Breaking Boundaries with Live Transcribe: Expanding Use Cases Beyond Standard Captioning Scenarios

Fernando Loizides
Cardiff University, UK
Cardiff
loizidesf@cardiff.ac.uk

Sara Basson
Google, USA
New York, New York
basson@google.com

Dimitri Kanevsky
Google, USA
Mountain View, California
dkanevsky@google.com

Olga Prilepova
Google, USA
Mountain View, California
olgaprilepova@google.com

Sagar Savla
Google, USA
Mountain View, California
sagr@google.com

Susanna Zaraysky
Google, USA
Mountain View, California
susanna.zaraysky@gmail.com

ABSTRACT

In this paper, we explore non-traditional, serendipitous uses of an automatic speech recognition (ASR) application called Live Transcribe. Through these, we are able to identify interaction use cases for developing further technology to enhance the communication capabilities of deaf and hard of hearing people.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility technologies**; **Accessibility**; Tablet computers; Mobile phones.

KEYWORDS

Deaf, DHH, Hard of Hearing, Interaction, ASR, Automatic Speech Recognition, Speech to Text

1 INTRODUCTION AND MOTIVATION

Not being able to hear in environments predicated on sounds and aural communication presents multiple challenges. For deaf and hard of hearing (DHH) people, this can lead to severe communication breakdowns. For children, this can hinder social and emotional development [2, 6, 22, 27]. Developmental gaps can persist to adulthood [3] with higher instances of depression [14]. Communication gaps have historically created barriers to employment and to informal human interactions [8, 17, 20]. Some positive changes in employment have emerged with legislative changes, but significant gaps in employment and career development persist for DHH people [13]. Reducing communication friction between a DHH person

and other DHH or hearing people can improve quality of life for the DHH individual and for their extended community of contacts. Sign language provides excellent communication opportunities within deaf signing communities, but is generally unavailable outside of these communities. Some deaf children are trained in oral communication and may not themselves be fluent in sign language. Adults who have lost hearing later in life may not adopt sign language, and may prefer text based communication [4]. Fluent sign language users can experience communication blocks when sign language interpreters are not available[18]. For casual communication, there is a high cost barrier. Users need to schedule interpreters in advance, limiting opportunities for spontaneous conversations with non-signing people. DHH people that do not sign and communicate with the speaking community often depend heavily on lip reading. There is a steep learning curve to develop good lip reading skills, since only about 30% of speech gestures are discernable from the face and lips [25]. Users depend on a lot of contextual information to fill in the gaps. Lip reading success depends on good lighting and direct line of sight as well as the speaker facial expressions, accents and pace.

Contemporary research has focused attention on technical solutions to improve interactions between DHH and speaking/hearing communities. Medical advancements such as Cochlear implants are increasingly prevalent[9]. The expectation is that users with Cochlear implants will "learn to identify enough sounds" to communicate in the speech/hearing community, supplemented by lip reading. This is not always the case, as acoustic or visual conditions may be less than ideal for the DHH with Cochlear implants to fully engage in conversation. Speech captioning is often a necessary supplement to improve the communication flow, and automatic speech recognition has become a critical assistive technology [12]. Bell Labs in the 1950s [21] was one of the early inventors and developers of speech recognition technology. Early applications of speech recognition targeted controlling computers by voice, interacting with services over the phone through voice menus, and speech dictation (generating documents through speech rather than using a keyboard.) Telecommunications relay services also used speech recognition (https://www.washington.edu/accesscomputing/what-are-telecommunications-relay-services - accessed June 2020). DHH users would have a phone with a text screen. A third party intermediary would re-speak the conversation of the hearing party into speech recognition that was trained for the intermediary's

voice. Text from the speech recognition output would appear on the DHH user's screen. As processing power increased and algorithms advanced, speech recognition improved such that trained third party intermediaries have become less necessary, and the speech of the hearing parties in the conversation can be transcribed directly (although sometimes there are challenges to doing so automatically [10]) with speech recognition [11]. As speech recognition technology has improved, it has been used more widely to caption lectures, videos, and conversations between DHH and hearing people [23, 26]. Usage is expanding beyond the typical, anticipated scenarios, as users become more comfortable with the technology. Live Transcribe is an Android mobile application that was released to the public in 2019. It was initially intended to help DHH and hearing people communicate more effectively. Live Transcribe provides free, real time speech-to-text transcriptions to make everyday conversations more accessible. Users can choose from over 80 languages and dialects, and quickly switch between languages. Adopting new interactive technology is a community process, often predicated on usefulness and likability.

