Human Factors in Accessibility and Assistive Technology, Vol. 37, 2022, 125–138
https://doi.org/10.54941/ahfe1001651

Speech Recognition Technology for Users with Apraxia: Integrative Review and Sentiment Analysis

Aimee Kendall Roundtree

Texas State University San Marcos, TX 78666, USA

ABSTRACT

This research reveals the need for more user experience and usability research on speech recognition technology for users with apraxia. First, it offers an integrative review of findings from 2009 to 2020. Only 9 of 120 provided sufficient detail about the 20% of the users diagnosed with apraxia of speech. The studies covered therapeutic rather than mundane settings. Second, the research gathers Twitter feedback about speech recognition products and speech impediments. Most of the 143 tweets were negative about the performance of speech recognition technologies. Future studies should include more homogeneous samples in terms of speech conditions and more heterogeneous samples in terms of demographics. Future studies should also gather more direct user feedback, compare technologies, and modify user experience and usability research methods.

Keywords: Human factors, User experience, Usability, Speech processing, Speech recognition, Apraxia of speech, User feedback, Twitter

INTRODUCTION

This integrative review examines the user experience and usability research on speech recognition software for users with apraxia. Apraxia is a disorder that, among other things, can make it hard to speak (Parker 2021). Apraxia is a broad neurological disorder that sometimes affects speech production, at which point it is referred to as speech apraxia. No good data exists concerning apraxia in different age groups or the prevalence of apraxia versus apraxia of speech (Chawla 2020). However, it is one of the most common speech disabilities in children. A recent study confirmed that one child in 1000 has childhood apraxia of speech (Shriberg et al. 2019). General apraxia is also estimated to be common in older age groups with high frequencies of stroke and dementia, leading to apraxia of speech. Apraxia of speech makes it difficult for people to make specific motor movements necessary for full speech articulation or the process of forming words. The nature of apraxia makes it an excellent site for technological intervention. Its prevalence is high among a growing number of children. It is also high among speech disorders acquired by older adults due to illness and aging. The relatively high prevalence of apraxia of speech and other speech disorders underscores the potential need to help improve the population's quality of life and socialization. Speech recognition and processing software might help with both. However, voice-controlled software and personal assistants can only enhance lives if they provide parity of user experience.

User experience research and usability research of speech recognition and voice-controlled technologies are critical, given the growing demand in the market. The global speech and voice recognition market is projected to rise at a compound annual growth rate of 19.8% to over $25 billion by 2026 (Fortune Business Insights 2020). As the market grows, so does the likelihood that these technologies will be leveraged and used by disability communities. Furthermore, disability communities might benefit from therapy and quality of life provided by new technologies. New technologies may make tasks and processes more accessible. Recent coverage has identified some limitations in speech and audio processing software. For example, the Washington Post reported that Alexa and Google Assistants are 30% less likely to understand non-American accents (Harwell 2018). Furthermore, Slate reported several cases of users with speech impediments for whom voice-controlled software failed (Corcoran 2018). Users with cerebral palsy and stuttering, for example, report experiencing great difficulty with Alexa, Google, and other popular personal assistant technologies working for them. The extent to which these pain points persist across communities of users with speech differences and impairments is a lingering question that remains unanswered by the literature. One study of voice-controlled personal assistants by people with disabilities reviewed several Amazon reviews of these technologies by self-reported people with disabilities ranging from visual impairment, motor or mobility impairment, speech impairment, cognitive impairment, and hearing loss (Pradhan et al. 2018). This study also found that almost 20% of all the reviews mentioned limitations, criticisms, or suggestions such as missing features, limited use, and affordability. A review of ambient assisted living products designed for users with disabilities, including apraxia, also enlisted user tests (Colomer et al. 2014). The study found that 67 participants could use the technology as expected, and ten sessions were discontinued due to system errors (Colomer et al. 2014). It is important to consider the ethics and usability of new software and technologies to serve better all communities who might benefit (Lazar et al. 2017, Roundtree 2021). These studies suggest that disabilities, particularly speech conditions, might negatively impact the user experience of voice-controlled technology.

