

SPeach: Automatic Classroom Captioning System for Hearing Impaired

Andres Cedeño, Riya Fukui, Zihe Huang, Aaron Roe, Chase Stewart, Peter Washington

Problem Definition

Over one in seven Americans have some form of hearing disability [2]. Hearing disabilities are a significant barrier in classroom learning; studies have shown that Deaf and Hard of Hearing students lag behind their hearing peers in education [1]. In order to improve comprehension and understanding in lecture settings, SPeach creates real-time speech-to-text transcription with the additional capability of correcting the output for improved accuracy. SPeach will leverage on a student volunteer to perfect a lecture transcript in real-time, with the option for the professor to finalize edits. Other students will be able to view the transcript in real time. The transcripts can then be saved for later. Through our simple, yet powerful automatic captioning system, SPeach improves on existing accommodations for those with a hearing disability.

Existing Competition

Manually-typed remote captions

Remote captioning attempts to provide accurate transcripts for classroom lectures, without requiring the captioner to be present, by using an web-based audio connection. However, a trained professional is needed for accurate and quick captioning and the specialized equipment makes the service more expensive than necessary. In addition to the high price, the service can be unreliable. If the acoustics of the room are poor or if the captioner has problems connecting through the software, then the students will be without captions.

Sign-language Translator

Using sign interpreters is an interactive, human method of translation that is effective for small groups. However, hiring an interpreter is expensive and less convenient for institutions because of the necessary specialized training. In university lecture settings, interpreting is taxing; in some cases, multiple interpreters are required to take turns interpreting. Interpreters are expensive for universities to hire and oftentimes are not familiar with the more technical nomenclature, leading to long delays in translation. Using a sign-language translator also leaves no record of the past, increasing the chance of the student being left behind.

Feedback from Potential Customers

Jeff Anderson, Assistant of Director of Rice's Disability Support Services

During his time at Rice's Disability Support Services (RDSS), Mr. Anderson has made multiple arrangements for the hearing impaired. On one occasion, RDSS hired a remote captioner for a student. However, the acoustics in the room were of such low quality that the captioner could not hear well. Also, the software for connecting remotely was difficult to use; it required more than fifteen minutes of preparation time. In addition, the captioner was unfamiliar with the technical nomenclature used in the student's classes, so the captioner spelled out those terms phonetically. All in all, the system was inconvenient and the resulting transcript was useless for the student. Currently RDSS requests volunteers to take notes for students who are unable to take notes, including those who are hearing-impaired. Mr. Anderson suggested using a single volunteer system to handle any error corrections.

He was impressed by the range of vocabulary that SPeach's demo was able to recognize, and overall gave us positive feedback about our project; he said he could see it being useful for RDSS. After seeing our current demo, he was impressed by how easy it was to set up our system especially in relation to the remote captioning. He also like the feature where the viewer could request corrections.

Rice Computer Science professors

We met with a few of the Rice Computer Science faculty who teach massive open online courses (MOOC), to see if our technology could be of use. As the lectures are recorded, live captioning would not needed, but we were interested in seeing if we could give the videos automatic captions. What is currently being done for websites like Coursera is that they pay captioners to make closed captions. The professors said the idea was a good idea, and is actually what Rice uses for some of their inhouse videos; they have a captioning system that generates automatic captions, and then they have someone perform corrections. They said that if we ever produced a prototype of our automatic captioning and correction system tailored for recorded videos, they would like to test it against Rice's.

SPeach System Architecture

Mobile Phone - Andres Cedeño, Peter Washington

- Android application that can begin the session and handles the speech handling and recognition. Sends the converted speech to the server in text form.
- Clip-on microphone is used.

Front End - Riya Fukui, Chase Stewart

- Website with separate URLs for the volunteer editor and the transcript viewers.

- The viewer's page updates as edits and additions are made.
- The editor's page includes a text pane where they have the ability to make edits.

Back End - Zihe Huang, Aaron Roe

- Web server that handles the concurrent shared transcript, edits from the volunteer, and additions from the speaker.

Current Status of SPeatch

Mobile Phone

- Automatic transcription of speech using an Android phone as the input device.
- Google's free speech to text API is used.
- The translated text is sent over to the viewer's webpage.

