5-2023

Speech Intelligibility Assessment: Predicting “Noncompliant” Listener Behavior

Briggs Kroff
Utah State University

SPEECH INTELLIGIBILITY ASSESSMENT: PREDICTING "NONCOMPLIANT" LISTENER BEHAVIOR

by

Briggs Kroff

Capstone submitted in partial fulfillment of the requirements for graduation with

University Honors

with a major in
Communicative Disorders and Deaf Education

in the Department of Communicative Disorders and Deaf Education

Approved:

Capstone Mentor Dr. Annalise Fletcher	**Departmental Honors Advisor** Dr. Brittan Barker

University Honors Program Executive Director
Dr. Kristine Miller

UTAH STATE UNIVERSITY
Logan, UT

Spring 2023

Abstract

Purpose: When researching speech intelligibility among people with dysarthria, convenience sampling has typically been used to recruit listeners. A new online crowdsourcing method, Mechanical Turk (MTurk), results in ecologically valid results, but outlier results are often removed from the analysis and considered "noncompliant". This study aims to examine whether there is a relationship between age, gender, speech/language/hearing impairment, and whether someone is "noncompliant".

Methods: 16 speakers, both with and without dysarthria, were recorded while they read prewritten sentences. Research participants found through MTurk then listened to the sentences and transcribed them. They also were asked questions including their age, gender, and if they have a speech/language/hearing impairment.

Results: There was no correlation found between either age or gender and if the participant was "noncompliant". However, there was a relationship that suggests participants with communication disorders were more likely to be "noncompliant".

Conclusion: The results of this study suggest that people with communication disorders are more likely to be removed from studies involving people with dysarthria. This could affect the ecological validity of the results since many people with dysarthria often interact with people with communication disorders in their daily lives.

Acknowledgments

First and foremost, I would like to thank my faculty mentor, Dr. Annalise Fletcher. She has given me kind feedback and helpful advice from the very beginning and has been flexible along the way as plans changed and things needed to be adjusted. Her excitement for the field, and research in general, has inspired me and helped me find a love for research that I never thought I would have.

I would also like to thank the Honors Program, all the professors I've had, and faculty I've worked with during my time at Utah State University. Without Honors, I would never have pushed myself out of my comfort zone as much as I have, and I wouldn't be leaving college with the same relationships and connections I have now. There are many faculty in the Communication Disorders department that have inspired me to pursue this major and to be a better person. They are knowledgeable, kind, and incredible at what they do.

Finally, I would like to thank my friends and family who have been there for me long before college. My parents have always believed in my ability to accomplish anything I put my mind to, and whether it's a high school play or college graduation, they are always there for me. My partner Natalie has been there from the start of this project to the end of it and has sat with me through many late homework nights. And lastly, thank you to my close friends for being the comic and social relief I needed when my workload sometimes felt like too much.

Table of Contents

Acknowledgements……i

Introduction……1

Methods……3

Listeners and Speech Stimuli……3

Procedure……4

Transcription Scoring……5

Defining "Noncompliant" Listeners……..6

Statistical Analysis……7

Results……....7

Discussion……10

Reflection……..14

Bibliography……..18

Appendix A……21

Author Bio……22

Table of Figures

Figure 1: *Frequency of Average Transcription Accuracy Across All Listeners*……..7

Figure 2: *Relationship Between Participants' Gender and Compliance Status*……..8

Figure 3: *Relationship Between Participants' Age and Compliance Status*……9

Figure 4: *Relationship Between Speech, Hearing, & Language Status and Compliance Status*..10

Introduction

Parkinson's Disease (PD) is a neurodegenerative disease that often results in motor symptoms such as a tremor, stiffness, and slower movements. Most commonly, PD onset begins between the ages of 55 to 65, but it can be developed as early as 40 years old. Due to its neurodegenerative nature, PD symptoms get worse over time, and the disease usually lasts about 10 years (Moya-Gale & Levy, 2019). Hypokinetic dysarthria is a speech impairment commonly associated with PD. As the result of damage to the basal ganglia, hypokinetic dysarthria can affect a speaker's loudness, voice quality, articulation, or fluency (Darley et al., 1968).

