially designed for people with speech disfluency often focus on eliminating disfluencies, either by altering the behavior of the speaker or by modifying their speech. SpeechEasy[1] aims to "fix" speakers, training their users to speak more fluently – with significant effort and financial costs. Google's Project Relate[2] directly targets "non-standard" speech, providing functionality such as closed-captioning, repeating through a computerized voice, and text-to-voice generation, all of which aim to make the "speaker" more fluent.

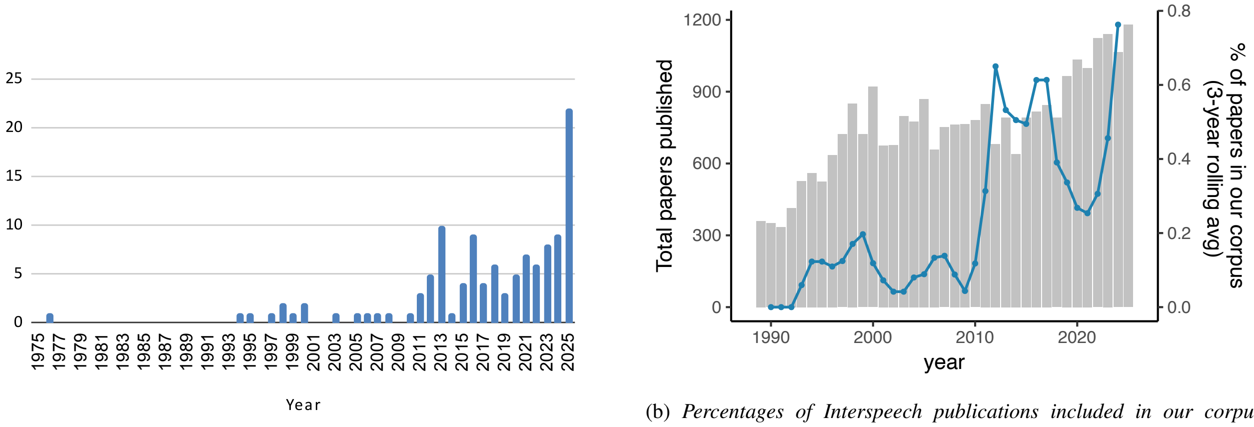
Disfluent communities have increasingly resisted the ableist desire for fluent speech, challenging the medical necessity of speech restructuring for merely auditory aesthetics [27]. Instead, many advocates promote an affirmative narrative around speech disfluency, which views stuttering and other speech differences not as defects to be eliminated but as natural forms of human communication [22, 28]. Central to this shift is the disability justice principle of "**nothing about us without us**" principle [29]; disabled people must be meaningfully involved in decisions, research, and technologies that shape their lives. This perspective has begun to reshape public discourse and technological development. For example, the stuttering community recently organized a successful public campaign against an AI-assisted fluency-first app marketed by Samsung, demanding that the voices of people who stutter be heard rather than suppressed by speech technology[3].

---

[1] `https://speecheasy.com`

[2] `https://sites.research.google/relate/`

[3] `https://www.spacetostutter.org/impulse`



(a) *Histogram of publications in our corpus*

(b) *Percentages of Interspeech publications included in our corpus (shown in the line and points), and histogram of total Interspeech (EuroSpeech and ICSLP before 2004) publications (shown in bars)*

Figure 1: *Temporal trends of publications in our corpus*

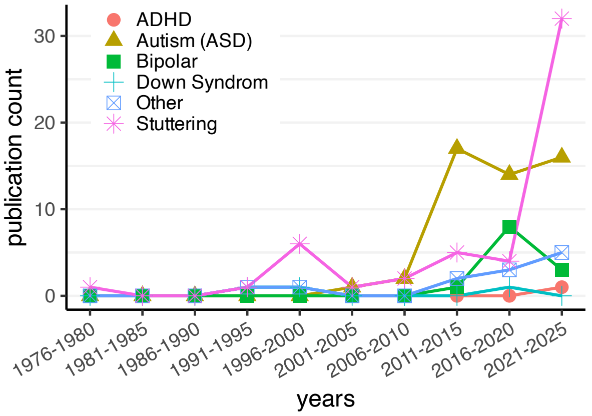


Figure 2: *Number of publications on different conditions over time*

3. Methods & Data

In an effort to steward future research objectives and inclusive practices, we set out to answer the following questions:

- How is disability and neurodivergence framed in speech technology papers?
- How involved are PWD and other stakeholders in the research process?
- Who primarily benefits from the technological innovations being described?
- How have approaches to disability and neurodivergence in speech technology research changed over the last 50 years?

We first identified papers about disfluent or atypical speech published at Interspeech and ICASSP, the top two publication venues in the Acoustics & Sound space [30]. Using the word stems 'stutter,' 'disfluen,' 'dysfluen,' 'disab,' 'disord,' and 'atyp' as search terms, our IEEE Xplore search yielded 605 ICASSP papers. Using a Python script to extract metadata and search the ICSA archives yielded an additional 565 papers published at Interspeech and its predecessors, EuroSpeech and the International Conference on Speech and Language Processing (ICSLP).

We reduced this initial corpus of 1170 papers to 155 papers by only including papers that explicitly mention developmental neurodivergent conditions affecting speech in their text: stuttering, Attention Deficit Hyperactivity Disorder (ADHD/ADD), Autism Spectrum Disorder (ASD), Bipolar Disorder (BPD), and Down Syndrome (DS). Upon closer reading by the research team, an additional 38 papers were excluded because they did not meet the inclusion criteria of being about speech and focusing on the conditions mentioned above (e.g., some papers only mention stuttering as an example of disfluency in their abstract or use electromagnetic – but no speech – signals).

