AMPlifying AV

Next Steps for the Audiovisual Metadata Platform

CNI Spring 2022
March 29, 2022

Jon Dunn / Indiana University / @jwdunn
Shawn Averkamp / AVP / @saverkamp
Project website: https://go.iu.edu/amppd

AMP
AUDIOVISUAL METADATA PLATFORM

Outline

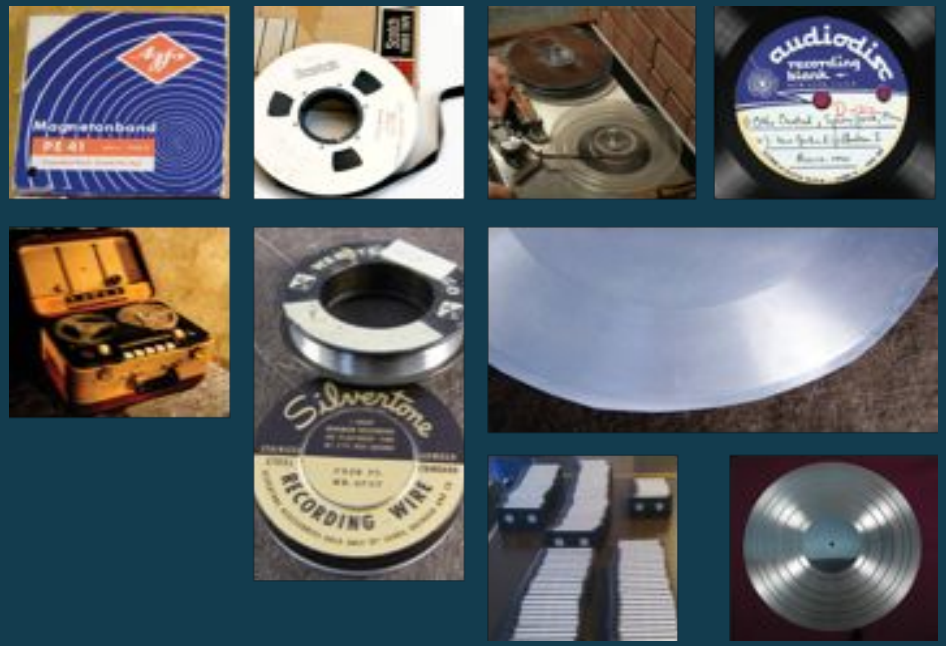
- Background
- Phase 2 accomplishments and lessons learned
- Phase 3 goals and progress
- Next steps

AMP
AUDIOVISUAL METADATA PLATFORM

Background

The Challenge

- Growing AV collections
  - Legacy formats
  - Explosion of born-digital
- Increased expectations for access
- Insufficient metadata
- Limited resources for cataloging



AMP
AUDIOVISUAL METADATA PLATFORM

The Opportunity

- Mass digitization approach extended to AV
  - “Digitize first”
  - e.g. IU’s Media Digitization and Preservation Initiative (MDPI): mdpi.iu.edu

- Emergence and continued improvement of machine learning and other automated tools

- How can we leverage the best of automated tools and human expertise in flexible and configurable ways?
  - Diverse collections demand diverse workflows

AMP
AUDIOVISUAL METADATA PLATFORM

AMP: The Vision

- Open source software platform to support metadata creation for AV collections
- Design and execute workflows combining automated and human steps
- Integrate multiple “Metadata Generation Mechanisms” (MGMs)
  - Automated, manual
  - Local, HPC, cloud
- Delivery of metadata to variety of target systems, e.g. online access systems (Avalon, Aviary), library catalogs, etc.