An important aspect of effective communication is the listener's ability to understand speech, which is called speech intelligibility. There are many different methods to assess a person's intelligibility, but conventionally the measurement involves counting how many words (or speech units) can be correctly understood by a listener (Kent et al., 1989). In the past, it has been common to use similar groups of people as listeners in intelligibility research, such as professionals in the field or groups of university students. However, some researchers have challenged the ecological validity of these methods (Lansford et al., 2016; Sescleifer et al., 2018).

Mechanical Turk (MTurk) is an example of crowdsourcing, a method that finds a broader group of people to participate in research tasks, who are more representative of the general population. People of varying demographics across the country sign up on MTurk with an account and are offered opportunities to earn money in exchange for small tasks. This may be more ecologically valid when assessing intelligibility in people with dysarthria because speakers with dysarthria typically interact with listeners of a variety of ages and education levels throughout their communities (Sescleifer, et al., 2018). Additionally, projects done using

crowdsourcing applications like MTurk have found similar overall intelligibility levels to more traditional methods (e.g., data collected in labs from college students). For example, an experiment was done that compared results between participants found through MTurk and laboratory participants after training was completed to improve their understanding people with dysarthria. The study concluded that both groups understood the same percentage of words spoken by someone with dysarthria (Lansford et al., 2016).

Although there are many benefits of crowdsourcing, there are problems with this method as well. It can be difficult to regulate the listeners' environment such as how loud it is and how focused they are on the task. There is also no way to make sure people's responses are not influenced by hearing, language, motor or cognitive difficulties. This makes it common when using convenience sampling techniques like MTurk to have a higher volume of outlier responses from participants.

Most studies that involve crowdsourced listeners will do some form of screening to remove these outlying listener responses (Sescleifer, et al., 2018). This is usually done with the rationale that "noncompliant" listeners need to be excluded. For example, studies may remove responses from participants who have a particularly low accuracy rate or people who leave their responses blank (Lansford, et al., 2016; McAllister Byun, et al., 2015; McAllister Byun, et al., 2016). However, outlying or atypical responses may also provide unique insights into the challenges faced by a more diverse listener population. People with dysarthria due to Parkinson's Disease are typically older, as well as people with communication disorders in general. Therefore, it is more likely that speakers with dysarthria will encounter other people with communication impairments in their daily lives (McAuliffe, et al., 2019). Analyzing these

participants' results can give us valuable information about the effectiveness of speech treatment methods for communication partners who have a communication disorder.

This study will examine the correlation between participants deemed "non-compliant" and different characteristics, namely gender, age, and hearing or speech impairments. We will answer the question of whether any of these characteristics are correlated with the appearance of "noncompliance" in participants. It is hoped that this will give us greater insight into the varying reasons for outlying listener responses and potential insight into the unique challenges faced by listeners with self-disclosed disabilities.

Methods

Institutional Review Board approval for this study was obtained through Utah State University.

Speech Stimuli

Speech stimuli was elicited from eight native speakers of American English with a diagnosis of Parkinson's disease and hypokinetic dysarthria (mean = 70.5 years, SD= 8.59). All speakers completed the Montreal Cognitive Assessment (Nasreddine et al., 2005), a cognitive screening assessment, and received scores of 20 or above (mean = 24.3, SD = 1.98). Eight age and sex matched native speakers of American English were also recruited as neurotypical control speakers (mean = 67.3, SD = 8.56). The control speakers reported no history of speech, language, or hearing impairments. Speakers with dysarthria and control speakers were prompted to read the Caterpillar Passage (Appendix A) which is an easily readable passage containing 16 sentences of varying complexity (Patel et al., 2013). Speakers read the passage one sentence at a time, in their everyday speaking voice. Digital recordings of the sentence productions were made

via a lavalier cardioid microphone positioned approximately 20 cm from the speakers' mouth with a sampling rate of 48 kHz and 16 bits of quantization.