The finalized corpus of papers our team analyzed in-depth includes 117 papers published from 1976 to 2025 at Interspeech (75 papers), ICASSP (33 papers), ICSLP (5 papers), and EuroSpeech (4 papers). Within this corpus, the number of publications addressing neurodivergence in speech technology started to increase in the 2010s with the highest number of papers published in 2025 (Figure 1a), indicating a rising interest in addressing topics of neurodiversity and accessibility. We also find that the rising interest in neurodivergent speech has outpaced the overall growth of the field. Although it remains a relatively small research area, the percentage of Interspeech publications on neurodivergent speech has increased in general since 2010, reaching nearly 0.8% of all Interspeech publications in recent years (Fig. 1b). Among this growing body of research, stuttering and autism were the most studied conditions, with 52 and 50 papers, respectively. Bipolar disorder (12 papers), Down Syndrome (3 papers), and ADHD/ADD (1 paper) were less frequent. 8 papers examined other conditions (e.g., depression, developmental language disorder (DLD), dysphonia) in addition to one of the target conditions. Figure 2 illustrates the temporal trends of different neurodivergence-related topics in our corpus, showing a substantial increase in research interests in stuttering and ASD in recent years. ASD gained significant traction in 2013 when Interspeech hosted the "Computational Paralinguistic Challenge" which included a sub-challenge to classify ASD and non-ASD speech [31]. Since then, the topic has received sustained interest from the Interspeech community. Stuttering, on the other hand, has only recently become a research focus. Between 2021 and 2025, stuttering was the topic of almost 62% of papers about neurodivergent speech, a large increase from 14% in the 2016–2020 period.

To understand how neurodivergent speech research is ini-

Table 1: *Paper evaluation rubric*

Dimension	Manual Coding Method
Inclusion Criteria	Is about speech? [Yes / No] Is about stuttering, ADHD, ASD, bipolar, or Down syndrome? [Yes / No]
Paper Type	Topic: Which topics related to neurodiversity does the paper address? [ADHD / Autism / Bipolar / Down Syndrome / Stuttering / Other] Other Topic (if "Other" selected): short description Paper Type [Full paper / Challenge paper / Show & Tell / Workshop Paper / Speech]
Main Contribution	Main Contribution [Model / Insight about condition / Dataset / Artifact / Other] Main Contribution: short description
Intended Use Case	Use Cases [Diagnosis/Detection / Intervention for technology / Intervention for PWD / Intervention for speech / Masking/Hiding / Prevention / Other] Use Case: short description Intended audience/end users [Professionals working with PWD / PWD / Other] Intended audience/end users: short description
Data Practice	Data Source [Existing dataset / New data] Data Source: short description of collection method and stakeholders involved Annotation [Used existing annotations / Added new/additional annotation] Annotation: short description of who annotated, what method was used Annotators [Researchers / Professionals/clinicians / PWD / Non-experts / Unknown/other]
Evaluation	Human Evaluation [Yes / No] If conducted: short description of who evaluated (PWD, clinicians, etc.) Subjective Metrics: short description of subjective metrics used, e.g., PWD experiences, PWD raters Objective Metrics: short description of objective metrics used, e.g., accuracy, F1
Language	Instances of ableism-associated language: use of terms such as "fix," "hide," "disease," "cure"

tiated, operationalized, and applied in relation to real-world contexts and stakeholders, we critically examined each paper's research topic, primary contribution, motivating use case, intended audience and end users, data sources, data annotation methods, evaluation processes and metrics, and use of ableist-style language (Table 1). Our analysis, drawing from both content and discourse methods, foregrounds not only the *content* of each individual paper but also how it *delivers* its content through words, tables, and imagery [32, 33]. Papers were split up and coded independently by the research team members. We met weekly to discuss observations and questions, and to ensure alignment on evaluation criteria. The corpus is provided in the supplemental material.

4. Results

The research in this corpus is largely technical and primarily aims to assist clinicians, creating an outsized impact of research from medicalized perspectives of neurodivergent conditions over those from PWD themselves. The following sections provide high-level observations of how neurodivergence is viewed and researched in speech science and technology. We find tendencies to medicalize, distance, and other neurodivergence in surveyed papers. As a result, existing studies often adopt ableist-style language (e.g., aiming to "correct" neurodivergent traits) and rarely consult with or involve PWD directly at any point in the research process. We use these insights to propose opportunities for fairer and more inclusive speech science and technology in the Discussion section.

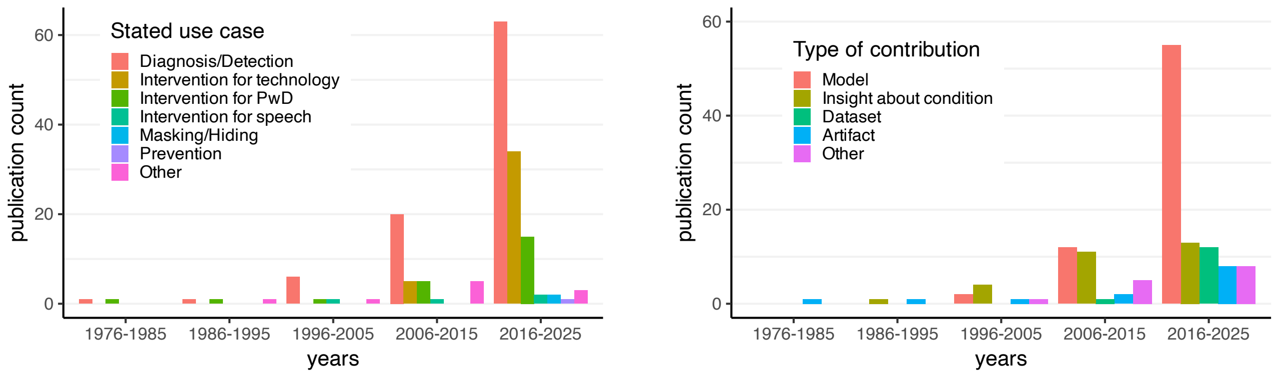
4.1. Medicalizing Conditions

The reviewed speech technology research often viewed neurodivergence through a medicalized lens. Despite the overall growing number of studies investigating neurodivergence and speech (Fig. 1a), the people benefiting from these novel models, insights, and systems were often *clinicians* working with PWD rather than PWD themselves.

Professionals working with PWD were the target audience of the proposed technological innovations in 56% of papers. In comparison, PWD were the target end users in 16% of papers. This focus on clinical contexts and audiences is also reflected in the intended use cases (Fig. 3a) and main contributions (Fig. 3b) of the corpus.

Almost 60% of papers contributed a novel model or training method ("Model") and 25% generated new insights into the condition under study ("Insight about condition"), which includes speech patterns and characteristics of people with the condition. 13 papers developed or tested an interactive system ("Artifact") and 13 papers presented a new dataset ("Dataset"). An additional 13 papers made other contributions, such as insights about speech production, data collection methods, or analyses of existing models, methods, and systems. Over time, and with advancements in AI and ML research, the number of model contributions increased significantly within the last decade, constituting $\sim$70% of publications from the last decade (2016–2025).