This integrative review extends the research by providing an overview and synthesis of existing findings on this subject. It also analyzes user feedback posted on social media. In collecting data for the review, the study exposes major limitations in reporting and coverage of usability and UX in this disability community. The research question for the review was as follows: What does current research tell us about the user experience and usability of speech recognition software for people with apraxia of speech? What is the quality of reporting of this body of research? What does user feedback on Twitter tell us? What are future directions?

METHODS

This study employs integrative review methods. An integrative review involves using research databases to compile and synthesize systematically the literature published on a topic. The integrative review allows researchers to integrate qualitative and quantitative findings (Whittemore and Knafl 2005).

Study characteristics included peer-reviewed articles and excluded theses and books for this review. They included English language peer-reviewed articles published from 2009 to 2020. I used EBSCO, Web of Science, Scopus, ACM, IEEE, and Google Scholar to identify sources, and library holdings limited me by accessing the full-text versions of the articles. Search terms included the following: speech, voice, and audio recognition and control, apraxia, user experience, user test, usability, users, and participant (root words). Databases were limited to finding these words in the abstract and titles of the article to isolate those studies that would most likely contain pertinent information. Database previews and article abstracts were screened for eligibility.

Once eligible studies were screened, I read the entire article to extract information, including the number and demographics of participants, the methods used, and study findings. Articles that did not pertain and were not written in English were eliminated. Student assistants and I organized and synthesized details from the literature with a spreadsheet. Inductive coding analysis and consensus were used to handle and interpret data and combine the study results (Fereday et al. 2006). To analyze the impact and quality of the articles, I also used Web of Science and Google Scholar to find how many times each included article has been cited and to determine each article's h-index, which measures the productivity and citation impact of the publications.

To cross-reference user feedback, I also gathered Twitter content about apraxia and speech impediments about Siri, Bixby, Google Assistant, Alexa, and Cortana, speech recognition software deployed in everyday life. We used the names of the programs and the keywords apraxia and speech impediment to search Twitter Advanced Search. Each tweet was read and characterized as positive and negative using close reading. LIWC and R-Studio were used to analyze the content. Linguistic Inquiry and Word Count (LIWC) codes files on several dimensions (Dey et al. 2018). Tweets received a set of 87 scores, each indicating the percent of occurrence relative to the total number of words in the file. RStudio is an open-source R programming language with a graphic user interface that facilitates R libraries and coding access. The interface supports direct code execution and a variety of robust tools for plotting, viewing history, and debugging. The R packages OpenNLP, NLP, and CRAN Task View enabled calculations of n-gram and word frequencies. N-grams are contiguous sequences of n items from a given sample of text or speech; they can determine sentiment. NLP methods confirmed the close reading using topic modeling to confirm categories.

RESULTS

The keyword search yielded a total of 120 articles across all databases. Several articles were eliminated for the following reasons: Some were inaccessible (n=2). Some were in a language other than English (n=3). Some only mentioned apraxia rather than it being the focus of the study (n=39). Some only mentioned speech recognition software without it being the focus of the study (n=8). Some included study citations rather than original user experience, usability testing, or research (n=56). Two were repeats of other articles, and one was not peer-reviewed. Nine articles met all the inclusion criteria for this integrative review.

Study Participants

The articles included a total of 484 participants (median=20.5). See Table 1. The participants ranged from children to the elderly, and they had a variety of conditions, including apraxia and other speech conditions (n=109, 22.5%). One study mentioned apraxia as a target condition for the application but tested users with aphasia, which impairs linguistic capabilities differently than apraxia. Some forms of aphasia make constructing clear grammatical sentences difficult. These issues can also thwart speech recognition technologies' expected use and accuracy, but differently from apraxia. Neither apraxia nor aphasia was represented in the studies for their analyses. It was unclear whether the participants were diagnosed with apraxia in three studies. Only two studies mentioned and considered the level of disability (i.e., mild to severe). Four studies recruited children as participants. Three included therapists as a part of testing, and two included parents as a part of testing. Only one study considered participants with comorbidities and other issues that might impact user experience and usability. Five studies also did not report gender demographics, and no studies reported race or ethnicity demographics, which might affect user experience and usability.