Listeners

Listeners were recruited via Amazon Mechanical Turk (MTurk), an online platform for recruiting workers to complete small jobs. Participation was geographically restricted on MTurk, so only workers with U.S.-based IP addresses were able to view and participate in the experiment. Participants were all over 18 years old and were required to have a prior task approval rating of 98% or better, and approval from a minimum of 500 tasks. These criteria were chosen to increase the quality of responses, as the task was only available to workers who historically adhered to task instructions. Workers all received $3 in exchange for their participation.

Procedure

The experiment was completed remotely by the listeners on their personal device. Details of the study were posted to MTurk outlining the time commitment (20 minutes) and basic tasks the participants would be asked to complete. Interested participants were then instructed to access the experiment, hosted on a secure, university-based web server, via an embedded link in the task description. After clicking the link embedded in the task description, participants reviewed a consent form and indicated their consent by clicking the "Agree" button on the screen. Following consent, participants completed a demographic survey that contained the following questions:

1) Gender
2) Current Age (Years)
3) Ethnicity

4) Race
5) Native Language
6) Do you currently have, or have you ever had, a speech or language impairment?
7) Do you currently have, or have you ever had, a hearing impairment?
8) Do you have significant experience communicating with people with speech disorders (for example, apraxia, dysarthria, or stuttering)?

Following the survey, participants listened to 16 different sentences by the same speaker. They were instructed to type the sentence they heard from the speaker; if they could not understand anything from the sentence they were instructed to type "X". The instructions specified that their computer volume should be on, and the participant should be wearing headphones to complete the activity. As part of a larger study, after listening to the sentences, they were instructed to rate the speaker on factors of likeability, intelligence, and confidence. Even though this data was collected, it was not analyzed for our study since it was not relevant to the research questions.

Transcription Scoring

Transcription accuracy was automatically calculated using the open-source tool, Autoscore, which reliability compares the word spoken by the talker to the word that was transcribed by the listener (http://autoscore.usu.edu/; Borrie et al., 2019). Responses were scored on a sentence-by-sentence basis. The highest possible score was equal to the number of words in the sentence, and for every incorrectly transcribed word, the score was reduced by one. For example, if the sentence was "Do you like amusement parks?" and the participant wrote "Do you like news in a parks?" they would receive a score of 4/5, or 80%. So, listener responses were considered 100% correct when they were identical to the word spoken by the talker. Autoscore

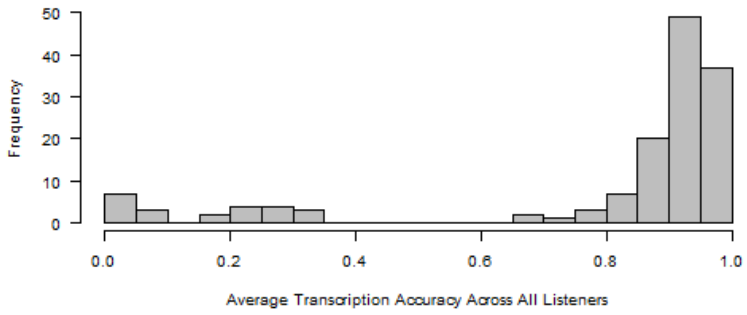
scoring rules including the tense and plural rules (see Borrie et al., 2019 for rule descriptions). Following the use of Autoscore, a research assistant manually screened the files for any common spelling errors or homonyms that were not in the default acceptable spell list (e.g., tic vs. tick). If an incorrect word was a homonym or was clearly attributable to a spelling error, the response was manually re-coded as correct.

Defining "Noncompliant Listeners"

After completing the transcription scoring, we examined the percentage of words each listener correctly transcribed. This information is visualized in Figure 1. Out of 142 listeners, 119 of them were able to accurately transcribe an average of more than 70% of words, regardless of whether they were listening to the control speakers or the speakers with dysarthria. Given that most of the speakers with hypokinetic dysarthria had relatively mild dysarthria, average scores above 70% were expected in this study. Not a single listener transcribed between 35% and 70% of words correct. However, 23 listeners correctly transcribed fewer than 35% of words. Since these were outlying listener results, we defined this group as "noncompliant" with the listening task and these were the participants who became the subject of analysis for this project. The behavior and responses of these participants is considered further in the results section.

Figure 1.