As evident from Figure 3a, diagnosis and detection are by far the most common use cases. In fact, 78% of papers in our corpus framed their contribution around diagnosing and detecting disfluencies or other speech characteristics (e.g., prosody, mood). This trend has been consistent over time: In the three decades since 1996, "diagnosis and detection" remained the top use case that motivated the neurodivergent speech research covered by our review (see Fig. 3a). Many of these papers propose systems, methods, or models to help clinicians and speech therapists count, classify, and annotate disfluencies to aid the diagnosis of stuttering or ASD (e.g., [34, 35, 36, 37]). The overall goal of these contributions is to make diagnosis and detection



(a) *Distribution of stated use cases*

(b) *Distribution of main contributions*

Figure 3: *The evolution of claimed contributions and stated use cases by surveyed papers over five decades*

less labor-intensive for clinicians. There was little focus on how PWD would be affected by these innovations.

In comparison, fewer papers focused on interventions targeting existing technologies (e.g., improving ASR for stuttered speech [38]), PWD (e.g., developing a symptom monitoring system to be used by PWD [39]), or speech itself (e.g., modifying speech to increase prosody [40]). However, such intervention-oriented use cases were again frequently positioned in service of detection and diagnosis. In fact, two-thirds of papers that described an intervention use case explicitly linked it to diagnostic or detection goals. Many contributions were framed as technological improvements intended to make clinical assessment, diagnosis, or monitoring more accurate and efficient (e.g., [41, 37, 42]). For example, [43] proposes a speech recognition system for speech therapy clients (i.e., an intervention for PWD) for the purpose of early detection and diagnosis of stuttering.

A small number of papers aimed to "treat" the symptoms of neurodivergent conditions by training PWD to better hide or mask them. For example, [44] proposes a method for the automatic discrimination and evaluation of soft/hard consonant articulation. They propose using this method to develop self-guided speech therapy interventions for people who stutter to help them practice speech patterns that mask their stutter and sound more fluent.

Altogether, pathological and medicalized approaches have dominated existing speech science and technology research on neurodivergent speech. Although only a relatively small proportion of the surveyed papers explicitly aimed to "cure" PWD, many tied their contributions to medical use cases such as diagnosis, symptom monitoring, or treatment of neurodivergent conditions. In addition, clinicians were frequently identified as the target users of technological innovations. This sustained emphasis on medical use cases and professional audiences reinforces the medical model of disability.

4.2. Distancing Stakeholders

Similar to findings in other fields [45, 9], many papers in our corpus took a "distanced position," where PWD were rarely involved in the research process and evaluation metrics were usually limited to technical benchmarks. This approach directly conflicts with calls for meaningful participation from the population of interest, as articulated in the "nothing about us without us" principle [29], often resulting in undermined agency of PWD and interventions not directly responsive to the needs of the community [46, 45].

Distancing was partially achieved by approaching the diagnosis of neurodivergent conditions from a purely technical perspective. In other words, papers were framed as contributing to (improved) algorithmic diagnosis and detection of "symptoms." Such contributions were primarily evaluated through quantitative benchmarks, such as accuracy, recall, or F1 scores. Without discounting the merits of this approach, we argue that the *sole* reliance on technical perspectives distances the reader and researcher from PWD, reducing them to data points and algorithmic features. Only a few papers discuss the limitations of this computation-centered approach (e.g., [47, 48, 49, 50, 51, 52]). [52], for example, states that their approach to ASD detection using acoustic and text data "is not yet valuable as a method for diagnosing ASD, which still needs to be diagnosed by expert clinicians." Limitations beyond technical capabilities, model architecture, and data availability were rarely mentioned.

Distancing is also evident in the lack of participation from PWD in research problem formation, design, and evaluation. No paper mentioned anyone outside of the research team being involved in the problem formation.

Noting the importance of accurate and scalable data in speech science and technology research, we investigated the data practices described in surveyed papers, categorizing the sources of data used as well as the methods and stakeholders involved in the data annotation process (see Table 1).

In most cases, researchers engaged PWD solely as data contributors, either directly through participant recruitment for data collection [37, 52, 44] or indirectly by utilizing existing datasets (e.g., SEP-28k, CPSD, PRIORI, USC CARE). In a few cases, when PWD's data was sourced from third parties, it was unclear whether participants' explicit consent was obtained. For example, [53] sourced videos of children with autism experiencing a meltdown from YouTube by searching for terms like "autism child shouting" but does not mention seeking consent from the uploader or people in the videos. Similarly, [54], one of the most highly cited papers in our corpus, presents a dataset of stuttered speech from podcasts with people who stutter without mention of seeking consent from the podcast hosts or guests.

In the data annotation phase, 55% of studies (65 papers) in our corpus conducted new or additional annotations to their dataset. Figure 4a shows the breakdown of the groups that performed the data annotations in these studies. Most of the time, the research team did the annotation (62%). When additional annotators were involved, they were most commonly professionals such as clinicians or speech language therapists (SLPs) (26%). Non-experts, such as Amazon Mechanical Turk workers, were involved in the annotation process in 6% of papers that conducted annotation. PWD aided in data annotation in only 2 papers (3%).

When examining whether and how the proposed methods, systems, or models were evaluated (Figure 4b), we find only about 20% of papers (n = 23) involved stakeholders – such as clinicians or PWD – in the evaluation phase. Among them, only two ($< 2\%$) directly solicited feedback from PWD during evaluation [55, 56].

Though human evaluation is not common in technology-focused venues such as Interspeech, a few notable examples in our corpus show it is possible for such papers to be more inclusive – to involve other stakeholders, acknowledge participants' needs and contributions, and discuss limitations inherent in a technology-centered approach. For example, Heeman et al. [35] aims to identify instances of stuttered speech by presenting a computer tool that assists speech-language therapists (SLP) in counting and annotating clients' disfluencies. To assess whether the tool aligns with the needs of speech therapists, the researchers conduct a user study with therapists who evaluate and correct the automatically generated transcriptions and annotations. Based on insights from a follow-up questionnaire, they were able to implement improvements detailed in [36]. Valente et al. [57], on the other hand, present a novel annotation scheme for the FluencyBank dataset based on clinical standards. They hire three SLPs to annotate disfluency types, behavioral signals, and tension levels in audiovisual samples of adults who stutter. In addition to objective evaluation metrics, the authors describe challenges reported by the SLPs during the annotation process, highlighting the subjectivity of disfluency detection and classification. Both of these publications make largely technical contributions while incorporating human evaluation techniques.