AMP
AUDIOVISUAL METADATA PLATFORM

Work to Date

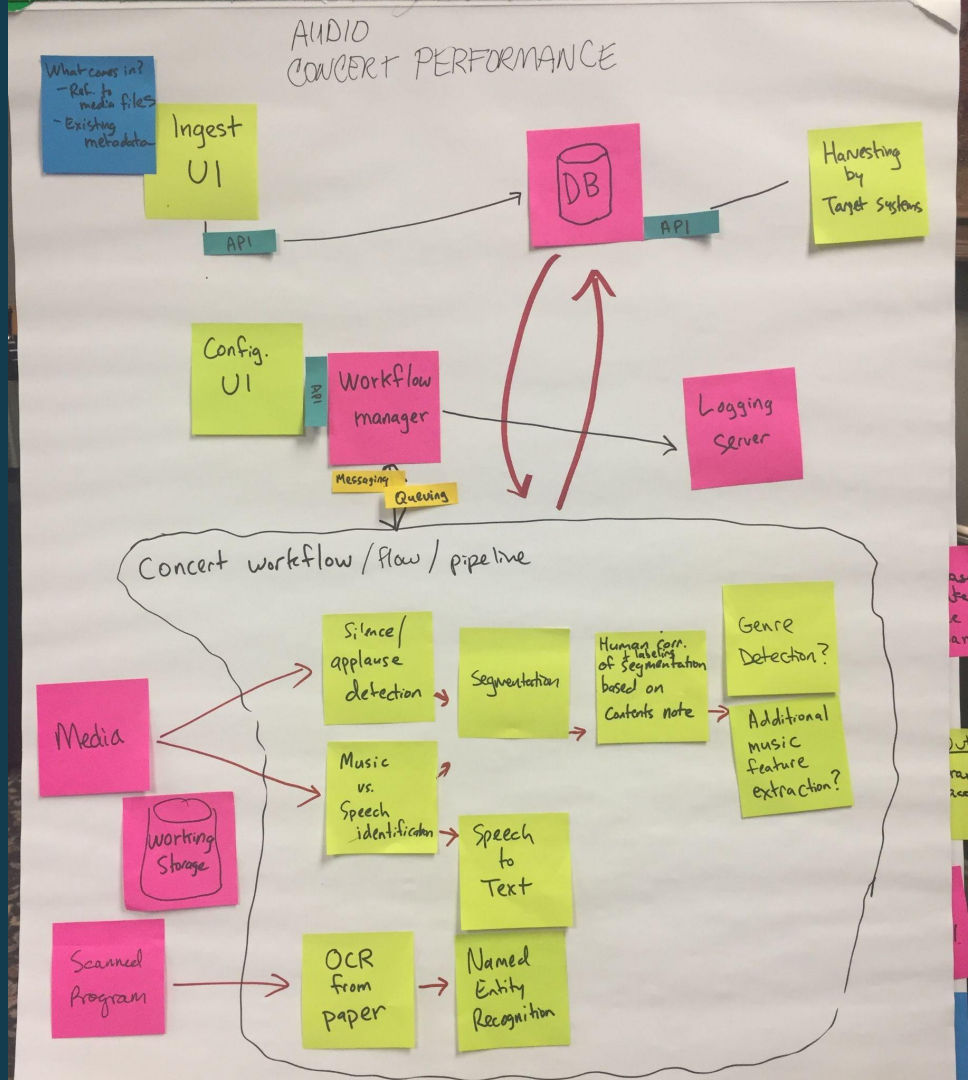
- 2017-2018: AMP Planning Project (Phase 1)
  - Developed initial AMP concept
  - Convened technical architecture planning workshop
  - White paper: *AMP Planning Project: Progress Report and Next Steps*

- 2018-2021: AMPPD - AMP Pilot Development (Phase 2)
  - Built proof-of-concept AMP system
  - Pilot tested using two collections from IU and one from NYPL
  - White paper: *AMPPD Final Project Report*

- **2021-2022: AMP Phase 3 (current)**
  - Additional system development
  - Packaging and deployment work
  - Testing with additional collections from IU and NYPL

AMP
AUDIOVISUAL METADATA PLATFORM

AMPPD (Phase 2)



Phase 2 Accomplishments

- Developed AMP application
  - Web-based user interface: Java, Vue.js, React.js
  - Workflow management and execution: Galaxy
- Evaluated and implemented MGMs (Metadata Generation Mechanisms)
  - Automated and Human MGMs
  - Evaluation criteria
- Tested with 100 hours of audio and video from each of three collections
  - Indiana University Archives
  - Indiana University Cook Music Library
  - NYPL Gay Men's Health Crisis collection

AMP
AUDIOVISUAL METADATA PLATFORM

AMP Dashboard

Start a new workflow

Unit | Collection | Workflow | Output | Date range | Submitter | External ID | Item | Primaryfile | Step | Status | Show Relevant Results Only

