

---

# Metadata Works From Home

How Student Workers Continued to Work  
While Increasing Collection Accessibility

---

Emily Vinson [evinson@uh.edu](mailto:evinson@uh.edu)  
Andrew Weidner [ajweidner@uh.edu](mailto:ajweidner@uh.edu)  
University of Houston Libraries

# Project Background

## Transcription Accuracy and Cost Study

- Internal grant awarded late 2018
- Examined accuracy of various transcription methods

# Transcription Accuracy and Cost Study

## Methods

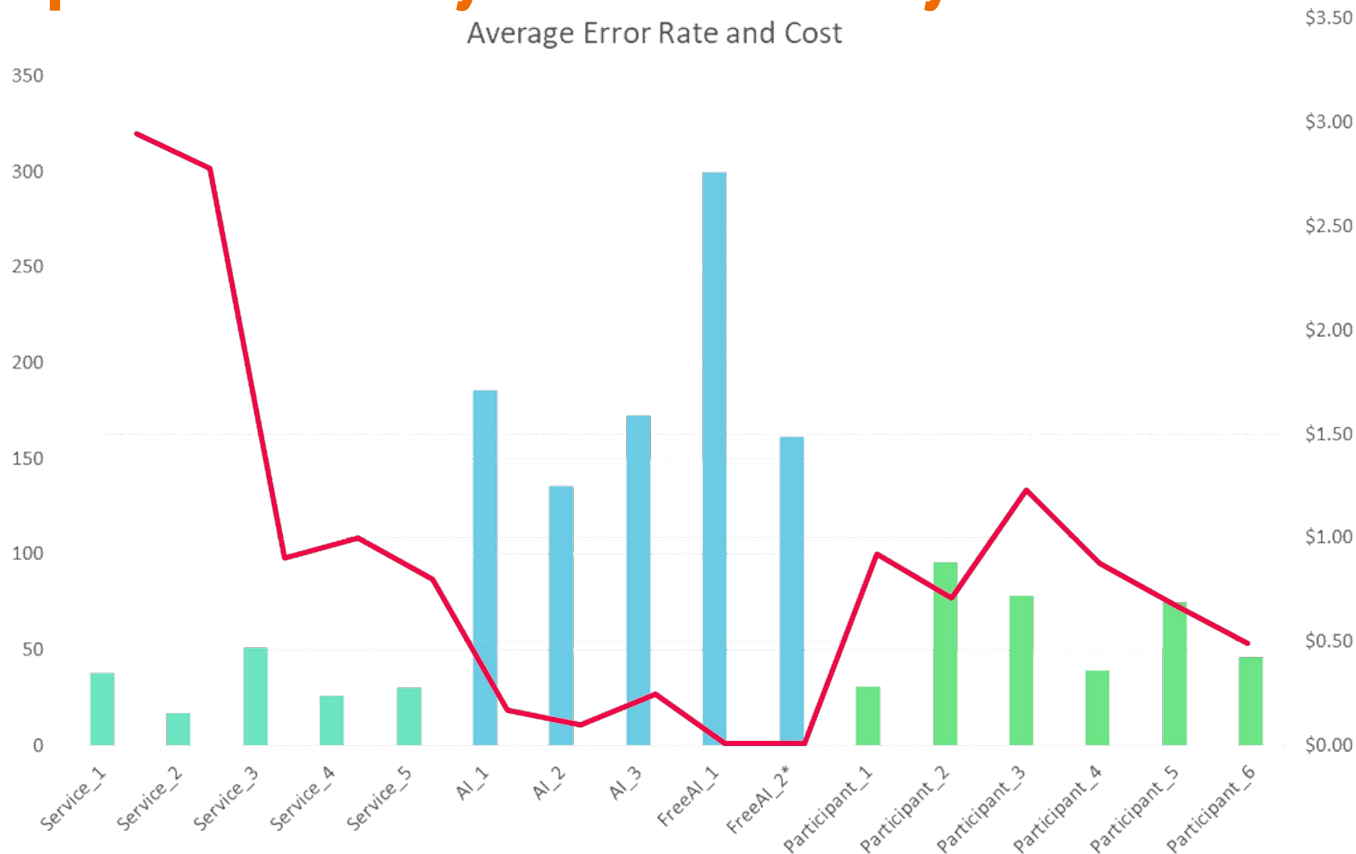
- Professional transcriptionist services
- Online Artificial Intelligence-based services
- Free “hacks”
- Untrained study participants

# Transcription Accuracy and Cost Study

## Sample Videos

- Single speaker, clear, scripted speech
- B-roll, crosstalk, off-mic speech
- Multi-person, off-mic and unclear speech
- Bilingual scripted speech
- Background music, repetitive language

# Transcription Accuracy and Cost Study



# Intro to Closed Captions

Timestamp in hh:mm:ss:ttt format

Line length and line breaks will be reflected in caption formatting

12  
00:00:52.440 --> 00:00:58.440  
sit tight and not move. They  
decided if one person moved, it  
would be a big panic.