Study Methods

Table 2 summarizes the study methods. Most studies tested some form of game or application for therapy (n=7). The exceptions were a study of user attitudes about a voice-controlled personal assistant and another about a rehab system for assisted living facilities. Two articles used formative and summative or pre-and post-tests. Two involved user engagement for more than one session—one asked the users to engage with the software for two weeks and another for a month. Most of the studies were typical usability tests, including some form of task scenario, product engagement, or observation during product use. One article was a retrospective study that used content analysis to analyze pre-existing reviews of users with disabilities. Four of the studies were tested with users under therapist supervision. Five studies solicited standard usability feedback on user satisfaction, employing the system usability scale or its dimensions to assess user satisfaction with the interface.

Six of the studies evaluated both technology performance and user attitudes. Performance metrics included automatic speech recognition-human agreement, task completion and duration, recording, rhythmic, and speech accuracy. User attitude metrics included ease of use, likes, dislikes, and interest level. One study evaluated an adapted version of the technology

Table 1. Participant numbers and demographics.

Study	N	Demographics
Ahmed et al. 2018	23	10 children with apraxia, mild to severe. (9 male and 1 female. Mean age: 7.9 years. Range: 6–11 years.) 6 typically developing children (1 male; 5 female; mean age: 8.7 years; range: 7–11 years). 7 speech-language pathologists (7 female; median experience: 12 years; ages 8–28 years)
Ballard et al. 2019	5	5 participants with apraxia and aphasia secondary to stroke
Hair et al. 2018	21	21 English speakers. 14 children with speed disorders ranging from mild to severe (7 motor-speech and 7 phonological impairments; 13 male and 1 female; mean age: 7.4 years, range: 4-12 years old).7 children reported by parents to be TD (4 male and 3 female; mean age: 8.7 years; range: 5-12 years old).
Lan et al. 2014	6	6 children. 5 male. Ages 4-12 years. 3 CAS, 3 control.
Pastorino et al. 2014	36	36 users in groups according to their age and neurological condition. 27 typically developing. 9 with apraxia and action disorganization syndrome. Mean age = 51.20 years. 21 females. 15 males.
Pirovano et al. 2014	7	7 subjects between ages 68 and 82. With therapists and physical comorbidities, such as past strokes, hypertension, heart problems (pacemaker or arrhythmia), arthritis, hip prosthesis, and glaucoma with macular dystrophy.
Pradhan et al. 2018	346	346 Amazon Echo reviews that include users with disabilities: visual impairment (37.9% of reviews), motor or mobility impairment (30.6%), speech impairment (13.6%), cognitive impairment (11.8%), and hearing loss (4.6%)
Shalash et al. 2015	20	20 in two groups of users: 5 therapists and 15 parents as one group and 15 children their ages range from two and half years to seven years as second group
Shepherd 2014	20	Tested on people with normal speech to verify. Pilot study on speakers with fluency deficits. Say thirty words depicted using photographs of various objects. Pre-test & post-test for accuracy.

acceptance model, which asked about dimensions in closer alignment with user experience research, such as whether the technology will enhance job performance.

Study Findings

Table 3 gathers the study findings. All studies found improved technology and user performance, such as improved word production, automatic speech recognition-human agreement, improved word production accuracy, about 80% success rate, and positive spoken and rhythmic accuracy trends. However, five studies reported problems with performance, such as ASR-human

Table 2. Methods and evaluation metrics.

Study	Tech	Methods	Evaluation
Ahmed et al. 2018	Speech driven mobile games	Task scenarios. During therapy. 1 session.	Automatic speech recognition-human agreement. Word length. Likes. Dislikes. Ease of Use. Interest level.
Ballard et al. 2019	Tablet app for therapy	Summative test. During therapy. Several sessions (one month).	Automatic speech recognition-human agreement. Word length. Satisfaction.
Hair et al. 2018	Speech therapy game	Hierarchical Task Analysis. 1 session.	Concentration. Completion. Satisfaction.
Lan et al. 2014	Mobile game for speech timing, prosody skills	Play and survey. 1 session.	Satisfaction
Pastorino et al. 2014	Cognitive rehab system	Task scenarios. Summative test. 1 session.	Audio cue and recording accuracy. Satisfaction.
Pirovano et al. 2014	Intelligent games	Formative & summative usability tests. Therapist supervision. 1 session	Technology acceptance. Ease of use. Success rate.
Pradhan et al. 2018	Voice-controlled personal assistants	Content analysis	Ratings. Comments.
Shalash et al. 2015	Therapy app	Usability tests. Therapist supervision. 1 session	Completion. Ease of use. Clarity.
Shepherd 2014	Therapy game	Pre- & post-test. In game scores. 2 weeks.	Rhythmic & speech accuracy.