Export to CSV

Show 10 Entries

Prev 1 2 3 4 5 6 7 8 9 ... 15 Next

Search

Date	Submitter	Collection	External Source	External ID	Item	Primaryfile	Workflow	Step	Output	Status
2022/03/24 01:58:40 pm	amp@iu.edu	Admin Collection			Lunchroom Manners	Lunchroom Manners	MW - Transcription and NER with HMGM for NER	transcript_to_webvtt	web_vtt	Complete
2022/03/24 01:55:03 pm	amp@iu.edu	Admin Collection			Lunchroom Manners	Lunchroom Manners	MW - Transcription and NER with HMGM for NER	ner_to_csv	amp_entities_csv	Scheduled
2022/03/24 01:51:40 pm	amp@iu.edu	Admin Collection			Lunchroom Manners	Lunchroom Manners	MW - Transcription and NER with	hmgm_ner	task_info	Sched

Tools

search tools

Inputs
Get Data
Send Data
Audio Extraction
Segmentation
Speech to Text

AWS Transcribe Speech to Text Transcribe speech to text via AWS Transcribe

Gentle Forced Alignment Align an audio file's speech with a transcript

Kaldi Speech to Text Local Kaldi Speech-to-Text transcription

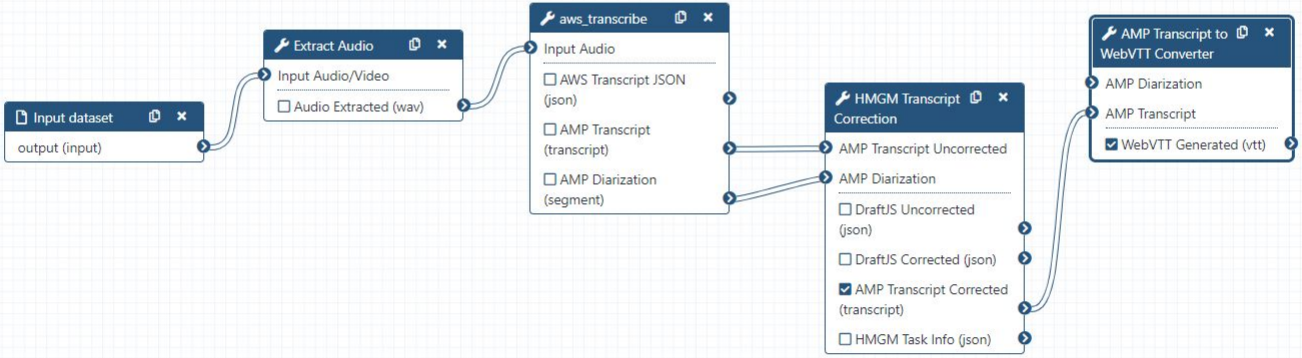
Kaldi on HPC Kaldi Speech-to-Text transcription on IU's HPC

AMP Transcript to WebVTT Converter Convert AMP transcript/diarization to WebVTT

Vocabulary Tagging Tag relevant words in a transcription with a provided vocabulary

Named Entity Recognition
Video Indexing
Shot Detection
Facial Recognition
Video Optical Charater Recognition
Human MGM Editor
Obsolete MGMs

AMP - Transcript-HMGM



100%

AMP Transcript to WebVTT Converter
Convert AMP transcript/diarization to WebVTT (Galaxy Version 1.0.0)

Label

Add a step label.

Step Annotation

Add an annotation or notes to this step. Annotations are available when a workflow is viewed.

AMP Diarization

Data input 'amp_diarization' (segment)

AMP diarization input with speaker diarization info to be added to VTT

AMP Transcript

Data input 'amp_transcript' (transcript)

AMP transcript input with speech-to-text info to be added to VTT

Email notification

No

An email notification will be sent when the job has completed.

Output cleanup

No

Upon completion of this step, delete non-starred outputs from completed workflow steps if they are no longer required as inputs.

AMP
AUDIOVISUAL METADATA PLATFORM

Transcript Editor

Complete | Reset | Save and Close

How does this work?

