Translation Tutorial: Inclusive and Fair Automatic Speech Recognition with People Who Stutter

Shaomei Wu[1], J. Scott Yaruss[2], Rong Gong[5], Ann Marie Ryan[2], Hope Gerlach-Houck[3], Caryn Herring[4], and Nihar R. Mahapatra[2]
[1]AImpower.org [2] Michigan State University [3] Western Michigan University
[4] Friends: The National Association for Young People Who Stutter [5]StammerTalk.net
shaomei@aimpower.org, jsy@msu.edu, rong.gong@stammertalk.net, ryanan@msu.edu, hope.gerlach@wmich.edu, caryn@friendswhostutter.org, nrm@egr.msu.edu

Team

The tutorial is prepared by an interdisciplinary team of speech language pathologists (SLPs), AI researchers, and people who stutter. The tutorial will be presented by the following team members:

J. Scott Yaruss is a professor of communicative sciences and disorders at Michigan State University. A board-certified SLP and specialist in stuttering, Dr. Yaruss has published more than 125 peer-reviewed papers, as well as more than 350 other publications, on stuttering. He has also given roughly 1000 presentations (invited and refereed) about stuttering and stuttering therapy.

Jia Bin is a person who stutters and a PhD student in Communicative Sciences and Disorders at Michigan State University. Jia received her M.A. in Secondary Education and Speech Language Pathology. Jia is the co-founder of StammerTalk, board member of World Stuttering Network, National Stuttering Association chapter leader, and vice president of the Spartan Stuttering Group.

Shaomei Wu is a person who stutters and the founder and CEO of AImpower.org. Previously, she was a staff research scientist at Meta AI. Shaomei frequently speaks and organizes academia conferences on the topics of accessibility and human-centered AI.

Nihar R. Mahapatra is a professor of electrical and computer engineering at Michigan State University. He is a principal investigator of an NSF grant on accessible and fair voice-activated AI systems for people who stutter. He has conducted research, published widely, and presented on a number of AI topics including AI fairness in recruitment and AI and the future of work.

Tutorial Description and Impact Statement

Introduction

Stuttering affects approximately 1% of the population worldwide [3]. The condition is typically characterized by the speech behaviors that people who stutter (PWS) may exhibit, such as repetitions ("li-li-like this"), prolongations ("lllllike this"), and blocks ("l—ike this"). Beyond these observable "speech disfluencies," many PWS also experience significant reductions in their quality of life due to the communication challenges that they face everyday [25]. These difficulties are compounded by the responses of listeners, who may react negatively to stuttered speech. In fact, ample research has demonstrated that PWS regularly face negative listener reactions, such as stigma [5] and discrimination [6], which can affect all aspects of life, including socializing with others [4], achieving educational goals [7], and pursuing employment opportunities [8].

One increasingly common "listener" for people who stutter (and all speakers) is the AI-powered, voice-activated automated speech recognition (ASR) system [2]. With 90% of the US adults familiar with voice-enabled products and over 60% using them heavily, ASR systems have become a ubiquitous part of today's communication environment, increasingly mediating our interactions with people, organizations, and the environment [19]. Despite the popularity and benefits of AI-powered ASRs for many people, research has shown that ASRs have significant difficulty in decoding stuttered speech, resulting in three to four times higher word error rate (WER) compared to non-stuttered speech [15]. In practice, ASRs may misinterpret the speech of PWS, cut them off while they are speaking, or simply be unable to provide valid responses. The inability of ASR systems to work with stuttered speech can exacerbate the communication difficulties that PWS experience when interacting with smart speakers, in-car navigation systems, AI-answered service phone lines, and automated video interviews. Therefore, **the increasing use of ASRs without considering speech diversity further disadvantages PWS, creating structural barriers for them to fully participate in all aspects of life.**

Impact Statement

While it is imperative to develop fair ASR systems for people who stutter, we see two significant challenges with the current technical approach. First,the characteristics of stuttered speech, with its irregularities and unique temporal pattern, defy many existing assumptions about human speech that were built-in into current ASRs. Second, while most existing ASR systems are designed to facilitate information exchange, the subjective experience of speaking is under-supported, demanding a paradigm change in the design and evaluation of ASRs that is more emphatic and less transactional. **The purpose of this tutorial is to provide necessary background information about stuttering and the needs of people who stutter, to drive research and engineering efforts to improve automatic recognition of stuttered speech and therefore increase access, fairness, and equity for people who stutter when they interact with ASR systems.**

Tutorial Description

The tutorial will begin with (1) an **interactive activity** that demonstrates the challenges of ASRs with stuttered speech, and then provide (2) an overview of the **typical characteristics of stuttered speech**, followed by (3) a summary of **helpful strategies** that listeners–and AI systems–can use to better understand and respond to people who stutter. The tutorial will also describe (4) **current technologies** for improving the understanding of stuttered speech by ASRs. Finally, we will engage the participants in (5) **group working sessions** to develop a stakeholder-driven framework for inclusive and fair ASRs, and brainstorm strategies and next steps. Additional details about these components are provided below.