Overall, the reviewed papers showed relatively low non-technical/non-clinical/PWD stakeholder involvement. This, in addition to the tendency to reduce the complex experiences of neurodivergent individuals to a small number of metrics, distances neurodivergent communities from research that is precisely meant to benefit them.

4.3. Othering Neurodivergence

Neurodivergence was often viewed through an "othering" lens, distinguishing it as an edge case for technology and using language to characterize PWD as "atypical." Under this lens, data from neurodivergent individuals served to showcase the robustness of technical contributions beyond "normal" cases. This approach to neurodivergence is especially reflected in the language of papers.

Language that reflects ableist ideas and assumptions was frequently spotted in the reviewed papers. Neurodivergent conditions were often described as "abnormal" and "pathological." The behavioral characteristics of people with the condition were termed "symptoms," "impairments," and "errors," illustrating a deficit approach to neurodivergence. In contrast, data collected from control groups consisting of people without neurodivergent conditions were often labeled as "normal" and "correct". Notably, across the span of 50 years, ableist-style language remained common in the reviewed publications (e.g., [58, 59, 34, 60, 41].

In some cases, introduction sections contained outdated information about the condition of interest to underpin the value of their contribution. For example, a number of papers about stuttering [61, 62, 34] perpetuate the notion that "it is critical to address speech disorder problems in early childhood, since delays in medical interventions can result in wide-ranging social and mental difficulties" [34]. While this assertion is often used to motivate stuttering detection research, it reflects an outdated understanding of stuttering. Studies in SLP and neuroimaging have established stuttering as a neurodevelopmental condition, with data showing that 80% of children who stutter will naturally grow out of it as their brains develop, regardless of interventions [63, 64]. People who stutter into adulthood often experienced negative reactions toward stuttering during their younger years, driven by social stigma and the pressure to "correct" it as soon as possible. In fact, current SLP research and clinical practice have advocated for *disfluency-affirming* approaches, which are linked to long-term mental-health benefits [65, 28]. While there is real value in developing effective stuttering detection models (e.g., making speech AI systems more accessible for people who stutter), the incorporation of outdated clinical knowledge only serves to further reinforce the social stigma associated with neurodivergence.

5. Discussion

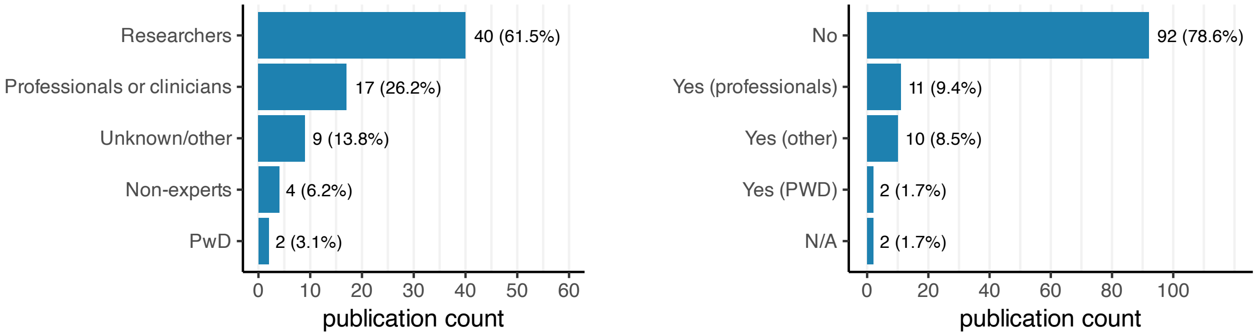
Our review revealed that there has been an increased interest over the past 50 years in working with diverse groups and promoting inclusivity and accessibility in speech science and technology research. Despite this, the existing body of research on speech and neurodivergent conditions has frequently taken a medicalizing, distancing, and othering approach to PWD impacted by these technologies, adopting the medical model of disability as a neutral default. In response, we propose three action items that aim to draw critical awareness towards the byproducts of current approaches and highlight opportunities to move towards structural fairness and meaningful inclusion of PWD in speech science and technology research.

5.1. Incorporating the Social Model of Disability

Our analysis reveals that existing research has framed neurodivergence predominantly through a medical model lens, where speech differences are characterized as deficits to be corrected. Past work has largely focused on clinical use cases, such as early detection, diagnosis, and monitoring of neurodivergence, often positioning clinicians as the intended audience for technical innovations. PWD, despite being the subjects of diagnostic tests and therapeutic interventions, were rarely directly targeted. While this approach has contributed valuable assistive technologies and interventions, its position as the implicit default framework limits the field's potential to address the full spectrum of needs, preferences, and lived experiences of PWD.

Our caution against the medical model of disability does not require rejecting medical perspectives entirely. The medical model of disability has generated important insights into speech production mechanisms and clinically valuable interventions. In some use cases, technologies focused on improving intelligibility or providing alternative communication methods represent important tools for participation and autonomy.

However, our work calls attention to the medical model as the unmarked default and dominant framework for approaching such technologies. Such dominance narrows the scope of research questions exclusively to individual pathology over en-



(a) *Distribution of who performed data annotations, if new or additional annotations were conducted. Non-experts are often crowd workers, and Unknown/other represents papers that did not provide details about who annotators were.*

(b) *Distribution of how often humans were involved in system evaluations, and if so, who the evaluators were. N/A represents papers where no evaluation was conducted, e.g., a method or dataset was introduced/proposed but not evaluated.*

Figure 4: *Distribution of personnel involved in result evaluation and data annotation, as described in the survey papers. Note that the total percentages in (a) add up to $>$ 100% because some of the 65 papers with new annotations had multiple types of annotators.*

vironmental or systemic barriers, normalizes a presumed universal goal of fluency, and limits technology's capacity to engage with the full spectrum of speech expression. Moreover, by prioritizing clinician perspectives of disability over PWD's lived experiences, the field risks developing technologies that inadvertently create new burdens or reinforce stigmatizing assumptions.