00:03:21:11 | 00:32:20:00 × 1

SPK_1	00:01:10	Chancellor Wells. Thank you for inviting us out this afternoon to visit with you. Uh the community of course is very closely tied with Indiana University. And to me, I guess you are Mr Indiana University. We like this afternoon, if we could to maybe visit with you concerning the history uh your comments about Bloomington. I know you came here as a student in like 1921 Tell us what bloomington was like in 1921
SPK_0	00:01:38	booming in 1921 was of course much smaller than it is now, it was much more typical of an indiana uh county seat town than it is now. Uh the growth since never fails to amaze me that I drive out around the growth of population, the growth of of of the extent of the city, the the building and so on. Uh But it was it was it was a thriving, is uh um active county seat center in those days with the close relationship between the university on the one hand and the downtown leadership particularly on the other, I found it very well it was a beautiful place then as it is now uh the there were certain features of that period that especially attractive. The the the fifth Avenue or kirkwood coming from town to town was arts over with big trees all the way out, It was rather rural and even greener than it is now possible. The old campus here was not quite a stick with trees as it is now 50 or 60 years made a difference. They the trees grow like all the rest of us and they all they they not only grow taller, they grow around there, don't
SPK_1	00:03:14	we? All
SPK_0	00:03:16	the uh bloomington had in those days, as I remember the very strong and active leadership group down down men like heavily and Graham Bill Graham, uh uh blame, brad, cute judge Wilson, the Haze Buskirk Morris, Riley, the whole group of very strong, all the Adams is, the whole addams family were both, both the elder Adams were alive at that time, it was a very strong and active leadership group.
SPK_1	00:03:59	The students in those days did they, were they involved as much in the community as we see the students in the last 15 years where they active in political causes and the social causes.
SPK_0	00:04:12	And I don't think they were in as much involved with with social causes as they became after World War Two. When when the Colonel Shoemaker was dean of students, he got started these various projects in which students would undertake to sponsor uh uh philanthropic and humanitarian activities downtown. Uh it started with, he got, it started with the Apo houses, I recall exchanging their hell week to Help week and the word
SPK_1	00:04:49	fraternities and
SPK_0	00:04:49	sororities. And then they
SPK_1	00:04:51	were active on the

Phase 2 MGMs Implemented

- Speech-to-text:
  - Kaldi (OSS, HPC)
  - AWS Transcribe (CCL)
- Named entity recognition:
  - SpaCy (OSS)
  - AWS Comprehend (CCL)
- Video OCR:
  - Tesseract (OSS)
  - MS Azure Video Indexer (CCL)
- Segmentation:
  - PySceneDetect (OSS)
  - MS Azure Video Indexer (CCL)
  - INA Speech Segmenter (OSS, HPC)
  - Applause Detection (AMP)
- Human MGMs:
  - BBC Transcript Editor (OSS)
  - Audio Timeliner for NER editing (OSS)
- Other:
  - Dlib face_recognition.py (OSS)
  - FFmpeg (OSS)
  - Vocabulary tagging (AMP)
  - Gentle forced alignment (OSS)

OSS=Open source software
CCL=Commercial cloud
HPC=High Performance Computing
AMP=AMP-developed

AMP
AUDIOVISUAL METADATA PLATFORM

MGM Evaluation Criteria

- Accuracy
- Input formats
- Output formats
- Growth rate
- Processing time
- Computing resources
- Social impact / ethical considerations
- Cost
- Support
- Integration capabilities
- Training

AMP
AUDIOVISUAL METADATA PLATFORM

Phase 2 Lessons Learned

- Challenges with proprietary tools
  - Unpredictability, undocumented behavior
  - “Black box” aspects of process
  - Lack of clarity in terms of use
  - Opt-in vs. opt-out for use of data for product improvement
  - …but for certain tasks they work really well!

- Tools for language more robust and available than those for music
  - Speech-to-text, NER vs. Music genre, instrumentation detection
  - Many music use cases require training
  - Opportunity for training data sets from libraries, collaboration with music IR community

AMP
AUDIOVISUAL METADATA PLATFORM

Phase 2 Lessons Learned (continued)

- Use of existing tools/models vs. training of new models
  - Applause detection for music concert segmentation
  - Facial recognition - ethical considerations
- Technical implementation
  - Unpredictable wait times for batch-oriented HPC jobs
  - Integrating Human MGMs into Galaxy workflows
- Librarian/archivist engagement
  - Lots of excitement about potential
  - More difficult to think about practical implementation

