HearMore: Addressing Hearing Loss through Automated Live Captioning

Katy Arnold
University College London
UCL Interaction Centre
London, WC1E 6BT
katherine.arnold.15@ucl.ac.uk

Albert Higgins
University College London
UCL Interaction Centre
London, WC1E 6BT
albert.higgins.16@ucl.ac.uk

Matej Kaninsky
University College London
UCL Interaction Centre
London, WC1E 6BT
matej.kaninsky.16@ucl.ac.uk

David Mühlebach
University College London
UCL Interaction Centre
London, WC1E 6BT
david.muhlebach.16@ucl.ac.uk

Caroline Trimm
University College London
UCL Interaction Centre
London, WC1E 6BT
caroline.trimm.15@ucl.ac.uk

Each submission will be assigned a unique DOI string to be included here.

Abstract

Hearing loss is an invisible disability which affects millions of people's experience of education. Without assistance children fall behind in school and the gap in achievement widens as they progress through education.

Here we present a new and innovative approach to real-time capture of spoken content for those who are hard of hearing. HearMore is a live captioning service that people can have with them every day. Based on research findings, it addresses the needs of those who are struggling to hear in a learning environment, encourages greater participation and drives up engagement for all.

Our approach differs from other solutions currently available as it relies on automatic speech recognition software rather than human captioners, which are expensive and inflexible.

Author Keywords

Hard of hearing; speech to text; automatic speech recognition; captions; wearable.

ACM Classification Keywords

H.5.m [Information interfaces and presentation (e.g. HCI)]; Miscellaneous

User research and analysis

Figure 1: A Palantypist at work

Figure 2: Creating an affinity diagram

Figure 3: An affinity diagram – output from in-depth interviews

Introduction

There are an estimated 11 million people in the UK (Action on Hearing Loss, 2015), and 37 million people in the US (Blackwell, Lucas & Clarke, 2014) living with some form of hearing loss. Often hidden from view, hearing loss has a significant impact on daily life. It leaves those affected with feelings of shame, embarrassment and can lead to social exclusion.

Hearing is critical to language development and communication, and good communication between teacher and student are critical to successful learning. Research shows that children with hearing impairments do less well in school and the gap in achievement increases over time (Packer, 2015). Interventions such as real time speech to text support services have been shown to significantly help the hard of hearing to take part in mainstream education (Elliot, Stinson & Easton, 2008).

However, live speech to text services rely on highly trained professionals - palantypists - which makes them extremely expensive, and they must be booked in advance at specific times. This pushes them out of reach for many who need them.

This study followed a user centred design process to create an assistive technology to help hard of hearing students get more from classroom based learning. We present evidence from our user research that a solution employing automatic speech to text software would be beneficial in this context.

Scope was limited to those who are hard of hearing and in further education. The profoundly deaf were also excluded as they differ both culturally and socially and have very different sets of needs.

Literature review

Some work in this area has focused on supporting interactions between students and their teachers. Fan et al. (2015) used a mobile phone system to gather and share in situ feedback between hearing students and teachers who were struggling to communicate in larger, noisier classes.

Alternatives to live captioning include the use of Amazon's online marketplace, Mechanical Turk, to provide crowd captioning in place of expensive palantypists. Several papers (Bernstein et al. 2011; Murphy et al. 2013; Lasecki et al. 2012; Lasecki & Bigham 2012) employed workers to simultaneously capture speech to text at events and produced results more accurate than voice recognition software at that time.

However, in the time since these papers were published, there have been great strides in voice recognition software. Commercially available technologies have improved dramatically and companies claim recognition software can achieve between 88% and 89% accuracy if operated under optimal conditions (Tran, Vincent, & Jouvet 2015). This technology is expected to soon achieve similar accuracies 'in the wild' which will open new possibilities for people who are hard of hearing.

Mostly, work in this area has focused on human captioners, whether paid for palantypists or crowd-sourced. We propose a new approach to real-time content capture in the classroom – using automatic

User Research Key findings

1. Hearing loss is an invisible disability
2. People with hearing loss don’t want to be singled out because of it or defined by it
3. The classroom is exhausting for students with hearing loss
4. Saving transcriptions for later is important
5. Keeping things simple is important

“I look like a freak – always turning round to ask ‘who said that?’”
Quote, P2 (moderate hearing loss).

“it’s exhausting. I’m missing parts of words so my brain fills them in – very tiring!”
Quote, P4 (moderate hearing loss).

“...just having a screen up there with someone typing what they say. It would make all the difference in the world.”
Quote, P1 (severe hearing loss).

speech recognition software (ASR) in place of trained palantypists - and the constraints they impose.