We have catalogued a number of novel uses of Live Transcribe speech recognition technology (https://www.android.com/intl/en_uk/accessibility/live-transcribe/ - accessed June 2020). We summarize a number of them in this paper. Live Transcribe has demonstrated rapid adoption, in settings that are predictable, but also in settings that were unexpected. In some scenarios, the social constraints surrounding the COVID-19 pandemic have spawned new use cases. The authors are either involved in the development of Live Transcribe or active users of the application with several of the experiences being a first hand account of the author's use of the app. The reports were gathered by the Live Transcribe team at Google. Both Googlers and non-Googlers who used the app sent notes about some of their experiences about how they use the technology to the Live Transcribe team. Some of the reports came from researchers at Gallaudet University in Washington DC in the United States. Gallaudet University is a university for the Deaf and hard-of-hearing that collaborated with Google to make the app. More specifically, the app was inspired by a dinner in a poorly lit restaurant with Dimitri Kanevsky and Sara Basson's family in New York. Dimitri is deaf. It was too dark for Dimitri to be able to lip-read. Sara's son took out his Android phone and voice typed in English into Google Docs. Dimitri read what Sara's son was saying and got the idea for a transcription app that would type out what people were saying in real time. Dimitri Kanevsky is one of the co-creators of Live Transcribe along with Chet Gnegy. Susanna Zaraysky uses the app with her father in Russian and English to communicate through the glass doors of his nursing home. Prior to the Covid lockdown, she used the app when her father's hearing aids weren't working or for him to follow along better in doctor's visits when he didn't hear clearly. Olga Prilepova uses Live Transcribe both in English with people in California and in Russian with people in Moldova when visiting. To avoid holding a device in one's hand while speaking in Live Transcribe, she has come up with ways to wear devices on one's body to display the transcriptions and how to use multiple devices to display transcriptions during video calls. Sagar Savla is the Product Manager for Live Transcribe and other sound and accessibility research efforts. He was motivated by wanting to help his grandmother with her age-onset hearing loss.

Fernando Loizides is an assistant professor of computer science and human-computer interaction who is gathering use cases for Live Transcribe and looking for paths to better interaction for accessibility. He uses Live Transcribe in traditional and non-traditional ways to gather live transcription in meetings and to "catch-up" to conversation of a meeting that has already started by re-reading the previous minutes' transcription.

2 COMMUNITY ACCEPTANCE AND ADOPTION

Any technology that is novel, faces a hurdle, big or small, of being accepted and used. This is modelled and frame-worked by the technology acceptance model [16]. Technology acceptance and adoption varies depending on several variables such as perhaps (among others) age, previous experience and exposure, trustworthiness, perceived usefulness and likeability. In the following examples, we demonstrate different scenarios where technology acceptance and adoption is demonstrated across different scenarios.

2.1 Employment opportunities that were previously unavailable

A hearing colleague went to a sports event. He needed directions for parking and asked an usher. The usher initially did not respond. When he got her attention, she pulled out her phone and started using Live Transcribe to converse. As a deaf person communicating with a hearing audience within the event context, the flexibility associated with Live Transcribe enabled her to fulfil a job as an usher.

2.2 Ability for deaf people to spontaneously communicate with a hearing person

Two deaf men on a business trip were walking to their hotel from the conference site. They were signing to each other and not planning to communicate with a hearing person at that time. A woman that seemed distressed approached them, asking for help. It was too dark to read her lips and they couldn't understand her. One of the deaf travellers took out his Android phone, opened Live Transcribe, and asked her to repeat her problem. Live Transcribe transcribed her perfectly. She was lost and couldn't find her hotel. The two deaf travellers were able to accompany her back to where she needed to be. The woman was able to adapt seamlessly to Live Transcribe and was extremely grateful for the help in getting to her hotel safely.