AMP
AUDIOVISUAL METADATA PLATFORM

AMP Phase 3

Jira Software | Dashboards | Projects | More | Create | Search

AMP board

Backlog

Board

QUICK FILTERS: Only My Issues | Recently Updated | All Issues

AMP Sprint 81 9 issues ACTIVE 6 17 0

22/Mar/22 9:20 AM - 05/Apr/22 9:20 AM View linked pages

AMP-1503	Gather information for scoring tools	Evaluation Tool	3
AMP-1751	Figure out how to get current session_csrf_token for logout from Galaxy	Workflow Editor	2
AMP-1770	Proxy traffic between AMP UI and Galaxy during WF edit session (no URL filte...	Workflow Editor	3
AMP-1772	High level design for MGM evaluation tools	Evaluation Tool	3
AMP-1768	POC of using tomcat server rather than apache as the web server	Packaging and Depl...	4
~~AMP-1693~~	Workflow Editor UI - implement Edit button	Workflow Editor	2
AMP-1798	Structure for test config information	Evaluation Tool	2
AMP-1730	Update AMP staging/prod with all changes since last update	Technical Work	1
AMP-1765	Contact sheets for VOCR	MGM development	3

Backlog 95 issues

Create sprint

AMP-1689	Test workflows covering all MGMs in galaxy	MGM development	2
AMP-1364	Discuss data needs	Evaluation Tool	3
AMP-1775	Workflow Editor UI - implement Done button	Workflow Editor	2
AMP-1756	Sprint 80 - Update amp-prod with all changes since last update	Technical Work	1
AMP-1661	Gentle Force Alignment MGM failed in PROD	MGM development	3
AMP-1779	Figure out how build processes work	Packaging and Depl...	2
AMP-1777	Scene Detection and blank scenes	MGM development	2
AMP-1674	Update azure credentials in each Galaxy instance	MGM development	1
AMP-1697	How busy?	Reporting	2

AMP Phase 3

Funded by a new grant from the Mellon Foundation
July 2021 – December 2022

Focus on:

- System Robustness and Resilience
- Packaging, Deployment and Documentation
- UX Evaluation and Improvement
- MGM Evaluation Module
- Collection Testing

AMP
AUDIOVISUAL METADATA PLATFORM

UX Evaluation and Improvement

Collection Partners

Indiana University

- Archives of African American Music and Culture
- Archives of Traditional Music
- Black Film Center & Archive
- IU Libraries Moving Image Archive
- University Archives

New York Public Library

- Research Division (Rights, Archives, Metadata)
- Schomburg Center for Research in Black Culture (Moving Image and Recorded Sound Division)