1. Interactive Activity: Experience Demo

As we anticipate that the majority of our participants would be speech technology developers without prior experience with stuttering, we will start the tutorial by sharing lived experience of PWS with ASRs to set the context and build empathy. In particular, we will play recorded speech of PWS in real-world scenarios (e.g. phone calls) and have both the participants and popular ASR systems transcribe the speech. We will compare the transcriptions to better understand the capacities and limits of current ASRs.

2. Characteristics of Stuttered Speech

Research over many years has sought to define and describe the overt speech behaviors that characterize stuttered speech (see review in [26]). Various categorization schemes (e.g. [12, 18]) have been developed, based on how listeners perceive interruptions in the flow of speech. While existing schemes typically represent stuttering based on seemingly distinct categories, in reality, disfluency types often overlap (so-called "clustered" disfluencies; see [21]) or occur in patterns that are not easily described based on canonical forms [10]. In order for ASRs to be able to appropriately identify and account for stuttered speech disruptions, it is necessary to ensure that the types of disfluencies and their myriad complications are understood [11]. This tutorial aims to provide a clear description of stuttered speech characteristics in support of this effort.

3. Helpful Strategies for Listeners

Research has identified many ways that listeners can improve their ability to understand stuttered speech, such as allowing speakers enough time to speak and active listening. These strategies can benefit ASRs as well. For example, people who stutter frequently report being cut off by the arbitrary time-limits of the response period allowed by the ASR, and find ASRs attempt to guess the words without fully listening to their messages. To address these issues, ASRs could allow a variable and adaptive response time that is consistent with the needs of an individual speaker who stutters, and allows the speaker to express their entire message. These adjustments to current ASRs can not only increase the accuracy for decoding stuttered speech, but also show respect to speakers who stutter.

4. Existing Technical Efforts

Current work on improving ASRs for stuttered speech has primarily focused on: 1) collecting and curating datasets of stuttered speech; 2) tuning existing ASR models to better detect and recognize stuttering speech. In our tutorial, we will first review existing datasets of stuttered speech (e.g. FluencyBank [20], UCLASS [9], SEP-28K [15], and the LibriStutter dataset with synthesized stutters [14]) and their pros and cons [15]. We will also share current techniques for better detecting stuttering events [1, 13, 14, 15, 22] and transcribing stuttered speech [1, 11, 16, 17, 23]. We will discuss the challenges with current approaches, and the gap in measuring and improving stuttering users' experience of ASRs in real world scenarios [2]. We will advocate for the epistemic shift to centering the subjective experience of stuttering in the development of ASR systems.

5. Group Working Sessions

We will lead working sessions with all participants and breakout groups to develop frameworks and roadmaps for inclusive and fair ASRs.

1) Develop framework for inclusive and fair ASRs

We will hold a structured, interactive session that involves participants in the analysis of the full AI life cycle in the context of ASRs, including ASR design, development, deployment, and evaluation. We will consider both technical and social attributes of inclusive and fair ASRs to develop guiding principles, guidelines, success criteria, conformance standards, and metrics and benchmarks.

2) Brainstorm solutions and roadmap

We will divide the participants into breakout groups (with at least one person who stutters in each group) to brainstorm solutions and next steps to improve the experience of PWS with ASRs. After sharing and discussing the ideas with the big group, we will present a set of design recommendations and a development roadmap for inclusive and fair ASRs.

Timeline (Proposed Length: 90 mins)

1. (5 mins) Interactive: demo of ASRs
2. (20 mins) Characteristics of Stuttered Speech
3. (20 mins) Helpful Strategies for Listeners and AIs
4. (20 mins) Current Technical Work on Improving ASR for Stuttered Speech
5. (20 mins) Group Working Sessions
   - ➢ (10 mins) Stakeholder framework development
   - ➢ (10 mins) Brainstorm in breakout groups
6. (5 mins) Reconvene and conclude

Acknowledgment: This material is partly based upon work supported by the NSF Award No. 2235916.

References

[1] Sadeen Alharbi, Anthony J H Simons, Shelagh Brumfitt, and Phil Green. 2017. Automatic recognition of children's read speech for stuttering application. In Proc. 6th Workshop on Child Computer Interaction (WOCCI 2017). 1–6. https: //doi.org/10.21437/WOCCI.2017- 1

[2] 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.