2.3 Shared responsibility for communication: Hearing users enabling Live Transcribe

A deaf user of Live Transcribe works in technical support, assisting employees with technical problems. In this particular instance, her phone had broken, and she did not know how she would communicate with her colleagues. Several hearing colleagues had already installed Live Transcribe on their phones, enabling her to communicate with them. This also freed her from her typical burden of constantly carrying her phone with her.
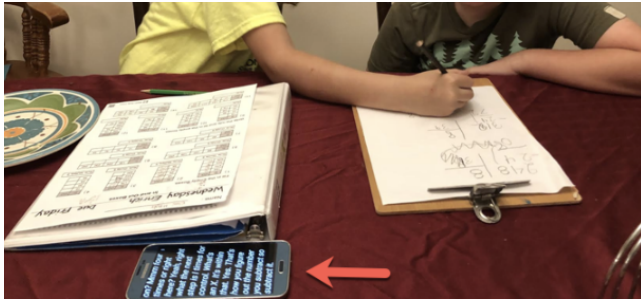
Figure 1: Using Live Transcribe to 'listen In' on the children's conversation

2.4 Deeper understanding between deaf parents and hearing children

There are complex relationships and psychological factors between parents and children, when some are DHH and others are hearing [24]. In the case where the parents are deaf, this can lead to the children being the primary spokespeople of the family and can sometimes leave the children feeling as though they have been deprived of their childhood [15]. In a family where all members are hearing, parents can engage in direct conversation with their children as well as overhearing "side" conversations (See Figure 1). A deaf professor commented that Live Transcribe enabled the parents to better understand how their children communicated, and how they helped each other with math. Live Transcribe enabled the parents to follow along to casual conversations as passive observers rather than active participants. As the parent puts it, "Now that's a nice application. I finally can see how the kids help each other with math".

2.5 Changing the communication style of a hearing person speaking to deaf people

A colleague indicated that she had communicated for many years with a deaf friend that depended on lipreading. Over the years, she had become aware of which sounds were more visible on the lips than others (only about 30% of English sounds are lip visible.) She had unconsciously become adept at opting for words with more visible sounds, given a choice. She also learned how to build up context, so that less familiar words would be recognisable. As a result, her own language choices would shift when she communicated with him. If there were many unusual and unpredictable words that needed to be communicated, she shifted to written/chat formats. Nonetheless, she couldn't be certain whether 100 per cent of what she was saying was intelligible. With Live Transcribe, verbal communication became more natural. She no longer monitors word choices, consciously or subconsciously. Since she can also view the Live Transcribe output, she knows what is or is not accurately understood, and when to repeat or re-state.

3 EMERGENCY COMMUNICATION DURING COVID-19

Due to the lockdowns, DHH people were sometimes left scrambling to find ways to communicate, as social distancing, quarantine and mask wearing rendered their previous communication methods impossible (e.g., standing close to their interlocutors and lip reading). These examples show how DHH users, their family members, and strangers adapted to using Live Transcribe, including people previously unfamiliar with the app, in order to increase inclusivity [1].

3.1 DHH people that typically lip read, interacting with hearing people wearing masks

Mask wearing is pervasive during the COVID-19 pandemic, making lip reading impossible for DHH people that depend on it. A video (accessed June 2020) has become popular on Facebook, showing Live Transcribe providing conversational access when lipreading is unavailable (https://www.facebook.com/drew.ledson/videos/vb. 1090025427/10221234434025640/?type=2&theater).