AMP
AUDIOVISUAL METADATA PLATFORM

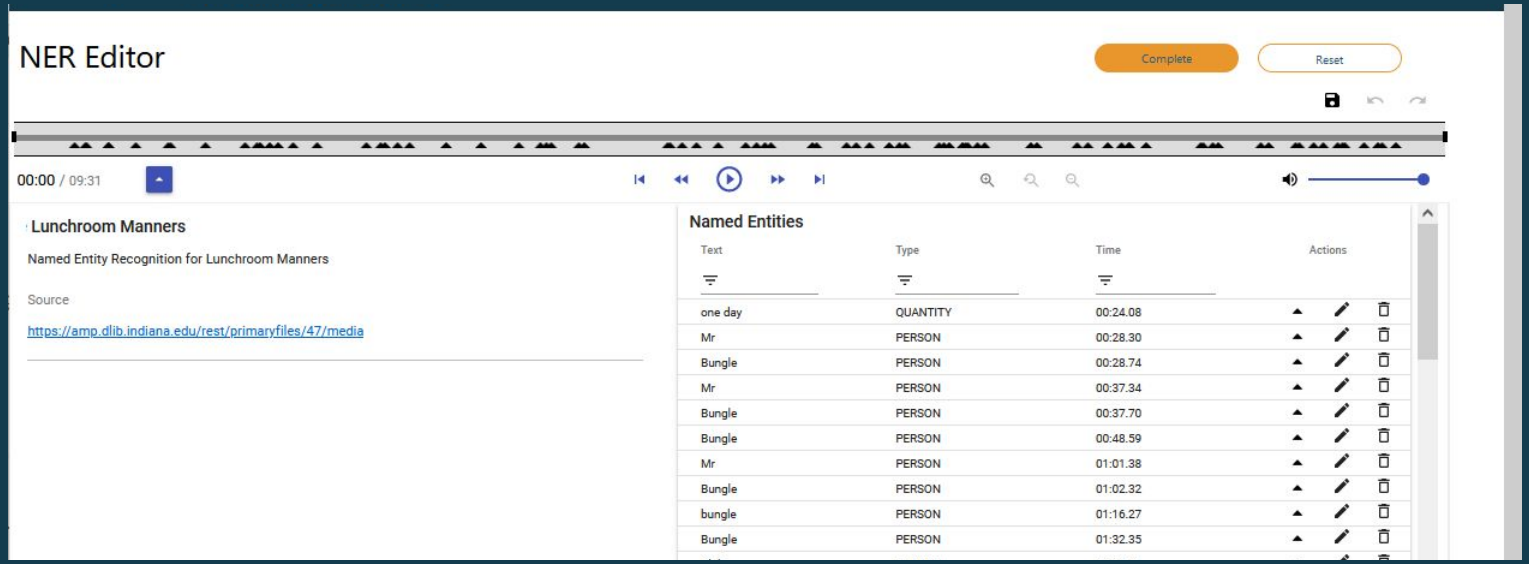
Phase 3 Feature Development

- Workflow creation
- Batch upload
- Improved navigation of collections
- “Intermediary” files as MGM input

AMP
AUDIOVISUAL METADATA PLATFORM

Phase 3 Feature Development

- HMGM tool improvements
  - BBC Transcript Editor debugging
  - Addition of video player to Timeliner (editing NER, Audio segmentation, others)
- Independence of HMGMs from AMP workflows



Engaging with Collections Managers

- System demos
- Hands-on training
- All-team meetings
- Targeted discussions
- Focus groups

AMP
AUDIOVISUAL METADATA PLATFORM

Use cases for AMP

- **Archivists** want to efficiently accession collections
- **Catalogers**, **archivists** and **metadata librarians** want to describe media or enrich metadata
- **Rights clearance staff** want to provide appropriate levels of access
- **Researchers** want minimally processed media to be further described on-demand

AMP
AUDIOVISUAL METADATA PLATFORM

System Requirements for File Upload

- **Archivists** want to efficiently accession collections
- **Catalogers**, **archivists** and **metadata librarians** want to describe media or enrich metadata
- **Rights clearance staff** want to provide appropriate levels of access
- **Researchers** want minimally processed media to be further described on-demand

Flexibilities in data model and API to accommodate:

- Single item, batch, and API upload
- 0 or more external identifiers

AMP
AUDIOVISUAL METADATA PLATFORM

Data Output Formats

A "contact sheet" that can be generated from shot detection, video OCR, face recognition, and other MGM outputs