[3] O. Bloodstein, N.B. Ratner, and S.B. Brundage. 2021. *A handbook on stuttering* (7th ed.). Plural Publishing.

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

[5] Michael P. Boyle 2018. Enacted stigma and felt stigma experienced by adults who stutter. *Journal of Communication Disorders* 73 (2018), 50-61.

[6] Geoffrey A. Coalson, Alexus Crawford, Shanley B. Treleaven, Courtney T. Byrd, Lauren Davis, Lillian Dang, Jillian Edgerly, Alison Turk. 2022. Microaggression and the adult stuttering experience. *Journal of Communication Disorders* 95 (2022), 106180.

[7] Isobel Crichton-Smith. 2002. Communicating in the real world: accounts from people who stammer. *Journal of Fluency Disorders* 27, 4 (2002), 333-352.

[8] Hope Gerlach, Evan Totty, Anu Subramanian, and Patricia Zebrowski. 2018. Stuttering and Labor Market Outcomes in the United States. *J Speech Lang Hear Res.* 61, 7 (2018), 1649–1663.

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

[10] Peter A. Heeman and Andy McMillin and J. Scott Yaruss. 2006. An annotation scheme for complex disfluencies. *INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing, INTERSPEECH 2006 - ICSLP*, *3*(1), 1081–1084. https://doi.org/10.21437/interspeech.2006-55

[11] Peter A. Heeman and Rebecca Lunsford and Andy McMillin and J. Scott Yaruss. 2016. Using clinician annotations to improve automatic speech recognition of stuttered speech. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *08-12-Sept*, 2651–2655. https://doi.org/10.21437/Interspeech.2016-1388

[12] Wendall Johnson & Associates. (1959). *The onset of stuttering*. University of Minnesota Press.

[13] Tedd Kourkounakis, Amirhossein Hajavi, and Ali Etemad. 2019. Detecting Mul- tiple Speech Disfluencies Using a Deep Residual Network with Bidirectional Long Short-Term Memory. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2019), 6089–6093.

[14] 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.

[15] 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

[16] Robert L. MacDonald, Pan-Pan Jiang, Julie Cattiau, Rus Heywood, Richard Cave, Katie Seaver, Marilyn A. Ladewig, Jimmy Tobin, Michael P. Brenner, Philip C. Nelson, Jordan R. Green, and Katrin Tomanek. 2021. Disordered Speech Data Collection: Lessons Learned at 1 Million Utterances from Project Euphonia. In Proc. Interspeech 2021. 4833–4837. https://doi.org/10.21437/Interspeech.2021- 697

[17] 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

[18] Ann Packman, and Mark Onslow. 1998. The behavioral data language of stuttering. In A. Cordes & R. J. Ingham (Eds.), *Treatment efficacy for stuttering: A search for empirical bases* (pp. 27–50). Singular.

[19] PwC. 2018.. Consumer Intelligence Series: Prepare for the voice revolution. Retrieved at pwc.com/cisvoiceassistants.

[20] 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.

[21] Michael P. Robb, Ainsley Sargent, and Greg A. O'Beirne. 2009. Characteristics of disfluency clusters in adults who stutter. *Logopedics Phoniatrics Vocology*, *34*(1), 36–42. https://doi.org/10.1080/14015430902748519

[22] Shakeel Ahmad Sheikh, Md. Sahidullah, Fabrice Hirsch, and Slim Ouni. 2021. StutterNet: Stuttering Detection Using Time Delay Neural Network. 2021 29th European Signal Processing Conference (EUSIPCO) (2021), 426–430.

[23] Olabanji Shonibare, Xiaosu Tong, and Venkatesh Ravichandran. 2022. Enhancing ASR for Stuttered Speech with Limited Data Using Detect and Pass. ArXiv abs/2202.05396 (2022).

[24] Kathryn Teesson, Ann Packman, and Mark Onslow. 2003. The Lidcombe Behavioral Data Language of stuttering. *Journal of Speech, Language, and Hearing Research*, *46*(4), 1009–1015. https://doi.org/10.1044/1092-4388(2003/078)

[25] Seth E. Tichenor and J. Scott Yaruss. 2019. Stuttering as defined by adults who stutter. *Journal of Speech, Language, and Hearing Research*, *62*(12), 4356–4369. https://doi.org/10.1044/2019_JSLHR-19-00137

[26] J. Scott Yaruss. 1997. Clinical Measurement of Stuttering Behaviors. *Contemporary Issues in Communication Science and Disorders*, *24*(Spring), 27–38. https://doi.org/10.1044/cicsd_24_s_27