14  
00:00:58.440 --> 00:01:02.690  
So everyone said "Okay, we  
won't sell our houses. We'll  
just stay right here,"

16  
00:01:02.690 --> 00:01:06.680  
and that was kind of what they  
did.  
Interviewer: Okay, cut.

18  
00:01:06.680 --> 00:01:09.960  
Man 1: Marker.  
Man 2: Two mark.  
Interviewer: How did your uh,



# Timeline

**March 16:**

UH Classes Go Online; Most  
Library Departments  
Transition to Remote Work

**March 18:**

Libraries Admin  
Approves Project

**March 19:**

First Project  
Managers Mtg

**March 24:**

Harris County Stay Home  
Order; All Library Depts  
Transition to Remote Work

**March 23:**

Second Project  
Managers Mtg

**March 26 & 27:**

Zoom Training  
with Students

**Transcript  
Correction  
Work  
Underway**

**April 10:**

Zoom Training  
with QC Students

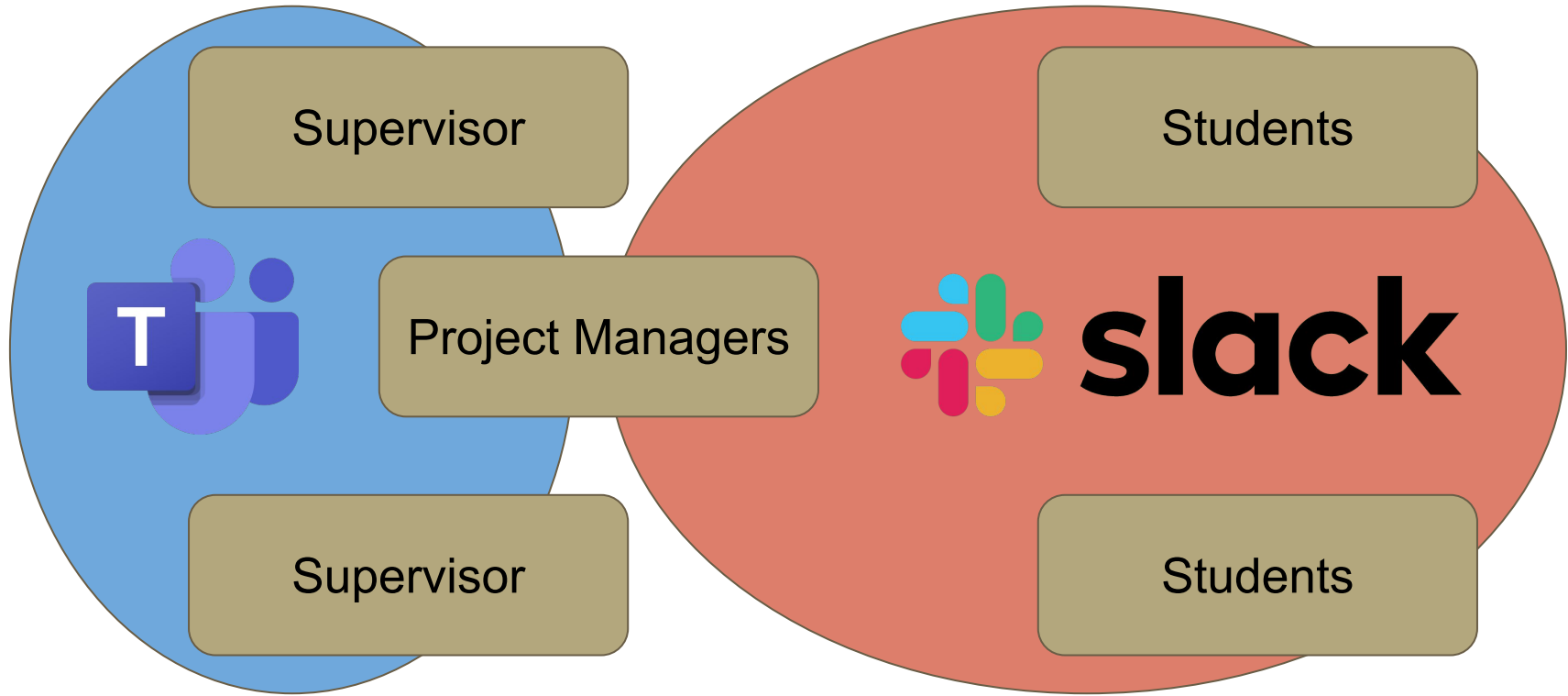
**Quality Control  
Work Underway**

**WEEK 1**

**WEEK 2**

**WEEK 3**

**WEEK 4**





# Communication Tools

**Microsoft Teams:** Group Chat with Supervisors

**Slack:** Group Chat with Students

**Zoom:** Planning & Training Meetings

**Email:** Onboarding Information & Personnel Issues

**Google Calendar & Drive:** Student Schedules & Documentation

# Workflow Tools

**Sonix:** Transcript Generation

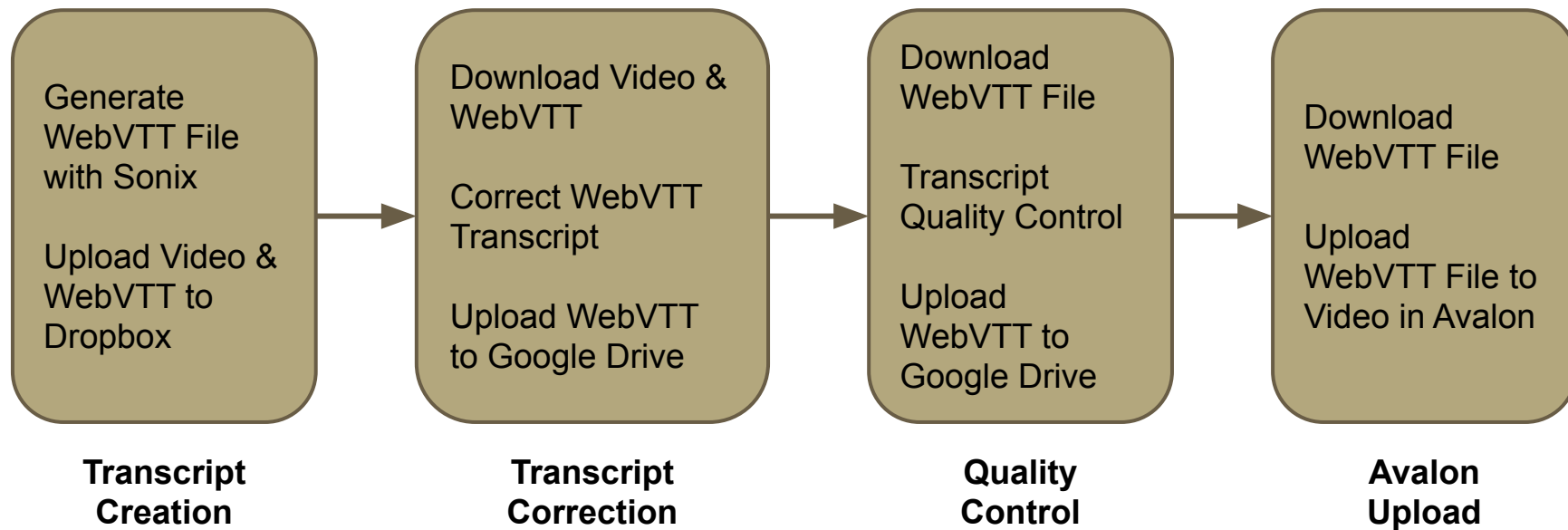
**Dropbox:** Video & Transcript Distribution