We thus advocate for future research agendas in speech science and technology to incorporate the social model of disability when engaging with the speech of PWD. The social model enables the field to focus on environmental and social barriers, as well as supporting a broader spectrum of communication diversity. Concretely, this means shifting attention from exclusively improving detection or correction toward examining how modeling assumptions, training data distributions, and benchmark design may disadvantage certain speech patterns. Research could, for example, address how communication environments construct disability for people with speech differences, develop adaptive systems that reduce the burden on speakers to conform to normative standards, or critically assess whether optimization objectives implicitly prioritize fluency over communicative effectiveness. The growing number of recent studies exploring interventions on speech technology to better interact with PWD suggests a promising trend in this direction (see Fig. 3a).

5.2. Increasing Direct Engagement with PWD Throughout the Research Process and Beyond

PWD have so far primarily served as data contributors, and their first-hand perspectives have rarely been consulted when forming research questions and evaluating the benefits and limitations of technical contributions. When stakeholders outside of the research team are invited to the table, they are most often clinicians, thereby implicitly prioritizing medical perspectives on neurodivergence over social ones. This pattern reflects and reinforces the medical model's dominance – positioning clinicians as the experts on disability experiences while marginalizing the lived expertise held by PWD themselves.

PWD are uniquely positioned to steward research related to their own experiences, knowing best what challenges they face, what solutions should address, and whether they will be effective. For example, in [66], the authors – several of whom are people who stutter – leveraged their connections within the stuttering community to collect and meticulously annotate 50 hours of Mandarin stuttered speech from 70 speakers, generating the first and largest Mandarin stuttered speech dataset contributed, enabling the evaluation and development of stuttering-friendly speech technology [67]. This example illustrates how lived expertise and knowledge enable not only access to communities but also a nuanced understanding of what the data mean and what they represent. Work that does not directly include the lived experience of these individuals as experts risks inaccuracy in representation, lack of robustness, inefficacy, and can even create additional burdens as PWD create their own workarounds for technology that does not expressly address their needs. Next, we propose concrete ways to directly engage with PWD throughout the research process.

Problem Formation: During this formative phase, researchers can leverage user research methods from the field of human-computer interaction [68], such as focus groups or surveys with PWD to generate relevant, meaningful, and valid research questions. Such participatory approaches can be highly beneficial to align research with real-world needs [69]. In addition, early and frequent iteration on research questions can yield fundamentally different research directions, such as ones that address social barriers, challenge normative assumptions, or prioritize goals that technical metrics alone cannot capture [70].

Data acquisition and annotation: During data collection and preparation, reporting on qualitative insights from participants, annotators, and researchers can offer new perspectives. PWD can aid in annotation, as they often have a unique perspective on data related to their own experiences, recognizing patterns, contributing context, and supplying much-needed nuance that might otherwise be missed. For example, Sridhar and Wu [55] – both identified as people who stutter – manually annotated a subset of the most prominent stuttered speech dataset (Sep-28K dataset [54]), identifying approximately 20% mislabled stuttering events by leveraging their embodied knowledge of stuttering. Similar to Valente et al.[57], researchers who do not have lived experience within their core team can involve external stakeholders in the annotation process. For example, PWD or, if needed, clinicians may be invited to review labels, provide contextual feedback, or reflect on annotation decisions [71]. Reporting these contributions can surface limitations in existing strategies, highlight overlooked nuances, and provide the research community with insights that would otherwise remain invisible. In this way, participatory annotation does not require that all researchers themselves be PWD, but it does create space for lived expertise to shape and inform the dataset.

Evaluation: In the evaluation phase, researchers can supplement technical benchmarks with assessments by human evaluators to better align systems with stakeholder needs. A small number of papers in the reviewed corpus incorporated subjective or qualitative evaluation metrics (e.g., [72, 73, 35, 56, 55]), demonstrating that such approaches are feasible within technical research contexts. Expanding the use of human evaluation methods not only yields richer data, it ensures that the technology *in deployment* sufficiently addresses users' needs. Technical metrics like error rate cannot fully capture how a system is experienced in use, whether it supports users' communication goals, or whether the technology may inadvertently introduce unintended social or usability consequences. Evaluation processes that center PWD perspectives enable technology to more accurately reflect users' needs and values, build trust and empathy between researchers and PWD communities, and ultimately make more meaningful contributions.

Ongoing stewardship: Meaningful engagement does not end with paper publication or technology deployment. Research artifacts – including datasets, models, and annotation schemas – continue to shape downstream applications, benchmarks, and design decisions. Researchers and organizations should therefore consider post-publication practices such as documenting dataset limitations, providing mechanisms for community feedback, remediating harms when they are identified, and clarifying appropriate use cases. Sustained engagement may also include establishing advisory relationships with stakeholders, compensating contributors beyond one-time data collection, and supporting governance mechanisms that allow affected communities to influence how their data are shared and used [74, 69, 75, 76]. In practice, this can mean involving PWD not only as participants and data contributors, but as collaborators, co-authors, or research partners in subsequent iterations of the work.

5.3. Reflecting on framing and social impact

Many of the reviewed papers use ableist-style language, treating PWD as an anomaly rather than a natural part of humanity. The ubiquity of this framing can lead to "othering," positioning people with disabilities as intrinsically different. This risks obscuring the fact that they are an important user group deserving of the same levels of utility, usability, and autonomy as their non-disabled peers. Deficit-based models obscure the fundamental reality that disability is not a fixed characteristic of a minority population, but a universal human experience [77]. Most people will experience temporary, recurring, or permanent disabilities, whether through injury, illness, aging, or other shifts in social expectations around capability.

Recognizing disability as a shared part of human experience rather than a signifier of "Other"-ness has profound implications for research and technology. When we position speech disability technology as something that could benefit anyone, the research questions shift from how to make a disabled person sound more "normal" to how to make speech technologies more accommodating of human diversity and variations. Instead of assuming a narrow form of expected speech, the technology must be robust in being able to truly handle all forms of speech. Importantly, this reframing does not erase the unique challenges of PWD; it challenges the false binary between "disabled" and "normal."

We therefore urge researchers in speech science and technology to examine and acknowledge the sociotechnical frameworks informing their work, especially when working on topics related to marginalized groups such as PWD. For example, by critically reflecting on whether the research has framed the speech by PWD as a "problem" or "impairment" to be fixed by technology, researchers can become more aware of the underlying model of disability (e.g., medical vs. social) and associated research methods (e.g., deficit-based vs. asset-based) shaping their work. More broadly, we call for greater recognition and transparency of which questions are not asked, whose perspectives are centered, and how research aligns with – or contradicts – the values and goals of impacted communities.