disagreement, particularly considering apraxia severity, interaction errors, equilibrium problems, limited control over speech output, and some completion errors. Despite these problems, participants in six studies reported positive user experience and satisfaction, including finding the technology fun, interesting, enjoyable, and potentially useful. One article only reported user attitudes about technology effectiveness and efficiency. Another did not report user experience or satisfaction information.

Most reported user enjoyment of some kind. However, those five studies only comprised 84 participants out of 484 across all studies. Two studies reported therapeutic improvement for participants. Two others reported some form of adverse event and challenges for participants, such as equilibrium problems. Only two reported that participants found the technology easy. Another reported ease for younger users and difficulty for older users.

Study Quality

Reporting quality was low in most of the studies. All the articles were missing some form of information or another. Many failed to provide basic information about users or enough details about the methods (including samples of questions asked). Furthermore, not all included user experience variables, such as satisfaction or system usability scales that focus more on users' immediate attitudes about the interface than therapeutic outcomes.

Table 3. Findings.

Study	Findings
Ahmed et al. 2018	Users said speech-controlled games are **interesting & fun**, despite **ASR–human disagreements**. Automatic speech recognition–human agreement **was higher for SLPs than children, similar between typically developing & apraxia unaffected by apraxia severity** (77% TD, 75% CAS = incorrect; 51% TD, 47% CAS, 71% SLP = correct). Manual stop recording yielded higher agreement than automatic. **Word length did not influence agreement.**
Ballard et al. 2019	ASR–human agreement on **accuracy averaged about 80%.** All **had improved word production accuracy over time** with the ASR-based feedback (after 1 month). All reported **enjoying using the app** with speech pathologist support.
Hair et al. 2018	Most children indicated that doing exercises after completing each level was **less disruptive and preferable to doing exercises** scattered through the level. **Children liked having perceived control over the game** (character appearance, exercise behavior)
Lan et al. 2014	**Six said it was "fun." One said it was a "bit hard to control at times." One said it was "not as hard."** Users with CAS liked the assisted mode; controls liked the free mode. Three did not like the headset microphone.
Pastorino et al. 2014	**Audio cue selected nearly 80% of the time. Text cue selected 20%. 90.1% of sessions were correctly recorded, with interaction errors.** AADS patients solicited technical support (M = 3.0, SD = 1.2). Mental workload very low for young (M = 1.2, SD = 0.4), elderly (M = 1.7, SD = 1.4). Arduous (M = 2.6, SD = 1.4) for AADS patients.
Pirovano et al. 2014	**All the subjects found the games engaging and stimulating and rated them positively. The success rate approached 80%.** No pain or stress was reported except for a subject with equilibrium problems.
Pradhan et al. 2018	**Some accessibility challenges exist; limited control over speech output settings for users with hearing loss.** Users with disabilities utilize the Amazon Echo, including for unexpected cases such as speech therapy and support for caregivers.
Shalash et al. 2015	**All five therapists and 13 of 15 parents did all the tasks directly and navigated easily without mistakes. Fourteen of 15 children played level 1 in a short time. 12 of 15 played level 2 in a short time. 11 of 15 played level 3 correctly.** Generally, the children's impression was good, and they liked the animated picture response a lot.
Shepherd 2014	**There was a definite upward trend in their rhythmic and speech accuracy. There was a positive trend showing participants' abilities improved.** Small trial shows how spoken accuracy increased. A trial showed how rhythmic accuracy increased after playing the game. The in-game scores showed that spoken accuracy became better after playing.

The articles also had low citation and h-index scores. The content analysis was cited the most by others (n=42). It achieved the highest h-index score of 5. The others were cited between 0 and 9 times only, and they reached no more than a one in an h-index score. All but one of the articles

Table 4. Study quality.