3.2 Communication through closed (glass) doors

A hearing colleague had a number of family celebrations in March 2020: It was the 40th anniversary of their departure from the former Soviet Union. It was also her parents' 50th wedding anniversary. Due to COVID-19, normal celebrations became impossible. Her father was in a nursing home, and she knew that he would not be able to hear them speaking to him from outside the nursing home, even with his hearing aids. She opted to try Live Transcribe on a tablet. "One of the nurses wheeled my dad in his wheelchair to the glass doors of the nursing home," she said. Her father wasn't expecting to see her, or his grandchildren, standing outside (See Figure 2 Left). They were able to hear and understand him. He was able to read the transcription on the tablet, as his granddaughter said "We miss you and we love you". It wasn't the big party at a nice restaurant they had originally planned, but it provided an opportunity to recognise a meaningful time for their family. As quarantine procedures at the nursing home became stricter, even visiting by the glass door proved impossible. The family replicated the "glass door" experience using a smartphone and video chat (See Figure 2 Right). They held the tablet with transcription such that her father could see the speaker and the Russian transcribed speech at the same time. They commented on the significance of using this technology to celebrate their family's departure from a country that didn't allow people to communicate freely.

3.3 Speaker feedback leads to better speech

A deaf patient with a Cochlear implant needed to see their audiologist at Johns Hopkins in Baltimore due to an emergency situation. It was the first day that the clinic had started seeing clients again after Maryland had lifted some restrictions, and many protocols for making the screening process for COVID-19 symptoms had been hastily put into place without regard for accessibility. At the entrance to the clinic, people in full Personal Protective Equipment (PPE) asked a set of screening questions, but it was impossible to understand anything they were saying. PPE also affects the accuracy of automatic speech recognition, so simply using Live Transcribe on their phone didn't quite work. The patient explained the situation, and
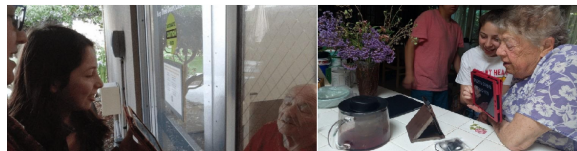
Figure 2: Left: Communicating through the glass with Live Transcribe Right: Speaking using Live Transcribe in Russian on their tablet while the father watches them via video call on the phone propped up on a homemade phone stand.

by holding the microphone really close to the worker in PPE, the worker could see the output in real-time and also correct herself in real-time if the PPE resulted in a garbled response. This feedback mechanism was critical to making this communication work.

3.4 Managing distractions

Online or face-to-face meetings provide an easy way for several parties to communicate in real time. Online meetings are the prevalent mode during COVID-19 lockdown, and it is easy to get distracted by other stimuli when working alone. One hearing user commented that they often switch to another window to reply to an email or a chat. When they switch back to the meeting, they realize that they have missed a portion of the conversation. With Live Transcribe, they are able to "catch up" on the portion of the conversation they have missed. They now use Live Transcribe more frequently as a meeting tool to provide note taking.

3.5 Experimenting with communication at a distance

A deaf colleague with a Cochlear implant had to ship two laptops and set up packing and shipping in the store. The process took about 20 minutes. He could not lip read given the store employee's mask, and did not want to ask the store employee to lower his mask. Nor did he want to engage in prolonged close contact by back-and-forth writing. He wasn't sure how well Live Transcribe would work, given the 6 feet (2 meters) of separation between them. He put his phone on the counter so that it was in visual range for himself as well as the store employee. The transcription worked well, and it was helpful that the employee could see his own transcription in real time.

4 CAPTIONING SPEECH AS A TOOL FOR SPEECH IMPROVEMENT FOR DHH USERS

DHH people often have particular speech characteristics associated with hearing loss, often referred to as "deaf speech". Research has explored the characteristics of deaf speech and various methods to improve articulation of deaf speakers, which is typically managed by speech therapists [5, 7, 19]. Practice with a speech therapist is inherently intermittent, however, and the therapist is not always available to correct a speaker's pronunciation. In the following experiences, Live Transcribe has been used as a tool to determine whether the deaf speaker's articulation is sufficiently accurate to be recognized correctly. Where it is not recognized correctly, the deaf speaker can dynamically tweak their own pronunciation until it is recognizable.