Beyond research merits and quality, attending to social disability approaches can also bare significant social impact. The language researchers use, the narratives we construct, and the way we relate to research artifacts and technology all shape public understandings of disability. These messages affect not just researchers but how disabled people feel about themselves – through internalized ableism, self-stigmatization, and self-acceptance [78, 79]. We thus encourage researchers to consider not just the scope of their field-specific achievements but their broader contributions to PWD's agency and autonomy. By attending to potential unintended impacts, researchers can also aim to ensure that their technology empowers, rather than constrains, members of disabled communities.

6. Conclusion

We present a scoping review of Interspeech and ICASSP papers on neurodivergent speech published over the past five decades, with particular attention to how these studies were motivated, operationalized, and evaluated in relation to key stakeholders such as people with disabilities. Our analysis uncovers the dominance of the medical model of disability in shaping research questions and contributions. We also find limited direct engagement with PWD in research design, developing, and evaluation, alongside the frequent use of ableist-style language and deficit-based framings in current work.

Lastly, we acknowledge that our analysis is one with the benefit of hindsight. Each paper is inescapably a product of its time, as is our own. Our intention is not to call out individual papers but to take this moment to encourage the speech research community to reflect upon the past and present. Researchers have the opportunity to shape a more just future where speech technology supports PWD through academic research and industry partnerships. We hope that by embracing diverse perspectives, assessing the social impact of one's research, and building strong relationships with marginalized communities, the Interspeech community can truly achieve its goal to celebrate *"**all** voices"* and *"improve access to speech technology for **all** speakers."*

7. Acknowledgments

This material is based upon work supported by the National Science Foundation Award No. 2427710 and the Patrick J. McGovern Foundation.

8. Generative AI Use Disclosure

Generative AI tools have been used to polish language in parts of this manuscript. All underlying work and intellectual contributions were carried out entirely by the authors.

9. References

[1] S. E. Brennan and M. F. Schober, "How listeners compensate for disfluencies in spontaneous speech," *J. Mem. Lang.*, vol. 44, no. 2, pp. 274–296, 2001.

[2] C. Lea, Z. Huang, J. Narain, L. Tooley, D. Yee, D. T. Tran, P. Georgiou, J. P. Bigham, and L. Findlater, "From user perceptions to technical improvement: Enabling people who stutter to better use speech recognition," in *Proc. of CHI'23*, 2023.

[3] Q. Li and S. Wu, "Towards fair and inclusive speech recognition for stuttering: Community-led chinese stuttered speech dataset creation and benchmarking," in *Proc. of EA CHI'24*, 2024.

[4] A. Glasser, "Automatic speech recognition services: Deaf and hard-of-hearing usability," in *Proc. of EA CHI'19*, 2019, p. 1–6.

[5] S.-E. Kim, B. R. Chernyak, O. Seleznova, J. Keshet, M. Goldrick, and A. R. Bradlow, "Automatic recognition of second language speech-in-noise," *JASA Express Letters*, vol. 4, no. 2, 2024.

[6] K. Wenzel and G. Kaufman, "Designing for harm reduction: Communication repair for multicultural users' voice interactions," in *Proc. of CHI'24*, 2024.

[7] A. Koenecke, A. Nam, E. Lake, J. Nudell, M. Quartey, Z. Mengesha, C. Toups, J. R. Rickford, D. Jurafsky, and S. Goel, "Racial disparities in automated speech recognition," *PNAS*, vol. 117, no. 14, pp. 7684–7689, 2020.

[8] N. Doyle, "Neurodiversity at work: a biopsychosocial model and the impact on working adults," *Br Med Bull*, vol. 135, no. 1, pp. 108–125, 2020.

[9] R. M. Williams, L. Boyd, and J. E. Gilbert, "Counterventions: a reparative reflection on interventionist hci," in *Proc 2023 SIGCHI Conf Hum Factor Comput Syst*, 2023, pp. 1–11.

[10] A. Shew, *Against Technoableism: Rethinking Who Needs Improvement*. W. W. Norton, 2023.

[11] A. Hamraie and K. Fritsch, "Crip technoscience manifesto," *Catalyst: Feminism, Theory, Technoscience*, vol. 5, pp. 1–34, 04 2019.

[12] M. Oliver, "The politics of disablement—new social movements," in *The politics of disablement*. Springer, 1990, pp. 112–131.

[13] C. Constantino, P. Campbell, and S. Simpson, "Stuttering and the social model," *J. Commun. Disord.*, vol. 96, p. 106200, 2022.

[14] C. D. Constantino, "What can stutterers learn from the neurodiversity movement?" *Semin. Speech Lang.*, vol. 39, no. 4, pp. 382–396, Sep. 2018.

[15] I. K. Zola, "Medicine as an institution of social control," *The sociological review*, vol. 20, no. 4, pp. 487–504, 1972.

[16] S. Tremain, "Foucault, governmentality, and critical disability theory," *Foucault and the government of disability*, vol. 1, pp. 1–24, 2005.

[17] T. Shakespeare, "The social model of disability," in *The disability studies reader*. Routledge, 2006, pp. 16–24.

[18] D. Goodley, *Disability studies: An interdisciplinary introduction*. SAGE Publications Ltd, 2024.

[19] P. Conrad, *The medicalization of society: On the transformation of human conditions into treatable disorders*. Johns Hopkins University Press Baltimore, 2007, vol. 14.

[20] M. Oliver, "The social model of disability: thirty years on," *Disabil. Soc.*, vol. 28, no. 7, pp. 1024–1026, 2013.

[21] J. S. Pierre, "The construction of the disabled speaker: Locating stuttering in disability studies," *Can. J. Disabil. Stud.*, vol. 1, pp. 1–21, 2012.

[22] Patrick Campbell and Christopher Constantino and Sam Simpson, Ed., *Stammering Pride and Prejudice: Difference not Defect*. J&R Press, 2019.

[23] J. Singer, "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*, pp. 59–67, 1999.

[24] P. Dwyer, "The neurodiversity approach(es): What are they and what do they mean for researchers?" *Hum. Dev.*, vol. 66, no. 2, pp. 73–92, May 2022.

[25] P. Jaarsma and S. Welin, "Autism as a natural human variation: reflections on the claims of the neurodiversity movement," *Health Care Anal.*, vol. 20, no. 1, pp. 20–30, Mar. 2012.

[26] L. Jackson, A. Haagaard, , and R. Williams, "Disability dongle," https://blog.castac.org/2022/04/disability-dongle/, 2022, accessed: 2025-02-03.