Frequency of Average Transcription Accuracy Across All Listeners



Statistical Analysis

Data were grouped into the different demographic groups based on their responses to the study survey questions. Groups included age, gender, ethnicity, race, native language, and whether they have a speech-language or hearing impairment. Then, a series of Fisher's exact tests were used to analyze if there was a statistically significant difference between the number of "noncompliant listeners" with self-disclosed speech and hearing disorders as compared to other participants who did not have hearing and/or speech and language impairments. Another Fisher's exact test was used to analyze the relationship between gender and compliance status, and a linear regression was performed to analyze the relationship between age and compliance.

Results

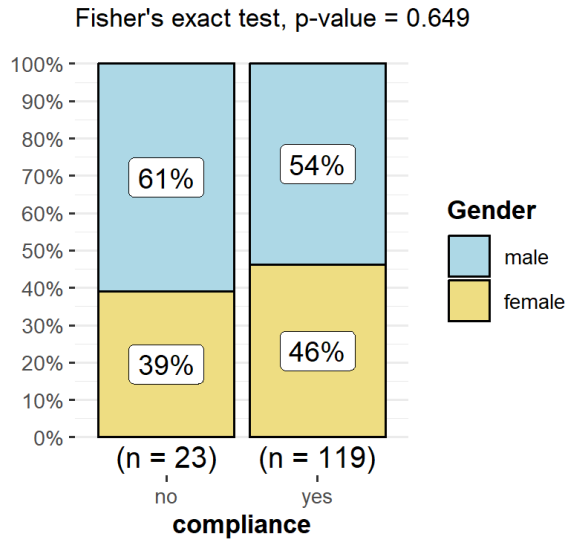
Of all the participants, 23 of them qualified as "noncompliant". As defined earlier, noncompliance was based on an overall transcription accuracy of less than 35%. There were several different listener behaviors that contributed to low accuracy. They included skipping trials without entering any information, and/or only writing one or two words in response to each sentence. Other behaviors included typing "X" for the majority of trials, describing the speaker in terms of likability or intelligence (i.e. "this person is nice" or "not intelligent") instead of transcribing their speech, or typing a number in the entry box instead of a sentence. Participants were instructed to type "X" in the entry box if they were unable to understand any of the words the speaker was saying, so this behavior did not necessarily indicate a failure to follow directions. However, typing X across the majority of trials did cause some participants to be classified as "noncompliant".

Did gender affect noncompliance?

A Fisher's exact test was used to determine if there was a significant association between the gender of the listener and their compliance status (see Figure 2). The test revealed that there was not a statistically significant association between gender and noncompliance (two-tailed, p=0.649).

Figure 2

Relationship Between Participants' Gender and Compliance Status

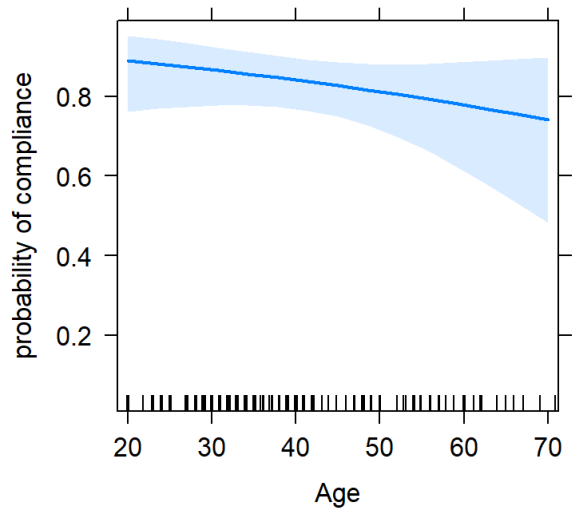


Did age affect noncompliance?

A logistic regression was used to analyze the relationship between the listeners' age and compliance status (see Figure 3). The model revealed a nonsignificant effect for listener age (b = -0.021, SE = 0.019, p = 0.27). This effect size indicates that the odds of a listener being classified as noncompliant were 0.98 times lower when there was a one-year increase in listener age. This difference was not statistically significant.