Front/Back End

- A volunteer can access a web page (“~/edit”) to edit the incoming transcript.
- Multiple students can access a web page (“~/view”) to view the incoming transcript as well as its correction on their laptops in real time.
- Viewers can request clarification by clicking on text which signals the editor with two stars
- Viewer can see translations of the transcript in 40+ major languages

Session Creation

- We now allow for the creation of multiple broadcast sessions. It is easy for a professor, student, and viewer to access sessions, as they only need the session ID.

Public Website

- SPeatch has been deployed to the web, and can be accessed at <http://speach.herokuapp.com/>.
- The website allows the user to easily access SPeatch on the server side, and provides a link for the user to download the mobile app.

How To Use SPeatch

To start off, navigate to the SPeatch website: <http://speach.herokuapp.com/>.

1) *Create a session.* To start a new SPeatch session, enter a session name in the left most input field and click "Start a New Session." Once you press the button, you will see a blank page. This is the viewers' page. First notice that you have a session ID. This is unique to your session. Viewers will use this ID to join your session.

2) *Join a session.* If you are an editor or a viewer, please obtain a session ID from the person who created the session. Enter the session ID and click on "Join a Session." Once you press the button, you will also be directed to the viewers' page.

3) *Start the mobile app.* If you are a speaker, please start the app. Click on the "Start Lecture mode" button, then enter the session ID and click on "start" button. You should hear a little jingle. The sound indicates that the app is ready for you to speak. Please speak into the phone.

4) *Edit an utterance.* Now back on the website, click on the "editor" link. You will see a speech transcript. Notice that you can edit the lines. This is where the editor corrects and errors in the speech transcription system.

Experiments

We ran several experiments with SPeach. The experiments involved two participants per trial - a viewer and an editor. For each trial, a speaker read an excerpt from a text filled with technical nomenclature. A hearing impaired student is simulated by placing the student in a separate room where sound from the lecturer cannot reach. The participants in each trial had five rotations:

1. Both participants view the transcript in a separate room from the speaker. No corrections are done to evaluate comprehension with just the raw text.
2. One participant is in the same room as the speaker and makes edits based on stars added automatically in front of the text. Stars are added based on confidence levels provided by the Google API. The other participant passively observes the transcript in real time in a separate room.
3. One participant is in the same room as the speaker and makes edits on request. The other participant observes the lecture and signals the editor which lines he wants corrected. Stars appear in the editor's screen signaling a correction request

Both participants do roles 2 and 3.

Following each trial, we provided a survey to each of the participants, asking them to rate their experience performing their tasks in terms of comprehension, stress level, and division of attention. We averaged these scores from all of our participants.

| <i>Experiment Group</i> | <i>Familiarity with topic (5 - very familiar)</i> | <i>Comprehension (5 - a lot)</i> | <i>Stress level (5 - relaxed)</i> | <i>Division of attention (5 - mostly lecturer)</i> | <i>Correction rate (5 - fast to correct)</i> |
|--------------------------|---|----------------------------------|-----------------------------------|--|--|
| control (no text edited) | 2.56 | 2.67 | 3.78 | x | x |
| viewers | 2.31 | 2.79 | 3.76 | x | x |
| editors | 1.70 | 2.65 | 2.90 | 2.40 | 1.15 |

It is clear that the human requested method of correction was preferred by the viewer, although it was difficult for the corrector. It is also clear that correction impose additional cognitive load to the editors.

General feedback we have received through these trials include:

- Familiarity with the topic covered (such as being a Teacher's Assistant) significantly affects the ability and the quality of correction
- It is hard to listen to the lecture and correct the incoming lines at the same time.
- As a viewer, the transcript in addition to the audio could enhance the learning experience.
- Editors reported correcting errors quickly and following the lecturer.

Future Directions

Google Glass

A future direction of SPeach would be to port the website as a Google Glass application. This would minimize the "additional cognitive load" on the users, eliminating the need for switching between looking at the lecture and a computer screen. Instead the students' gaze may be kept on the presenter.

Audio Broadcasting

Some students may benefit from having amplified sound directly broadcast into the ear through earphones as they follow along with the lecture. Local audio broadcasting could harness audio coming into the phone and send it to the student.

References

[1] Antia, Shirin D., Kathryn H. Kreimeyer, and Susanne Reed. "Supporting students in general education classrooms." *The Oxford handbook of deaf studies, language, and education 2* (2010): 72-92. Web. 16 March 2014.

[2] "Quick Statistics." National Institute on Deafness and Other Communication Disorders. U.S. Department of Health & Human Services, 16 Jun 2010. Web. 24 Feb 2014.