Researching the problem space

Early research included a questionnaire which went to a representative sample of 70 students to understand how many might be affected by hearing loss. Then, to build empathy, each member of the team took part in a bodystorming exercise, where they wore earplugs for the day

We sought the participation of members of a lip-reading class, in a higher education college in London and conducted contextual enquiry during a class with eight participants. This helped us better understand the challenges faced by lip-readers.

Recruiting through the UCL disability support network we conducted semi-structured interviews with eight hard of hearing individuals. Support workers, including a palantypist and disability support officer, were interviewed to understand the wider impact.

At each stage of research, the outputs were analysed and common themes identified using affinity diagrams. This gave us a set of prioritised user requirements to design for.

We found that hearing loss leads to feelings of abnormality. Lip-reading and note taking in class is exhausting and involves a lot of guesswork, and contextual information is important.

Designing the intervention

Brainstorming, sketching and ideation helped us generate multiple ideas. Early concepts were storyboarded and used to gather feedback during interviews. One idea was selected and we began prototyping in paper first - before moving to low-fidelity prototypes.

Early paper prototypes were tested with a hard of hearing tutor and three students at a lip reading class. A low-fidelity prototype was tested during a real lecture with six students and the lecturer as participants. Speech to text interaction was simulated with a clip-on microphone and ASR software which captured text to a shared Google document.

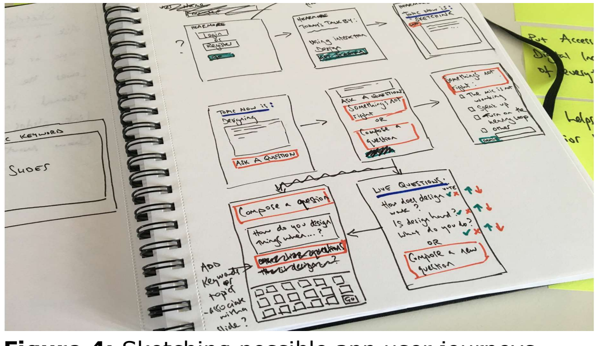

Figure 4: Sketching possible app user journeys

A clickable prototype was used in two further rounds of usability testing with three students who were hard of hearing. Following each round the design was iterated. Usability testing with these humanities students exposed a key assumption in our design; that students would be allowed to take laptops into class with them, but this was not always the case. Another ideation session looked at how to turn ideas for a digital, app based product into a physical, tangible product.

Usability testing and Co-design

Figure 5: Usability testing with hard of hearing student

Figure 6: Co-design session with two hard of hearing students

Figure 7: Role playing with Wizard of Oz prototype test

To help design a physical object and to challenge any internal bias around that, a co-design session was conducted with two hearing impaired students. Using a selection of craft materials such as modelling clay, we worked with the participants to create physical 'devices' to help with the challenges they faced while studying.

In that session, we also tested a Wizard of Oz prototype using HTML into which we fed a scripted 'lecture'. Using Role Playing, one of us read out the scripted lecture whilst someone operated the prototype. To participants it appeared that the spoken words appeared in real time on the screen.

User studies key findings

We found that for most people, being hard of hearing is an invisible disability. Participants told us even close friends and family forget that they can't hear well. Missing critical elements of social interactions led to feelings of abnormality and social isolation. One participant spoke of her acute embarrassment when other students try to talk to her. "I'm a freak" she told us. Above all, they want to avoid drawing attention to themselves and their impairment.

The classroom is exhausting for students who are hard of hearing. They operate under increased cognitive load and they told us that they have little capacity during the lecture to do anything more than lip read and process speech to text.

The lo-fi prototype test findings highlighted the need for accurate transcriptions, autoscroll, readable text, and the option to save the transcription as notes for later.

Usability testing with the clickable prototype emphasized the need de-clutter the screen and ensure the most important feature - the speech to text transcription - was prominent and clear.

Through Wizard of Oz testing we learned that some latency in the display is actually helpful for students who lip read. It enables them to look at the teacher before looking to the text for additional help.

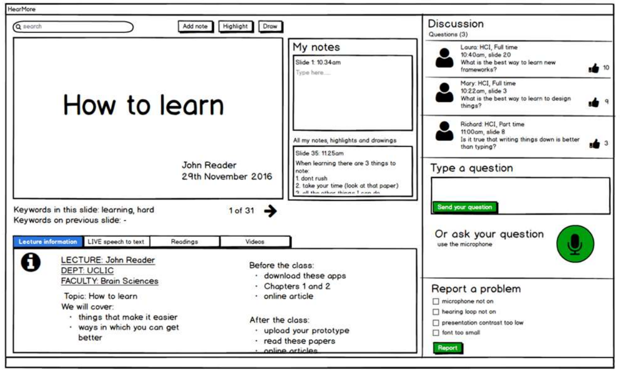

Figure 8: An early version of the clickable prototype. Usability testing identified it as being too cluttered

Co-design made us highly aware of our participants' affective states. In particular, we observed a striking reaction to the playful nature of a snap band bracelet which provoked tactile pleasure in both.