Study	Cites	h-index	IRB	COI
Ahmed et al. 2018	5	1	1	1
Ballard et al. 2019	0	0	1	1
Hair et al. 2018	4	1	1	1
Lan et al. 2014	11	0	0	1
Pastorino et al. 2014	7	1	1	1
Pirovano et al. 2014	9	1	0	1
Pradhan et al. 2018	42	5	0	0
Shalash et al. 2015	1	0	0	1
Shepherd 2014	0	0	0	1

declared conflicts of interest or acknowledged funding sources. Only four of the articles—just under half—reported receiving permission from Institutional Review Boards, Committees from the Protection of Human Subjects, or ethics committees to test users.

User Feedback

There was virtually no user feedback provided in the research articles, and the research articles reported primarily clinical uses of speech recognition rather than everyday uses. Twitter content provided some insight into user feedback. The search yielded 247 tweets from November 30, 2011, to January 23, 2021. Forty-nine (n=49) of the tweets did not pertain to users with apraxia or speech impediments; they mentioned the names of people rather than speech recognition technologies. Fifty-five (n=55) tweets were of users who felt that the robot voice of the speech recognition software itself had a speech impediment. These tweets fell outside of the scope of this study—namely, understanding the challenges faced by apraxia users use speech recognition technologies in their daily lives. Overall, 143 tweets reflected the attitudes about these technologies held by users with self-described, self-reported apraxia or speech impediments. Of these, 20 tweets reported a combination of both positive and negative evaluations by comparing two different technologies or by describing how frustration with the technologies instilled pride or positive feelings about themselves or their family members with apraxia or a speech impediment. Because some tweets contained both positive and negative sentiment, the total close reading count of the overall evaluations was larger than the number of tweets.

Within the 143 tweets, there were 116 (81.1%) negative evaluations and 47 (32.9%) positive evaluations. The evaluations were derived from close reading each tweet and manually coding the overall sentiment and overall technology assessment. Not all tweets indicated the use for which the technologies were deployed. However, some users discussed how they use these speech recognition technologies for general speech tasks (n=161), playing songs (n=8), writing (n=2), and playing videos (n=2). See Table 5.

Positive evaluations included complements for devices that understand users with disabilities. For example, this user was pleasantly surprised that

Table 5. User feedback.

	tweets	positive	negative	both
Alexa	85	39 (45.9%)	61 (71.2%)	15 (17.6%)
Bixby	1	0 (0%)	1 (100%)	0 (0%)
Cortana	7	1 (14.2%)	7 (100%)	1 (14.2%)
Google	4	1 (25%)	4 (100%)	1 (25%)
Siri	46	6 (13%)	43 (93.5%)	3 (6.5%)
Total	143	47 (32.9%)	116 (81.1%)	20 (14%)

Siri understood her child with a speech impediment: "My child was playing with my phone, and I was afraid Siri wouldn't understand him. But she did. Loud and clear. #Apraxia #speech #CAS." Still, others reported how the precision with which they had to speak to these technologies helped treat their speech impediments: "I think writing my jokes using Siri is starting to fix my speech impediment." An article reporting one person's memoir about how the technology seemed to cure their child was retweeted 22 times: "Alexa Cured My Daughter's Speech Impediment http://dlvr.it/PdNjjw." Some users perceived that everyday speech recognition technology had therapeutic benefits.

In evaluations containing both positive and negative feedback, users shared that one technology was better than the other at understanding users with disabilities. For example, this user expressed frustration at using other's technology because their technology was not imprinted with their unique voice recognition: "I hate using other people's Siri because only my Siri understands my weird speech impediment." Still another compared Siri to Cortana: "I must say I'm impressed with Cortana - it seems to understand everything I say, unlike Siri, which makes me feel I have a speech impediment." Finally, other tweets with positive and negative evaluations praised the tenacity and effort of the user with apraxia and speech impediments for persisting and finding affordances that help the technology work for them. For example, a parent expresses pride in their child's persistence in using Alexa despite the child's speech impediment: "8yo daughter has Down syndrome. Speech hard to understand (apraxia), yet she got @amazon Alexa to play her favorite @Eminem song this am!" Users with speech impediments used affordances and workarounds to use speech recognition technology.