Figure 3: CART on a computer screen, Live Transcribe on a smartphone, and the round Roger Select (an assistive listening device)

4.1 Real time transcription to train DHH users with Cochlear implants

Cochlear implants [9] can provide DHH people with some access to auditory information. Often users need to train themselves on interpreting the sound provided through the Cochlear implant. One user commented that he has historically done this training with speech transcription provided by CART writers; stenographers that caption speech to create access for DHH people (See Figure 3). The output from a CART writer generally lags the actual speech by a couple of seconds. The CART writer needs to hear the speech, and convert it to stenotype. A DHH Cochlear implant user commented that training is more complicated when there is a lag between what he hears, and what he reads as the caption. With Live Transcribe, the latency is minor. The word that he hears is still in auditory memory when the caption for the word appears. He shared his setup in the image below. He continues to use CART which appears on his computer screen. He believes CART provides highest accuracy when other people speaking have strong accents. In addition he has Live Transcribe on display, for auditory reinforcement with low latency.

5 HANDS-FREE INTERACTION

5.1 Wearables

Ideally, captioning information would be available without burdening the DHH user to always have a phone in hand. Holding a

Figure 4: Wearable innovations for Live Transcribe Interaction. Left: head and chest wear. Center: chest wear. Right: wrist wear.

Figure 5: Using a stand for hand's free interaction

phone limits the person's ability to do two-handed tasks. Also, if the captioning appears on a phone in the user's hand, then the user will appear less engaged with other people in the conversation. Staring at the phone in hand can erroneously signal lack of engagement. Finally, the DHH person staring at a phone will miss other useful cues from the speaker that are gleaned from lip reading, body language, and other emotional indicators. We have begun experimenting with a number of display alternatives. The hat-mounted phone setup uses a sticky magnetic plate, a magnet and a tiny Palm phone (See Figure 4 - Left). Wearing and displaying Live Transcribe for others to view. One can monitor transcription by watching another phone with Live Transcribe in their hands (See Figure 4 - Center). This self-loop allows directing audio from a speaker to a mic, and transcribes audio that is played on the phone (including phone calls). A bracelet can be made from a piece of fabric cut from a shirt sleeve. We experimented with several wrist/hand mounts using a running wristband or a thin glove (See Figure 4 - Right).

5.2 Touch-free access to captions

We put a Live Transcribe tablet on a boom mic stand, for use by an immobile user. There are many ways to mount a tablet. Here, we used sticky-back velcro on a board (See Figure 5). This enables limited mobility DHH people to have a line of sight with the Live Transcribe tablet and be able to communicate with others in the room and through calls.

6 SUMMARY AND FUTURE WORK

In this paper, we presented several experiences of non-standard use of an automatic speech recognition application, Live Transcribe. These experiences serve as use cases to explore both the interaction behavioural models of DHH and hearing people. We have discovered social and physical barriers that were broken down through these use cases. and identified where the lives of DHH people and their interactions were enriched. In future, we hope to identify more cases where automatic speech transcription

in Live Transcribe and other speech recognition tools advances communications beyond one to one interactions. A recent project, code named Euphonia,(https://ai.googleblog.com/2019/08/project-euphonias-personalized-speech.html - accessed June 2020) allows users to train speech recognition models for non-standard speech. We also plan to capture experiences with other features in Live Transcribe (such as environment sound detection, noise level indicator, haptic feedback) that were not presented in this paper.

ACKNOWLEDGMENTS

REFERENCES

[1] Vicki A Aaberg. 2012. A path to greater inclusivity through understanding implicit attitudes toward disability. *Journal of Nursing Education* 51, 9 (2012), 505–510.

[2] Drew Arnold and Ann Tremblay. 1979. Interaction of deaf and hearing preschool children. *Journal of Communication Disorders* 12, 3 (1979), 245–251.

[3] Gunnel AM Backenroth. 1995. Deaf people's perception of social interaction in working life. *International Journal of Rehabilitation Research* 18, 1 (1995), 76–81.

[4] Larwan Berke, Christopher Caulfield, and Matt Huenerfauth. 2017. Deaf and hard-of-hearing perspectives on imperfect automatic speech recognition for captioning one-on-one meetings. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility.* 155–164.

[5] Lynne E Bernstein, Edward T Auer, and Paula E Tucker. 2001. Enhanced speechreading in deaf adults: Can short-term training/practice close the gap for hearing adults? *Journal of speech language and hearing research* 44, 1 (2001), 5–18.

