

---

**EE/CprE/SE 491 BIWEEKLY REPORT 1**

**August 30 - September 13**

**Group number: sdddec21-06**

**Project title: DigiClips Media Design**

**Client: DigiClips**

**Advisor: Ashfaq Khokhar**

**Team Members/Role:**

**Sam Massey - Assignment planning, Research, Work on speech-to-text**

**Tyler Johnson - Planning and implementing test cases**

**Maxwell Wilson - Primary point of contact with client, Research, Work on speech-to-text, Team Leader**

**Max Van De Wille - Documenting architecture changes, Research, Work on video-to-text**

---

o **Weekly Summary**

This past week, our team tested out multiple different platforms for video-to-text and speech-to-text elements. One of the main forms of speech-to-text translation looked into was AWS speech-to-text. After additional research and discussion with DigiClips, we will be moving forward with development using the open sourced speech-to-text software Deep Speech. For video-to-text, our team, along with the client DigiClips decided to use Tesseract, an OCR. We have begun development as well to deliver results for video-to-text along with the speech-to-text to the client at our weekly meeting on Friday.

o **Past week accomplishments**

Max Wilson:

- Researched AWS Transcribe since our client showed interest in trying it out.
- Set up a small proof of concept to see how AWS Transcribe would work using our clients data
- Tested multiple audio files using our existing speech-to-text model and compared to AWS Transcribe

- Discovered AWS Transcribe is incredibly accurate and far better than our own custom speech model.
- Refactored last semester's code to use a production-ready web server Uvicorn. This will make our app faster and improve the way it handles requests. Also changed the web framework from Flask to FastAPI because Flask can be slower than FastAPI in some instances and works better with Uvicorn.

Sam Massey:

- Researched AWS Transcribe usage cost and compared/calculated total cost for running transcriptions on a massive scale.
  - Learned the total possible price for running data from DigiClips through this software.
  - Price was reported to DigiClips, though AWS was a cheaper option, it still will be too costly to use in a final product.
- Set up Deepspeech to initiate development of speech-to-text.using C++.

Max Van De Wille:

- Tested different OCR parameters for video-to-text software
- Downloaded sample footage from client and ran existing video-to-text implementation for benchmarking

Tyler Johnson:

- Compared multiple different options and configurations of software to determine best configurations
- Began creating test sets for speech-to-text and video-to-text to create a standard comparison point for further tests

## ○ **Pending issues**

- No unified/standardized testing set to compare performance of one iteration to the next makes it hard to benchmark progress/performance improvements.
- One of the main pending issues our team as well as the team at DigiClips is facing will be ensuring accuracy in our speech-to-text and video-to-text.
- Speech-to-text custom model is missing chunks of audio for some reason. Some debugging needs to be done to determine the problem.

## o Individual contributions

Team Member	Contribution	Weekly Hours	Total Hours
Sam Massey	Presentation, AWS Research, Deepspeech setup	6	12
Tyler Johnson	Presentation, test benching	6	12
Maxwell Wilson	AWS Testing, presentation	7	14
Max Van De Wille	Video-to-text development, generating benchmark samples for client	6	12

## o Plans for the upcoming week

Max Wilson:

- Look into the speech-to-text chunking issue and work on making various improvements to the speech-to-text part of the application.
- Work on improving the organization of the speech-to-text application. Code is a bit disorganized so improvements could be made.

Sam Massey:

- Continue development of speech-to-text software.
- Analyse code to improve speed when running.
- Test speech-to-text code with DigiClips audio recordings to test speed and accuracy.

Tyler Johnson:

- Continue curating a set of audio and video clips to be representative of usual expected input.
- Transcribe curated audio and videos to provide a human-accurate model to base all accuracies off of for testing purposes.

Max Van de Wille

- Test custom training tesseract OCR with samples of news footage to ideally improve accuracy
- Brainstorm methods to eliminate duplicates/noise from recovered text artifacts
- Test different image pre-processing methods (i.e. canny edge, grayscale, binary thresholding)