Finally, the negative feedback expressed frustration at the technology's inability to understand the voices of those with a speech impediment and with apraxia. One user called the technology useless to them: "Having a speech impediment makes Siri pretty much useless. Actually, very useless." Another complained about Google Assistant: "Using google assistant with a speech impediment is terrible." Still others complained about Alexa: "If you have any sort of speech impediment, then dealing with @amazon Alexa can be very disheartening and infuriating." Every technology—Google Assistant, Cortana, Bixby, Alexa, and Siri—had more negative evaluations than positive ones. Overall negative words used in the tweets increased over time from 2010 to 2021, except for one spike in positive feedback in February

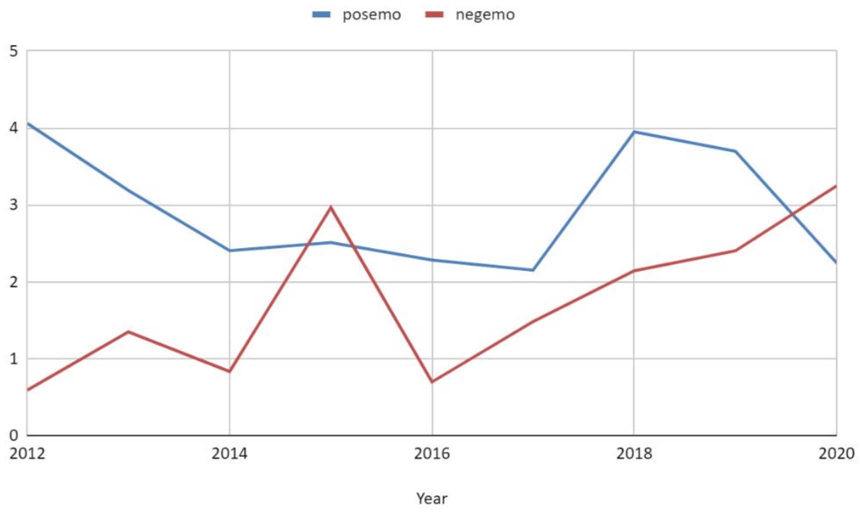

Figure 1: Positive and negative sentiment over time.

Table 6. Sentiment per product.

Tech	Tweets w/ interactions	Interactions	Positive	Negative
Alexa	53	195	20 (37.8%)	44 (83%)
Cortana	4	14	0 (0%)	4(100%)
Google	2	6	0 (0%)	2 (100%)
Siri	31	456	4 (12.9%)	29 (93.5%)
Total	90	671	24 (26.7%)	79 (87.8%)

2019 when the link to the memoir about the technology during a speech impediment was retweeted several times. See Figure 1.

Ninety tweets received some form of interaction, including replies, retweets, and likes. The interaction score was derived by adding replies, retweets, and likes. There were 671 interactions. Some tweets with interactions contained both positive and negative evaluations; therefore, the numbers of positive and negative interactions exceeded the raw total of tweets.

Overall, interactions were more negative (87.8%) than positive (26.7%). Alexa and Siri received the most interactions. Alexa received a more significant percentage of positive interactions than did Siri.

CONCLUSION

This research revealed that much more research is necessary for helping voice control and speech recognition software to better serve communities with speech therapy in general and apraxia of speech in specific. Overall, published studies are limited to clinical contexts for users with apraxia, where the technology is meant to help treat and train pronunciation and diction. The studies do not examine how users with apraxia deploy these technologies in their personal lives for their quality of life. The studies suggest that the technology has a minor benefit for users with apraxia in clinical settings. Still, the studies do not answer to what extent speech recognition systems work for users with apraxia.

Current research does not tell us about speech recognition software's user experience and usability for people with apraxia of speech. It reports clinical applications of the technology for therapeutic purposes. The quality of the reporting could be significantly improved insofar as current articles were missing information and were not published in prominent journals. Future research should provide reporting guidelines and investigate the everyday use of the technology by people with apraxia and other speech impediments. Speech recognition systems might work for clinical purposes to help in therapeutic activities for people with apraxia. Still, it is unclear if users with apraxia face more challenges with everyday use than their counterparts without speech impediments. It was also unclear from the literature how these systems fail this population in everyday settings. Future studies should include formative and summative investigations, user tests, feedback, and in situ observations, to determine more clearly how these systems fail and what changes are necessary to improve the performance of such speech recognition systems for users with apraxia.