Figure 3

Relationship Between Participants' Age and Compliance Status

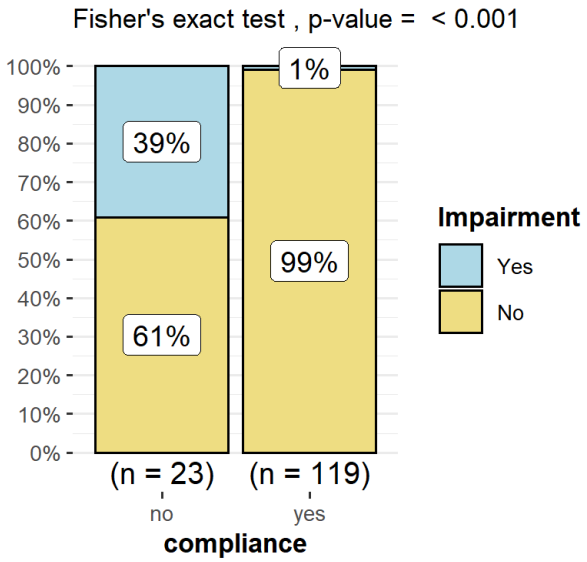


Do speech, language, and/or hearing impairments affect noncompliance?

Fisher's exact test was used to determine if there was a significant association between the occurrence of a self-reported speech, language, and/or hearing impairment and classification of participants as noncompliant (see Figure 4). In this dataset, there was a large overlap in participants who reported speech and language impairments and those who reported hearing impairments, so both groups were analyzed together. Results indicated a statistically significant association between these impairments and noncompliance (two-tailed, p=<0.001), indicating that people who reported communication impairments were more likely to be classified as noncompliant listeners.

Figure 4

Relationship Between Speech, Hearing, and Language Status and Compliance Status



Discussion

This study aimed to examine whether there was a relationship between listeners labelled as "noncompliant" and factors such as age, gender, and speech, language or hearing status. In this study, a listener was defined as "noncompliant" when their score fell below 35% transcription accuracy, which was considered an outlying result. Using this definition, there were a total of 23/142 participants who were labelled as noncompliant. In other words, 16% of the participants fell into the "noncompliant" category. Although the proportion of discarded data is not always presented in published research, this number seems consistent with prior studies. For example, Lansford et al., (2016) removed approximately 11% of otherwise eligible listeners for failing to provide input when asked to transcribe speech. Other Mechanical Turk studies that involve listener ratings have removed up to 20 or even 35% of participants for failing attention checks or missing data (e.g., McAllister Byun et al., 2015, McAllister Byun et al., 2016).

This research study showed that there was no correlation between a participant's age or gender and whether they were considered "noncompliant". This was not surprising, as we did not expect to find a correlation between any gender and a higher rate of noncompliance. Our

hypothesis is supported by a study demonstrating that the proportion of people with a 95% task success rate or higher on MTurk was equal between men and women (Peer, E. et al., 2014).

There was also no correlation between age and noncompliance, which was unsurprising. We speculated that there could possibly be a difference between younger and older groups of people's ability to use technology successfully, which could have resulted in more noncompliance among older people. We considered this because there are lower rates of technology use among people older than 65 years as well as a more negative perception of technology among these populations. (Charness & Boot, 2009). However, research has also demonstrated that generally there is not a big discrepancy between the task success rate of younger and older people using MTurk (Peer, E. et al., 2014).

There was a connection to participants with speech language, and hearing impairments and noncompliance. Thirty-nine percent of the noncompliant participants were people with communication difficulties, but only 1% of the compliant participants had these communication difficulties. So, it appears more likely that if someone had a speech, language, or hearing impairment their results would be considered an outlier and removed from data analysis.

These results could have implications for speech intelligibility studies involving people with communication difficulties, especially with research aimed at finding better methods of intervention and treatment. For example, online studies might measure how well listeners can transcribe or rate sentences read by a speaker with dysarthria before and after an intervention. If research studies are removing listener data that is low enough to be an outlier and considering these responses to be "noncompliant", it is likely that listeners with communication disorders are being disproportionately removed as part of that group. This means that the treatment gains assessed in these studies (e.g., the amount of intelligibility improvement experienced) may not be

applicable to listeners who experience their own communication difficulties (Nightingale, E. et al., 2020). AHSA has reported that 5-10% of all Americans have some kind of communication disorder (n.d.b). As people get older, they have a higher chance of developing Parkinson's Disease and communication disorders like dysarthria as a result. (n.d.a). It is reasonable to assume that people spend most of their time around people who are in a similar position in life, meaning that the older generations spend time around each other. If this is the case, senior adults with dysarthria and other communication disorders are spending time around other senior adults who might have similar communication disorders. The implications of removing a disproportionate number of participants with communication disorders are amplified because of this.