**Google Drive:** Transcript Versioning & Progress Tracking

**VLC Media Player:** Video Playback

**Atom:** Transcript Editing

# Workflow Overview



# Documentation

## Communication

- Email
  - We will occasionally communicate with supervisors and students via email
  - Students, please make sure your email address has been added here: <https://docs.google.com/spreadsheets/d/1JXMSy3D-TaAUVHscU18LVBdR6dFC09D7YnD6d4ucl7vUeSshmg/edit#gid=0>
- Microsoft Teams (supervisors)
  - Supervisors will be added to a Transcription Project team to coordinate administrative aspects of the project
- Google Calendar
  - Students will maintain their work schedule on a [Google Calendar](#)
  - Supervisors will have access to the calendar
- Slack
  - We will primarily communicate regarding production activities via [Slack](#).
  - Students will receive an invitation to the #av-transcribers channel at the email address you provided
  - This is the place to ask questions and coordinate work with others

## Documentation

- Google Drive
  - <https://drive.google.com/drive/folders/1Qq06dJnFIS9w3Y1JqA1h0uQCPn7u6nshwng>
  - All project documentation and production tracking will be available in the Google Drive folder linked above
  - All students and supervisors will have edit access
- Dropbox
  - Videos and WebVTT files will be made available for download via Dropbox
  - Video files are large. Please make sure you have sufficient hard drive space.

## Video Playback

- VLC Media Player
  - <https://www.videolan.org/vlc/index.html>
  - VLC has playback controls that make transcription work easier. Please download and install prior to training. Staff are available to assist with any issues.
  - Transcript Correction and QC should be done on your local machine with downloaded videos using VLC.
- Avion
  - <https://av.lib.ub.edu/>
  - All project videos are also available for streaming online if needed

## Text Editing

- WebVTT files will be edited in a simple text editor from the following approved list:
  - TextEdit (Mac)

Draft: March 24, 2020

## Closed Caption Project

### Getting Started - Technical Requirements

#### THE BASICS

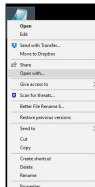
1. Use TextEdit (Mac), Notepad (Windows), or Atom (cross-platform) for editing WebVTT files or creating transcripts
2. Use [VLC](#) to play video, as this allows for slowing playback speed, jumping 10 seconds forward/back, and the addition of WebVTT files to QC for closed caption sync. It is free to download and works on both PC and MAC: <https://www.videolan.org/vlc/index.html>

#### DETAILED INSTRUCTIONS

##### Opening WebVTT files

In this project, you will be working with two types of files: videos and transcript files (called [WebVTT](#)) files, both downloaded from the project Dropbox.

When opening the transcript file, right-click, navigate to "Open with" and select Notepad (on PC) or TextEdit (on Mac), or Atom. All transcript editing will take place here.



Draft: March 25, 2020

## TRANSCRIBING GUIDELINES FOR CLOSED CAPTIONING PROJECT

Transcribing is a tedious process, but also very important. The transcripts you will be working with are formatted in order to add closed captions to make audiovisual materials accessible to people with hearing impairments. This file format is called WebVTT and includes timestamps on every line, ensuring the transcript is synced to the audio. These transcripts can also provide researchers a useful format to access information (via keyword search) of what is covered in a recording.

The files you are working with were machine-generated, and will include many errors - particularly with people and place names, homonyms (two words that sound the same but are spelled different), and in case when the speaker or audio is unclear. Correcting a transcript for a one-hour recording may take much longer than an hour, it is ok to take breaks and/or to not complete correction of a recording transcription during a single work shift! *Accuracy is more important than speed.*

Keep in mind that these are archival recordings, most created between 30 - 70 years ago. There may be some language that is unfamiliar. Other language may be dated and offensive. Please know that we simply present the archival item authentically, and do not edit out or change the materials. If you are uncomfortable working on any material, feel free to move on to another recording; just let Emily Vinson ([emvins@ub.edu](mailto:emvins@ub.edu)) know that you have done so.

For Technical Instructions see [Getting Started - Technical Requirements](#)

## Table of Contents

[Basic Transcription Guidelines](#)

[Formatting and Punctuation Style Guidelines](#)

[Speaker IDs](#)

[Abbreviations](#)

[Acronyms](#)

[Commas](#)

[Contractions](#)

[Dashes](#)

[Colons](#)

[Dates](#)

[Monies](#)

[Numbers](#)

[Quotes](#)

[Tips](#)

# Project Successes

- Phase 1 caption correction complete for over 30% of all online videos
- Raised awareness of accessibility issues within the UH Libraries and among student employees
- Continue to keep students employed remotely
- Developed effective workflow to continue creating accessible video content

## **In Memoriam**



**Crystal Cooper (1981-2021)**

# Questions