HearMore

The cumulative output of our research and design is HearMore, a live captioning service students can have with them every day. HearMore is designed to fit discreetly into the classroom environment. The HearMore ecosystem includes:

Prototyping

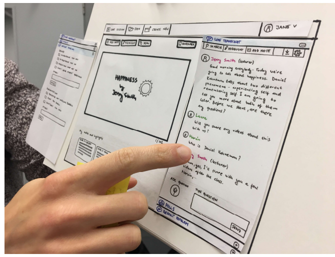

Figure 9: Paper prototyping

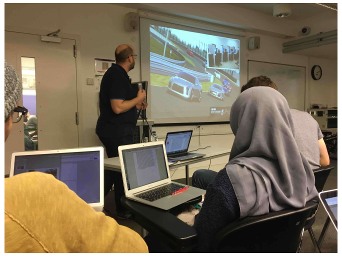

Figure 10: Lo-Fi prototype test during a real class

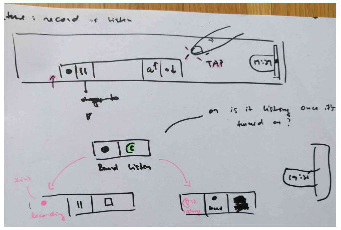

Figure 11: Ideation for a physical product

- **a wearable wristband** that displays text, allowing the user to read what is being spoken in the moment;
- **a mobile app** that stores transcriptions and allows editing after the event;
- **a microphone** that can be worn by a speaker at events or lectures.

The wristband encourages interaction and participation in the classroom environment. Playful interactions have been shown to encourage higher levels of engagement (Golseijn et al. 2015) so the flexible wristband is tactile and pleasurable to touch. It can be worn on the wrist or flattened out on a desk. It is not reliant on bringing a laptop into class and supplements learning without distractions.

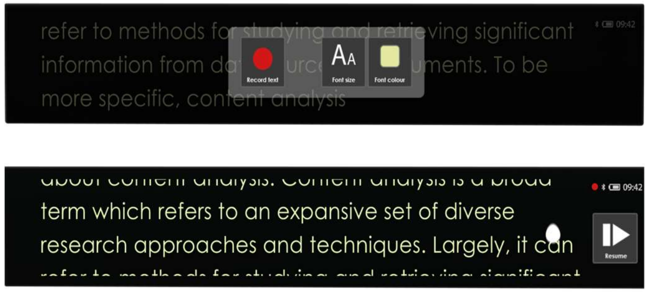

Figure 12: Functions on the wristband

The wristband is deliberately discreet and stylish to avoid singling the wearer out as hearing impaired. It is programmable to recognise and vibrate when someone nearby speaks the wearer's name. Context specific words, or those that are hard to lip read can also be added to the dictionary and are highlighted on the wristband when spoken. Users can configure the wristband to suit their needs, adjusting the font size, colour, number of words per line and the latency with which text appears on the wristband.

The mobile app stores the captioned lectures so users can go back to them later to re-read anything they missed. Users can change the formatting, edit and make additional notes. This feature allows for total concentration and negates the need to take notes during class.

Figure 13: The HearMore ecosystem

The HearMore wristband can pick up speech within a range of 1.5 metres (Bauman, n.d.) but in larger spaces such as lecture theatres the additional microphone worn by the speaker helps to avoid noise distractions and ensures more accurate transcriptions.

Discussion

The purpose of this study was to create an intervention for hard of hearing students that would help to level the playing field in the classroom environment. We conducted eight interviews with people who are hard of hearing and three rounds of design, testing and iteration. We found that for many people being hard of hearing is an invisible disability which has a negative

The stylish and playful wristband

Figure 14: The wearable wristband

Figure 15: The wearable wristband

impact on their ability to participate in mainstream education.

HearMore helps hard of hearing students 'hear' more of what's being said in a learning environment. It is a portable, flexible solution which addresses a real world problem, releasing hard of hearing people from the cost and constraints of booking palantypists. It provides instant support when needed through automatic speech recognition software.