[6] Rosemary Calderon and Mark Greenberg. 2011. Social and emotional development of deaf children: Family, school, and program effects. *The Oxford handbook of deaf studies, language, and education* 1 (2011), 188–99.

[7] Ruth Campbell and Helen Wright. 1988. Deafness, spelling and rhyme: How spelling supports written word and picture rhyming skills in deaf subjects. *The Quarterly Journal of Experimental Psychology Section A* 40, 4 (1988), 771–788.

[8] John B Christiansen and Judy Egelston-Dodd. 1982. *Socioeconomic Status of the Deaf Population.* Vol. 4. Gallaudet College.

[9] Graeme Clark. 2004. Cochlear implants. In *Speech processing in the auditory system.* Springer, 422–462.

[10] Martin Cooke, Phil Green, Ljubomir Josifovski, and Ascension Vizinho. 2001. Robust automatic speech recognition with missing and unreliable acoustic data. *Speech communication* 34, 3 (2001), 267–285.

[11] Kosei Fume, Taira Ashikawa, Nayuko Watanabe, and Hiroshi Fujimura. 2018. Implementation of Automatic Captioning System to Enhance the Accessibility of Meetings. In *International Conference on Computers Helping People with Special Needs.* Springer, 187–194.

[12] Biing-Hwang Juang and Lawrence R Rabiner. 2005. Automatic speech recognition–a brief history of the technology development. *Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara* 1 (2005), 67.

[13] Ronald R Kelly, Andrew B Quagliata, Richard DeMartino, and Victor Perotti. 2016. 21st-century deaf workers: Going beyond" just employed" to career growth and entrepreneurship. (2016).

[14] Marit H Kvam, Mitchell Loeb, and Kristian Tambs. 2007. Mental health in deaf adults: symptoms of anxiety and depression among hearing and deaf individuals. *The Journal of deaf studies and deaf education* 12, 1 (2007), 1–7.

[15] Harlan L Lane, Robert Hoffmeister, and Benjamin J Bahan. 1996. *A journey into the DEAF-WORLD.* Dawn Sign Press.

[16] Younghwa Lee, Kenneth A Kozar, and Kai RT Larsen. 2003. The technology acceptance model: Past, present, and future. *Communications of the Association for information systems* 12, 1 (2003), 50.

[17] Janet E MacLeod-Gallinger. 1992. The career status of deaf women: A comparative look. *American Annals of the Deaf* (1992), 315–325.

[18] Donna Jo Napoli and Rachel Sutton-Spence. 2010. Limitations on simultaneity in sign language. *Language* 86, 3 (2010), 647–662.

[19] Nancy A Neef and Brian A Iwata. 1985. The development of generative lipreading skills in deaf persons using cued speech training. *Analysis and intervention in developmental disabilities* 5, 4 (1985), 289–305.

[20] David L Passmore. 1982. Vocational and Economic Implications of Deafness. *Journal of Epsilon Pi Tau* 8, 2 (1982), 34–38.

[21] Roberto Pieraccini and ICSI Director. 2012. From AUDREY to Siri. *Is speech recognition a solved problem* (2012), 23.

[22] Robert L Schum. 1991. Communication and social growth: a developmental model of social behavior in deaf children. *Ear and hearing* 12, 5 (1991), 320–327.

[23] Matthew Seita and Matt Huenerfauth. 2020. Deaf Individuals' Views on Speaking Behaviors of Hearing Peers when Using an Automatic Captioning App. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–8.

[24] Jenny L Singleton and Matthew D Tittle. 2000. Deaf parents and their hearing children. *Journal of Deaf studies and Deaf education* 5, 3 (2000), 221–236.

[25] Quentin Summerfield. 1992. Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 335, 1273 (1992), 71–78.

[26] Emily Q Wang and Anne Marie Piper. 2018. Accessibility in Action: Co-Located Collaboration Among Deaf and Hearing Professionals. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–25.

[27] Mary T Weiner and Margery Miller. 2006. Deaf children and bullying: Directions for future research. *American Annals of the Deaf* 151, 1 (2006), 61–70.