[27] Z. Richter, "On stuttering activism and resistance," in *Stammering Pride and Prejudice: Difference not Defect*, P. Campbell, C. Constantino, and S. Simpson, Eds. J & R Pres, 2020.

[28] C. D. Constantino, "Fostering positive stuttering identities using stutter-affirming therapy," *Lang Speech Hear Serv Sch*, vol. 54, no. 1, pp. 42–62, 2023.

[29] J. I. Charlton, *Nothing About Us Without Us: Disability Oppression and Empowerment*, 1st ed. UC Press, 1998.

[30] Google. Acoustics & Sound - Google Scholar Metrics. [Online]. Available: https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=phy_acousticssound&inst=1581658869211165550

[31] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, M. Mortillaro, H. Salamin, A. Polychroniou, F. Valente, and S. Kim, "The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism," in *Interspeech 2013*. ISCA, 2013, pp. 148–152.

[32] N. Phillips and C. Hardy, *Discourse Analysis: Investigating Processes of Social Construction*, 2nd ed. Sage Publications, 2002.

[33] K. Krippendorff, *Content Analysis: An Introduction to Its Methodology*, 4th ed. SAGE Publications, 2019.

[34] S. Alharbi, M. Hasan, A. J. H. Simons, S. Brumfitt, and P. Green, "A lightly supervised approach to detect stuttering in children's speech," in *Interspeech 2018*, 2018, pp. 3433–3437.

[35] P. A. Heeman, A. McMillin, and J. S. Yaruss, "Computer-assisted disfluency counts for stuttered speech," in *Interspeech 2011*, 2011, pp. 3013–3016.

[36] P. A. Heeman, R. Lunsford, A. McMillin, and J. S. Yaruss, "Using clinician annotations to improve automatic speech recognition of stuttered speech," in *Interspeech 2016*, 2016, pp. 2651–2655.

[37] H. Kothare, V. Ramanarayanan, O. Roesler, M. Neumann, J. Liscombe, W. Burke, A. Cornish, D. Habberstad, A. Sakallah, S. Markuson, S. Kansara, A. Faerman, Y. Bensidi-Slimane, L. Fry, S. Portera, D. Suendermann-Oeft, D. Pautler, and C. Demopoulos, "Investigating the interplay between affective, phonatory and motoric subsystems in autism spectrum disorder using a multimodal dialogue agent," in *Interspeech 2021*, 2021, pp. 1967–1971.

[38] D. Mujtaba, N. R. Mahapatra, M. Arney, J. S. Yaruss, C. Herring, and J. Bin, "Inclusive asr for disfluent speech: Cascaded large-scale self-supervised learning with targeted fine-tuning and data augmentation," in *Interspeech 2024*, 2024, pp. 1275–1279.

[39] J. Gideon, E. M. Provost, and M. McInnis, "Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder," in *Proc. of ICASSP'16*, 2016, pp. 2359–2363.

[40] J. P. van Santen, L. Black, G. Cohen, A. B. Kain, E. Klabbers, T. Mishra, J. de Villiers, and X. Niu, "Applications of computer generated expressive speech for communication disorders," in *EUROSPEECH 2003*, 2003, pp. 1657–1660.

[41] S. Huang, J. Deng, J. Kang, and R. Zheng, "Leveraging LLM for Stuttering Speech: A Unified Architecture Bridging Recognition and Event Detection," in *Interspeech 2025*. ISCA, Aug. 2025, pp. 1843–1847.

[42] Ashwini B, V. Narayan, and J. Shukla, "SPASHT: Semantic and Pragmatic Speech Features for Automatic Assessment of Autism," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[43] E. Nöth, H. Niemann, T. Haderlein, M. Decher, U. Eysholdt, F. Rosanowski, and T. Wittenberg, "Automatic stuttering recognition using hidden Markov models," in *6th International Conference on Spoken Language Processing (ICSLP 2000)*. ISCA, 2000, pp. vol. 4, 65–68–0.

[44] K. Ochi, K. Mori, and N. Sakai, "Automatic evaluation of soft articulatory contact for stuttering treatment," in *Interspeech 2018*, 2018, pp. 1546–1550.

[45] J. Mankoff, G. R. Hayes, and D. Kasnitz, "Disability studies as a source of critical inquiry for the field of assistive technology," in *Proc. of ASSETS'10*, 2010, p. 3–10.

[46] K. Spiel, C. Frauenberger, O. Keyes, and G. Fitzpatrick, "Agency of Autistic Children in Technology Research—A Critical Literature Review," *ACM TOCHI*, vol. 26, no. 6, pp. 1–40, Dec. 2019.

[47] M. Jouaiti and K. Dautenhahn, "Dysfluency classification in stuttered speech using deep learning for real-time applications," in *Proc. of ICASSP'22*, 2022, pp. 6482–6486.

[48] S. P. Bayerl, D. Wagner, I. Baumann, F. Hönig, T. Bocklet, E. Nöth, and K. Riedhammer, "A stutter seldom comes alone – cross-corpus stuttering detection as a multi-label problem," in *Interspeech 2023*, 2023, pp. 1538–1542.

[49] S. Khorram, J. Gideon, M. McInnis, and E. M. Provost, "Recognition of depression in bipolar disorder: Leveraging cohort and person-specific knowledge," in *Interspeech 2016*, 2016, pp. 1215–1219.

[50] T. Chaspari, E. M. Provost, A. Katsamanis, and S. Narayanan, "An acoustic analysis of shared enjoyment in eca interactions of children with autism," in *Proc. of ICASSP'12*, 2012, pp. 4485–4488.

[51] S. P. Whiteside, R. A. Varley, T. Phillips, and H. Garety, "The production of real and non-words in adult stutterers and non-stutterers: an acoustic study," in *Proc. of ICSLP'00*, vol. 2, 2000, pp. 515–518.

[52] S. Cho, M. Liberman, N. Ryant, M. Cola, R. T. Schultz, and J. Parish-Morris, "Automatic detection of autism spectrum disorder in children using acoustic and text features from brief natural conversations," in *Interspeech 2019*, 2019, pp. 2513–2517.

[53] J. Casebeer, H. Sarker, M. Dhuliawala, N. Fay, M. Pietrowicz, and A. Das, "Verbal protest recognition in children with autism," in *Proc. of ICASSP'18*, 2018, pp. 301–305.