Failing to limit the user group caused challenges early in the design process, leading to an over-complicated design that met no-one's needs. Narrowing the target user group to those who are hard of hearing resolved this, allowing us to prioritise the key requirements and simplify the design.

Our sample size was not as large as we would have liked, despite obtaining some participants from the target user group, due to time constraints. A larger sample size, as well as a longer time-period over which to design and test, would lead to increased confidence in the robustness of our approach. Future work would include more research with a wider range and geographic spread of hard of hearing participants. Additional testing with a high-fidelity prototype would also enhance the external validity of our design.

Given the recent improvements in ASR (Hinton et al. 2012) fueled by an increased interest in voice based technologies, we are confident that the technical challenges around robustness to noise can be overcome, allowing our design to provide support to millions of hearing impaired people all over the world.

Contribution

Here we have presented a new approach to real-time capture of content for people who are hard of hearing and based in a classroom setting. We have shown through usability testing and co-design sessions how this approach meets the needs of those in the lecture room who are struggling to hear, but it can also be tailored to conferences and other events too. We have also presented a novel display interaction by creating a slap band-like flexible display which is easy to put on and take off, and also has an adaptive screen size. This approach encourages greater participation in lecture content and drives up engagement for all, which could lead to better outcomes for students of all levels.

Acknowledgements

We would like to thank our mentors, Ben Bengler and Judith Borghouts, for providing helpful resources and advice. We would also like to thank students and teachers from the lip reading class, and all the participants who helped us to test our prototypes and took part in our co-design sessions. The feedback was invaluable. We would especially like to thank Rachel Benedyk and Paul Marshall for their help and advice.

References

1. Action on Hearing Loss, 2015. Hearing Matters. Accessed 30 November 2016. Available from: https://www.actiononhearingloss.org.uk/your-hearing/about-deafness-and-hearing-loss/statistics.aspx
2. Blackwell D. L., Lucas J. W., Clarke T.C. Summary health statistics for U.S. adults: National Health Interview Survey, 2012. National Center for Health Statistics. Vital Health Statistics 10 (260). 2014. Available from:

http://www.cdc.gov/nchs/data/series/sr_10/sr10_260.pdf

3. Packer, L. 2015. How hearing loss affects school performance. Healthy Hearing. Accessed 30 October 2016. Available from:

   http://www.healthyhearing.com/report/52433-How-hearing-loss-affects-school-performance

4. Elliot, Stinson & Easton, 2008, College Students' Learning with C-Print's Educational Software and Automatic Speech Recognition. Paper presented at the American Educational Research Association Annual Meeting, March 25, 2008, New York, NY.

5. Fan, X., Luo, W., Menekse, M., Litman, D., & Wang, J. 2015. CourseMIRROR: Enhancing large classroom instructor-student interactions via mobile interfaces and natural language processing. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pp. 1473-1478.

6. Bernstein, M. S., Brandt, J., Miller, R. C., & Karger, D. R. 2011. Crowds in two seconds: Enabling realtime crowd-powered interfaces. In Proceedings of the 24th annual ACM symposium on User interface software and technology, pp. 33-42.

7. Murphy, M. J., Miller, C. D., Lasecki, W. S., & Bigham, J. P. 2013. Adaptive time windows for real-time crowd captioning. In CHI'13 Extended Abstracts on Human Factors in Computing Systems, pp. 13-18.

8. Lasecki, W., Miller, C., Sadilek, A., Abumoussa, A., Borrello, D., Kushalnagar, R., & Bigham, J. 2012. Real-time captioning by groups of non-experts. In Proceedings of the 25th annual ACM symposium on User interface software and technology, pp. 23-34.

9. Lasecki, W. S., & Bigham, J. P. 2012. Online quality control for real-time crowd captioning. In Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility, pp. 143-150.

10. Tran, D.T., Vincent, E., & Jouvet, D. 2015. Discriminative uncertainty estimation for noise robust ASR. In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5038-5042.

11. Golsteijn, C., Gallacher, S., Koeman, L., Wall, L., Andberg, S., Rogers, Y., Capra, L. 2015. In: Verplank, B and Ju, W and Antle, AN and Mazalek, A and Mueller, F, (eds.) TEI '15: Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction, pp. 201-208.

12. Bauman, N. n.d. Center for Hearing Loss Help. Five Facts About Hearing Loss and Hearing Aids That Make Using Loop Systems Important. Accessed 3 December 2016. Available from: http://hearinglosshelp.com/blog/five-facts-about-hearing-loss-and-hearing-aids-that-make-using-loop-systems-important/

13. Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., Kingsbury, B. 2012. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. In IEEE Signal Processing Magazine, Vol.29 (6), pp.82-97