The use of technology in all aspects of life has increased exponentially over the last few decades, and research sampling is no different. Many different fields have started to use MTurk and trends show that the use of MTurk and other electronic crowdsourcing will continue to increase as we learn more about them (Chan & Holosko, 2016; Keith, et al., 2017). When these crowdsourcing methods are used, the sample sizes are more diverse than they are when convenience sampling is used, especially because convenience samples in a university setting tend to result in sample populations that contain only young adults. The diversity offered by MTurk can result in higher numbers of participants being removed, which means it is crucial to learn more about why they are being removed and if something needs to change.

One change that could be beneficial to the ecological validity of relevant research projects is to adjust research methods so that participants with communication disorders are included in data analysis. Not all "noncompliant" listeners should be included in the results because, like we saw in our study, only 39% of them had a speech, language, or hearing

impairment. However, those 39% may contribute valuable information on which intervention methods are effective in the real world.

To get more information about the "noncompliant" participant behavior, one possible solution could be to conduct follow-up interviews or surveys with those who were "noncompliant" and had self-reported communication disorders. This way you could ask them more specific questions about their experience to see if it reflects what their experience would be like in the real world. For example, if someone with a language disorder struggled to process and remember the sentence long enough to transcribe it, that could represent a real-life situation when two people with communication disorders are interacting with each other.

Future studies should also be mindful of their use of the word "compliant" when describing outlying data as it implies that listeners are simply unfocused or not motivated to complete the task. This study has demonstrated that there may be other reasons why someone has not completed a task correctly other than their willingness to follow directions.

Word count: 3,100 words

Reflective Writing

This capstone project was a challenging and valuable capstone to my undergraduate education. It was the biggest and longest project I've ever been a part of, and it required more follow through and careful organization than anything I have done previously. The fact that the capstone is so open ended required me to consider many different options and find a project that would reflect me, my education, and the things I care about. The topic I ended with is a great capstone project for my personal experience because it was largely aimed at investigating whether we should be finding more equitable ways to research evidence-based interventions for people with communication disorders. This involves various topics I am personally invested in like speech disorders and inclusion. Around the time in my junior year when I started thinking about my capstone project marked a shift in my undergraduate experience. I got a job with Dr. Fletcher as a research assistant and worked with her for the remainder of my time at USU. From that point forward, research became a focal point of my academic experience. Because of this, working on a final research project with Dr. Fletcher is a fitting capstone experience.

My capstone project also added to my general education and helped me develop skills that will support future goals in life. Since I was young, doing well in school has come easy to me and, as a result, I have always struggled to follow through on things that are difficult and take a lot of time and/or effort. A project of this magnitude took a tremendous amount of work and dedication, and I pushed myself to stick with it. In fact, when I was completing the capstone preparation course that Honors offers, I considered dropping Honors and not trying to do the capstone project at all because it sounded too overwhelming. Thankfully, I dealt with those anxieties head on and learned to use the resources that are there to help me. Getting through the

moments of overwhelm like that have given me a new level of confidence in myself and my ability to stay with something until it's finished, which is crucial to achieving goals in the future.

Although it was one of the most intimidating aspects of completing a capstone project in the beginning, the relationship I built with my faculty mentor has been one of the biggest takeaways of the project. It has given me a safe space to ask questions and explore research in the field I want to go in. Especially because I worked with Dr. Fletcher before I started my capstone project, I have been able to build a positive and strong connection with her that I wouldn't have been able to complete the project without. It was a wonderful experience to work with someone who was so patient with me as I was learning and was so flexible when we had to refocus or adjust our plans.