The analysis of tweets of feedback from users with apraxia and speech impediments shows room for improving the precision and accuracy of these technologies as used in the lives of people with apraxia and speech impediments. There was far more negative feedback about the technologies and their inability to understand users with apraxia and speech impediments. The tweets did not reveal a wide range of activities users with apraxia and speech impediments undertake. This finding might suggest that the technology is only marginally helpful to users with apraxia or speech impediments or that users have abandoned the hope of using these technologies for anything but basic functions. The negative feedback outweighed positive feedback, but positive feedback suggested differences in the programming and responsiveness of name brand speech recognition products to the needs of users with apraxia and speech impediments. Mixed feedback also indicates that, for some conditions, it might be the case that repetition in interaction with the technologies might provide some form of practice or therapy for some, but it presents a point of frustration and inequity for others. Future research should deploy user experience and usability research methods to gather observational, formative, and summative user feedback to learn how users with apraxia use and adapt these technologies. It should also compare the effectiveness of different speech technologies used with apraxia in daily life.

Of 120 articles referencing the subject matter, only 9 provided enough detail to suggest that the technology developed for this community was tested with users in the community. Of the studies that met inclusion criteria, only about a fifth of the users and participants recruited were diagnosed with apraxia, a very particular speech disorder that directly impacts speech recognition accuracy and precision. Furthermore, while the samples were often heterogeneous regarding speech diagnosis, gender, and age, the sample may have been homogeneous in terms of race and ethnicity. These factors are essential to consider because they may impact tone, texture, intonation, and other speech detection variables. Study methods were primarily orthodox user testing involving task scenarios. However, most of the applications were designed for use in or for therapy rather than for use in the daily lives of users with apraxia.

The samples and methods of most of the studies were heterogeneous, which complicates generalizing the findings to the apraxia community specifically or at large. They were diverse in their sampling as well. The research also employed a mixed design, where participant health conditions or outcomes were pertinent in some cases, and the effectiveness of speech recognition was pertinent in others. In instances where therapists were tested alongside disabled users, the design was specific to clinic settings rather than quotidian settings. The diversity also made meta-analysis findings more challenging to obtain. Most of the studies examined therapeutic rather than lifestyle applications. And there were fewer longitudinal tests where user interactions were observed over time rather than in one session. These details suggest that we have more to learn from disabled users' daily interactions and engagement with speech recognition and voice-controlled technology.

Only one study examined voice-controlled personal assistants and other, more pervasive forms of speech recognition technologies. Furthermore, some of the studies were with users and therapists, and only two requested that users engage with the software over several sessions and weeks. In this way, the research studies do not help confirm or provide helpful information about the in-situ experience of users with apraxia. The therapy-focused applications also impacted the findings insofar as all studies reported user and technology performance metrics meant to confirm the integrity and internal validity of the software. Furthermore, we should also understand the almost unanimous user satisfaction through the lens of the therapeutic purposes of the applications. Users might have higher standards, expectations, and needs for technology that they use every day for their quality of life rather than for therapy.

Finally, overall, the studies did not report enough details to confirm the quality of the findings. These factors might also have impacted the articles' overall low h-index and citation numbers. It is essential to generate more insights and findings from user experience and usability research of people with disabilities to ensure parity and accessibility of the everyday technology that impacts and potentially improves the quality of life.

Recruitment can be a barrier when testing in disability communities, as can administering user tests in these communities (Shepherd 2014). Standard protocols may require modifications and accommodations, as might unique adaptations for informed consent and scheduling sessions. User testing can take more time and require changes to traditional research methods. The researchers of the articles included in this review might have encountered these barriers. However, including diverse participants, especially people in disability communities that could very much benefit from new technologies, is vital throughout the design process. Their input can improve technology design in ways that designers do not anticipate.

Future studies should include more homogeneous samples in terms of speech conditions and more heterogeneous samples in terms of demographics. Furthermore, more research studies reporting product design for this community should detail user experience and usability testing. Finally, product designers should test products with diverse populations, including those with disabilities, and develop personae to help them keep in mind their needs. While recruiting and retaining these users might be difficult, any extra effort will pay dividends in product quality and marketability.