[54] C. Lea, V. Mitra, A. Joshi, S. Kajarekar, and J. P. Bigham, "Sep-28k: A dataset for stuttering event detection from podcasts with people who stutter," in *Proc. of ICASSP'21*, 2021, pp. 6798–6802.

[55] C. Sridhar and S. Wu, "J-j-j-just Stutter: Benchmarking Whisper's Performance Disparities on Different Stuttering Patterns," in *Interspeech 2025*. ISCA, Aug. 2025, pp. 3753–3757.

[56] M. Corrales-Astorgano, D. Escudero-Mancebo, C. Gonzalez-Ferreras, Y. Gutierrez-Gonzalez, V. Flores-Lucas, V. Cardenoso-Payo, and L. Aguilar-Cuevas, "The Magic Stone: A Video Game to Improve Communication Skills of People with Intellectual Disabilities," in *Interspeech 2016*. ISCA, 2016.

[57] A. Valente, R. Marew, H. Toyin, H. Al-Ali, A. Bohnen, I. Becerra, E. Soares, G. Leal, and H. Aldarmaki, "Clinical Annotations for Automatic Stuttering Severity Assessment," in *Interspeech 2025*. ISCA, Aug. 2025, pp. 4318–4322.

[58] D. Bone, T. Chaspari, K. Audkhasi, J. Gibson, A. Tsiartas, M. V. Segbroeck, M. Li, S. Lee, and S. Narayanan, "Classifying language-related developmental disorders from speech cues: The promise and the potential confounds," in *Interspeech 2013*. ISCA, Aug. 2013, pp. 182–186.

[59] O. P. Skljarov, "Content-independent duration model on categories of voice and unvoice segments," in *5th International Conference on Spoken Language Processing (ICSLP 1998)*. ISCA, Nov. 1998, pp. paper 1149–0.

[60] V. Narasinga, P. Kommagouni, S. Vanga, K. S. S. Motepalli, S. Akarsh C, P. Barche, and A. Vuppala, "Enhancing Stutter Detection using Long-Term Average Spectrum Values," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2025, pp. 1–5.

[61] K. S. S. Motepalli, V. Narasinga, H. Pathuri, H. Khan, S. Mahesh, A. K. Abraham, and A. K. Vuppala, "Stuttering detection application," in *Interspeech 2023*, 2023, pp. 680–681.

[62] T. Kourkounakis, A. Hajavi, and A. Etemad, "Detecting Multiple Speech Disfluencies Using a Deep Residual Network with Bidirectional Long Short-Term Memory," in *Proc. of ICASSP'20*, 2020, pp. 6089–6093.

[63] A. Smith and C. M. Weber, "How stuttering develops: The multifactorial dynamic pathways theory." *JSLHR*, vol. 60 9, pp. 2483–2505, 2017.

[64] E. Yairi and N. G. Ambrose, "Early Childhood Stuttering I: Persistency and Recovery Rates." *Journal of Speech, Language & Hearing Research*, vol. 42, no. 5, p. 1097, 1999.

[65] V. Sisskin, "Disfluency-affirming therapy for young people who stutter: Unpacking ableism in the therapy room," *Lang Speech Hear Serv Sch*, vol. 54, no. 1, pp. 114–119, 2023.

[66] R. Gong, H. Xue, L. Wang, X. Xu, Q. Li, L. Xie, H. Bu, S. Wu, J. Zhou, Y. Qin *et al.*, "As-70: A mandarin stuttered speech dataset for automatic speech recognition and stuttering event detection," in *Interspeech 2024*, 2024, pp. 5098–5102.

[67] J. Li, Q. Li, R. Gong, L. Wang, and S. Wu, "Our Collective Voices: The Social and Technical Values of a Grassroots Chinese Stuttered Speech Dataset," in *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, ser. FAccT '25. Association for Computing Machinery, 2025, pp. 2768–2783.

[68] J. Lazar, J. H. Feng, and H. Hochheiser, *Research Methods in Human-Computer Interaction*. Morgan Kaufmann, Apr. 2017.

[69] A. Birhane, W. Isaac, V. Prabhakaran, M. Diaz, M. C. Elish, I. Gabriel, and S. Mohamed, "Power to the people? opportunities and challenges for participatory ai," in *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 2022, pp. 1–8.

[70] J. Li, S. Wu, and G. Leshed, "Re-envisioning remote meetings: Co-designing inclusive and empowering videoconferencing with people who stutter," in *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, 2024, pp. 1926–1941.

[71] X. Tang, J. Li, and S. Wu, "Disability-first ai dataset annotation: Co-designing stuttered speech annotation guidelines with people who stutter," in *Proc. of CHI'26*, 2026. [Online]. Available: https://doi.org/10.1145/3772318.3790405

[72] J. Harvill, M. Hasegawa-Johnson, and C. D. Yoo, "Frame-level stutter detection," in *Interspeech 2022*, 2022, pp. 2843–2847.

[73] A. Batra, M. Narang, N. K. Sharma, and P. K. Das, "Boli: A dataset for understanding stuttering experience and analyzing stuttered speech," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2025, pp. 1–4.

[74] E. Leifsen, M.-T. Gustafsson, M. A. Guzmán-Gallegos, and A. Schilling-Vacaflor, "New mechanisms of participation in extractive governance: between technologies of governance and resistance work," pp. 1043–1057, 2017.

[75] O. O. Queerinai, A. Ovalle, A. Subramonian, A. Singh, C. Voelcker, D. J. Sutherland, D. Locatelli, E. Breznik, F. Klubicka, H. Yuan *et al.*, "Queer in ai: A case study in community-led participatory ai," in *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 2023, pp. 1882–1895.

[76] J. Li, P. Liu, R. Lietz, N. Tang, N. M. Su, and S. Wu, "Govern with, not for: Understanding the stuttering community's preferences and goals for speech ai data governance in the us and china," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, vol. 8, no. 2, 2025, pp. 1548–1560.

[77] J. Dinishak, "The deficit view and its critics," *Disability Studies Quarterly*, vol. 36, no. 4, 2022.

[78] F. A. K. Campbell, "Exploring internalized ableism using critical race theory," *Disability & society*, vol. 23, no. 2, pp. 151–162, 2008.

[79] A. Pyszkowska and M. M. Stojek, "Early maladaptive schemas and self-stigma in people with physical disabilities: The role of self-compassion and psychological flexibility," *International journal of environmental research and public health*, vol. 19, no. 17, p. 10854, 2022.