There were many ways this project deepened both my research experience and my critical thinking about the communication disorders field. I've spent the past few years learning about different communication disorders, technical terms, methods of intervention, and more, but this project introduced me to a different side of the discipline than what I was learning in class. I learned more about gathering data around speech intelligibility, processing the data, and using a computer program (specifically "R") to analyze different aspects of it. This project shifted my view to look at the big picture by considering the methods of current research in the field and ways that we could change/improve those methods. Especially while I prepared for and gave my poster presentations and the undergraduate research symposium and the capstone showcase, I had to consider not only the data we analyzed, but real solutions that could help with the problem we were addressing. Asking questions and recognizing problems is an important part of critical thinking, but I also believe that problem solving is just as important. I also think that effective

research requires critical thinking because you must be willing to ask questions and consider ideas that we don't have answers to yet, which can be intimidating sometimes.

Although my capstone project had a narrow focus on the communication disorders field, the implications can be carried into many different fields as research continues regarding this subject. For example, a sociological/anthropological approach could be taken when analyzing this data to see how the results extend into real-world society and how things have been done historically. Or disability studies could replicate similar studies considering different disabilities such as autism spectrum disorder or mental illnesses like depression and anxiety. Because our society is set up in a way to consider certain characteristics "normal" and the rest of people "outliers" it is interesting to think about this issue through different disciplines to see how that is really affecting people in their everyday lives. In a small way, I also learned more about statistical analysis when we were analyzing the data, including learning about linear regression tests, Fisher's exact test, etc. Statistics also came into play when creating figures and tables.

Although I didn't work directly with the local or global community when I was working on this project, it was done with the wellbeing of the community in mind. The goal was to analyze reasons that people may be being left out of research studies and see how this could be affecting results of said studies. The biggest aim was to consider how this could be affecting our research to find new interventions and treatments and if this could be affecting the ecological validity of those studies. This means that the discussion section was focused on local and global communities, especially of older people and people with communication disorders, and improving their quality of life through appropriate intervention.

Overall, my capstone project was one of the most valuable things I participated in during my undergraduate experience. It helped me gain confidence in my own ability to see difficult

things through to the end and opened my eyes to an aspect of the communication disorders field that I wouldn't have considered otherwise. I will always be grateful for the push that the Honors Program gave me to complete this project, and, despite the stressful times, it was a very positive experience.

Word count: 1,058 words

References

American Speech-Language-Hearing Association. (n.d.a). *The Aging Population.* https://www.asha.org/slp/clinical/aging/

American Speech-Language-Hearing Association (n.d.b). *Quick Facts about ASHA.* https://www.asha.org/about/press-room/quick-facts/#:~:text=5%25%20to%2010%25%20of%20Americans,children%20have%20noticeable%20speech%20disorders

Borrie, S. A., Barrett, T. S. & Yoho, S. E. (2019). Autoscore: An open-source automated tool for scoring listener perception of speech. *The Journal of the Acoustical Society of America, 145,* 392-399. https://doi.org/10.1121/1.5087276

Charness, N., & Boot, W. R. (2009). Aging and Information Technology Use: Potentials and Barriers. *Current Directions in Psychological Science, 18*(5), 253-258. https://psycnet.apa.org/doi/10.1111/j.1467-8721.2009.01647.x

Chan, C. & Holosko, M. J. (2016). An Overview of the Use of Mechanical Turk in Behavioral Sciences: Implications for Social Work. *Research on Social Work Practice, 26*(4), 441-448. https://doi.org/10.1177/1049731515594024

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential Diagnostic Patterns of Dysarthria. *Journal of Speech and Hearing Research, 12*(2), 246-269. https://doi.org/10.1044/jshr.1202.246

Keith, M. G., Tay, L. & Harms, P. D. (2017). Systems Perspective of Amazon Mechanical Turk for Organizational Research: Review and Recommendation. *frontiers in Psychology, 8:*1359. https://doi.org/10.3389/fpsyg.2017.01359

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek J. C. (1989). Toward Phonetic Intelligibility

Testing in Dysarthria. *Journal of Speech and Hearing Disorders, 54,* 482-299. https://doi.org/10.1044/jshd.5404.482