REFERENCES

Ahmed, B., Monroe, P., Hair, A., Tan, C. T., Gutierrez-Osuna, R., & Ballard, K. J. (2018). Speech-driven mobile games for speech therapy: User experiences and feasibility. International journal of speech-language pathology, 20(6), 644–658.

Ballard, K. J., Etter, N. M., Shen, S., Monroe, P., & Tien Tan, C. (2019). Feasibility of automatic speech recognition for providing feedback during tablet-based treatment for apraxia of speech plus aphasia. American journal of speech-language pathology, 28(2S), 818–834.

Booher, H. R., & Minninger, J. (2003). Human systems integration in army systems acquisition. Handbook of human systems integration, 663-698.

Booher, H. R. (2003). Handbook of human systems integration (Vol. 23). John Wiley & Sons.

Chapanis, A. (1996). Human factors in systems engineering. John Wiley & Sons, Inc..

Colomer, J. B. M., Salvi, D., Cabrera-Umpierrez, M. F., Arredondo, M. T., Abril, P., Jimenez-Mixco, V., ... & Medrano, A. (2014). Experience in evaluating AAL solutions in living labs. Sensors, 14(4), 7277–7311.

Corcoran, M. (2018). When alexa can't understand you. Slate (online), Oct.

Dey, A., Jenamani, M., & Thakkar, J. J. (2018). Senti-N-Gram: An n-gram lexicon for sentiment analysis. Expert Systems with Applications, 103, 92-105.

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. International journal of qualitative methods, 5(1), 80–92.

Folds, D., Gardner, D. L., & Deal, S. (2008). Building up to the human systems integration demonstration. Insight, 11(2), 15–18.

Friedenthal, S., Moore, A., & Steiner, R. (2014). A practical guide to SysML: the systems modeling language. Morgan Kaufmann.

Hair, A., Monroe, P., Ahmed, B., Ballard, K. J., & Gutierrez-Osuna, R. (2018, June). Apraxia world: A speech therapy game for children with speech sound disorders. In Proceedings of the 17th ACM Conference on Interaction Design and Children (pp. 119–131).

Harwell, Drew. (2018) The Access Gap. Washington Post. Retrieved from https://www.washingtonpost.com/graphics/2018/business/alexa-does-not-understand-your-accent/

Honour, E. C. (2006, July). 2.3. 1 A Practical Program of Research to Measure Systems Engineering Return on Investment (SEâŁłROI). In INCOSE International Symposium (Vol. 16, No. 1, pp. 299–308).

Lan, T., Aryal, S., Ahmed, B., Ballard, K., & Gutierrez-Osuna, R. (2014, October). Flappy voice: an interactive game for childhood apraxia of speech therapy. In Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play (pp. 429–430).

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). Research methods in human-computer interaction. Morgan Kaufmann.
Meilich, A. (2008). INCOSE MBSE Initiative Status of HSI. MBSE Activity (Presentation).
Pastorino, M., Fioravanti, A., Arredondo, M. T., Cogollor, J. M., Rojo, J., Ferre, M., ... & Wing, A. M. (2014). Preliminary evaluation of a personal healthcare system prototype for cognitive eRehabilitation in a living assistance domain. Sensors, 14(6), 10213–10233.
Pirovano, M., Mainetti, R., Baud-Bovy, G., Lanzi, P. L., & Borghese, N. A. (2014). Intelligent game engine for rehabilitation (IGER). IEEE Transactions on Computational Intelligence and AI in Games, 8(1), 43–55.
Pradhan, A., Mehta, K., & Findlater, L. (2018, April). " Accessibility Came by Accident" Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In Proceedings of the 2018 CHI Conference on human factors in computing systems (pp. 1–13).
Roundtree, A. K. (2020). ANT Ethics in Professional Communication: An Integrative Review. American Communication Journal, 22(1).
Shalash, W. M., Sait, N., & Hashem, D. (2015). Three stagesA Interactive System for Solving Children Communication Disorder. International Journal of Computer Applications, 130(16).
Shepherd, J. J. (2014). Benefits of video games in multidisciplinary scientific research (Doctoral dissertation, University of South Carolina).
Taubman, P. (2008). Top engineers Shun Military; concern grow. The New York Times, 25.
Whittemore, R., & Knafl, K. (2005). The integrative review: updated methodology. Journal of advanced nursing, 52(5), 546–553.