Lansford, K. L., Borrie, S. A., & Bystricky, L. (2016). Use of Crowdsourcing to Assess the Ecological Validity of Perceptual-Training Paradigms in Dysarthria. *American Journal of Speech-Language Pathology, 25,* 233-239. https://doi.org/10.1044/2015_AJSLP-15-0059

McAllister Byun, T., Halpin, P. F., & Szeredi, D. (2015). Online crowdsourcing for efficient rating of speech: A validation study. *Journal of Communication Disorders, 53,* 70-83. https://doi.org/10.1016/j.jcomdis.2014.11.003

McAllister Byun, T., Harel, D., Halpin, P. F., & Szeredi, D. (2016). Deriving gradient measures of child speech from crowdsourced ratings. *Journal of Communication Disorders, 64,* 91-102. https://doi.org/10.1016/j.jcomdis.2016.07.001

McAuliffe, M. J., Schluter, P. J., & Jamieson, H. A. (2019). An epidemiological profile of communication disability among older adults with complex needs: A national cross-sectional study. International journal of speech-language pathology, 21(6), 537-546. https://doi.org/10.1080/17549507.2018.1495763

Moya-Gale, G. & Levy, E. S. (2019). Parkinson's disease-associated dysarthria: prevalence, impact, and management strategies. *Research and Reviews in Parkinsonism, 9,* 9-16. http://doi.org/10.2147/JPRLS.S168090

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool for Mild Cognitive Impairment. *Journal of the American Geriatrics Society, 53*(4), 695-699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Nightingale, C., Swartz, M., Ramig, L. O., & McAllister, T. (2020). Using Crowdsourced

Listeners' Ratings to Measure Speech Changes in Hypokinetic Dysarthria: A Proof-of-Concept Study. American Journal of Speech-Language Pathology, 29, 873-882. https://doi.org/10.23641/asha.12170112

Patel, R., Connaghan, K., Franco, D., Edsall, E., Forgit, D., Olsen, L., Ramage, L., Tyler, E., & Russell, S. (2013). "The Caterpillar": A novel reading passage for assessment of motor speech disorders. *American Journal of Speech-Language Therapy, 22,* 1-9. https://doi.org/10.1044/1058-0360(2012/11-0134)

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior research methods*, *46*, 1023-1031. https://doi.org/10.3758/s13428-013-0434-y

Sescleifer, A. M., Francoisse, C. A., & Lin, A. Y. (2018). Systematic Review: Online Crowdsourcing to Assess Perceptual Speech Outcomes. *Journal of Surgical Research, 232,* 351-364. https://doi.org/10.1016/j.jss.2018.06.032

Appendix A

The Caterpillar Passage

"The Caterpillar"

Do you like amusement parks? Well, I sure do. To amuse myself, I went twice last spring. My most MEMORABLE moment was riding on the Caterpillar, which is a gigantic rollercoaster high above the ground. When I saw how high the Caterpillar rose into the bright blue sky, I knew it was for me. After waiting in line for thirty minutes, I made it to the front where the man measured my height to see if I was tall enough. I gave the man my coins, asked for change, and jumped on the cart. Tick, tick, tick, the Caterpillar climbed slowly up the tracks. It went SO high I could see the parking lot. Boy, was I SCARED! I thought to myself, "There's no turning back now." People were so scared they screamed as we swiftly zoomed fast, fast, and faster along the tracks. As quickly as it started, the Caterpillar came to a stop. Unfortunately, it was time to pack the car and drive home. That night I dreamt of the wild ride on the Caterpillar. Taking a trip to the amusement park and riding on the Caterpillar was my MOST memorable moment ever!

Author Biography

Briggs will be graduating from Utah State University in May of 2023 with a major in communicative disorders and a minor in religious studies. They have enjoyed their involvement as an office assistant for the Honors Program, a research assistant in the Motor Speech Disorders Lab, and a member of the Honors Student Advisory Board representing the College of Education and Human Services. After graduating, Briggs hopes to complete a graduate degree in audiology and continue to be involved in research in the field. In their free time, Briggs enjoys playing tennis, going on walks, and writing and performing music.