AMP
AUDIOVISUAL METADATA PLATFORM

User Feedback – Surfacing New Issues

PySceneDetect (shot detection) contact sheet

Too many shot transitions detected in areas with no content

AMP
AUDIOVISUAL METADATA PLATFORM

User Feedback – Surfacing New Issues

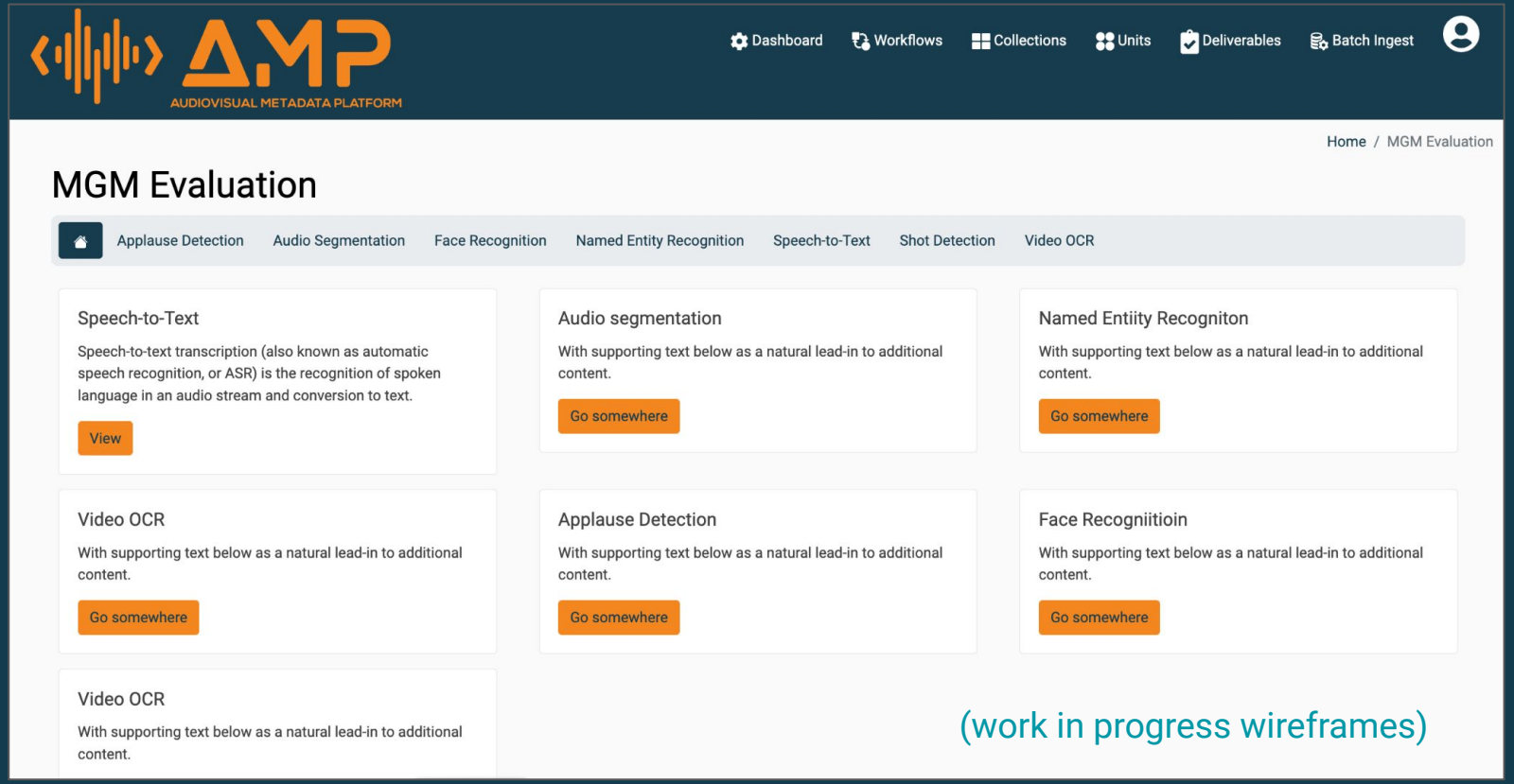
```
00:28:36 --> 00:28:39
<v spk_0> T. I. U or call monday through friday

00:28:39 --> 00:28:40
<v spk_4> during normal business hours.

00:29:05 --> 00:29:05
<v spk_4> Okay.

00:29:14 --> 00:29:14
<v spk_4> Okay.

00:29:30 --> 00:29:34
<v spk_4> Okay. Okay.

00:29:51 --> 00:29:54
<v spk_4> Okay. Okay.

00:30:01 --> 00:31:06
<v spk_4> Okay. Okay thank you. Okay. Okay.

00:31:27 --> 00:31:27
<v spk_4> Okay.

00:31:39 --> 00:31:49
<v spk_4> Okay. Okay. Okay. Okay.

00:32:01 --> 00:32:08
<v spk_4> Okay. Okay.

00:32:37 --> 00:32:44
<v spk_4> Okay. Okay.

00:32:58 --> 00:33:04
<v spk_4> Okay. Okay.
```

Amazon Transcribe (speech-to-text transcription) WebVTT file

Words transcribed from segments of silence

AMP
AUDIOVISUAL METADATA PLATFORM

MGM Evaluation Module



(work in progress wireframes)

Evaluation – Ground Truth Testing

AMP
AUDIOVISUAL METADATA PLATFORM

Dashboard | Workflows | Collections | Units | Deliverables | Batch Ingest

Home / MGM Evaluation

MGM Evaluation

Applause Detection | Audio Segmentation | Face Recognition | Named Entity Recognition | Speech-to-Text | Shot Detection | Video OCR

Speech-to-Text

Speech-to-text transcription (also known as automatic speech recognition, or ASR) is the recognition of spoken language in an audio stream and conversion to text.

→Confluence documentation Sample Use Cases

+ New Test | Evaluation Results

Select a New Test

- STT Word Error Rate Test
- STT Accuracy Test

STT Word Error Rate Test

Word error rate measures speech-to-text accuracy by comparing the generated transcript against a ground truth transcription and calculating the number of errors (substitutions, additions, and deletions) as the cost of restoring the output word sequence to the original input sequence. Scores are measured as percentages, with lower scores representing high accuracy (low word error rate). Scores may exceed 100%, especially if the STT engine produced many insertions. Related to WER, character error rate (CER) is similar to WER, but based on characters, the word information loss (WIL) measures the proportion of word information lost in a transcription, and word information processed (WIP) measures the inverse of WIL.

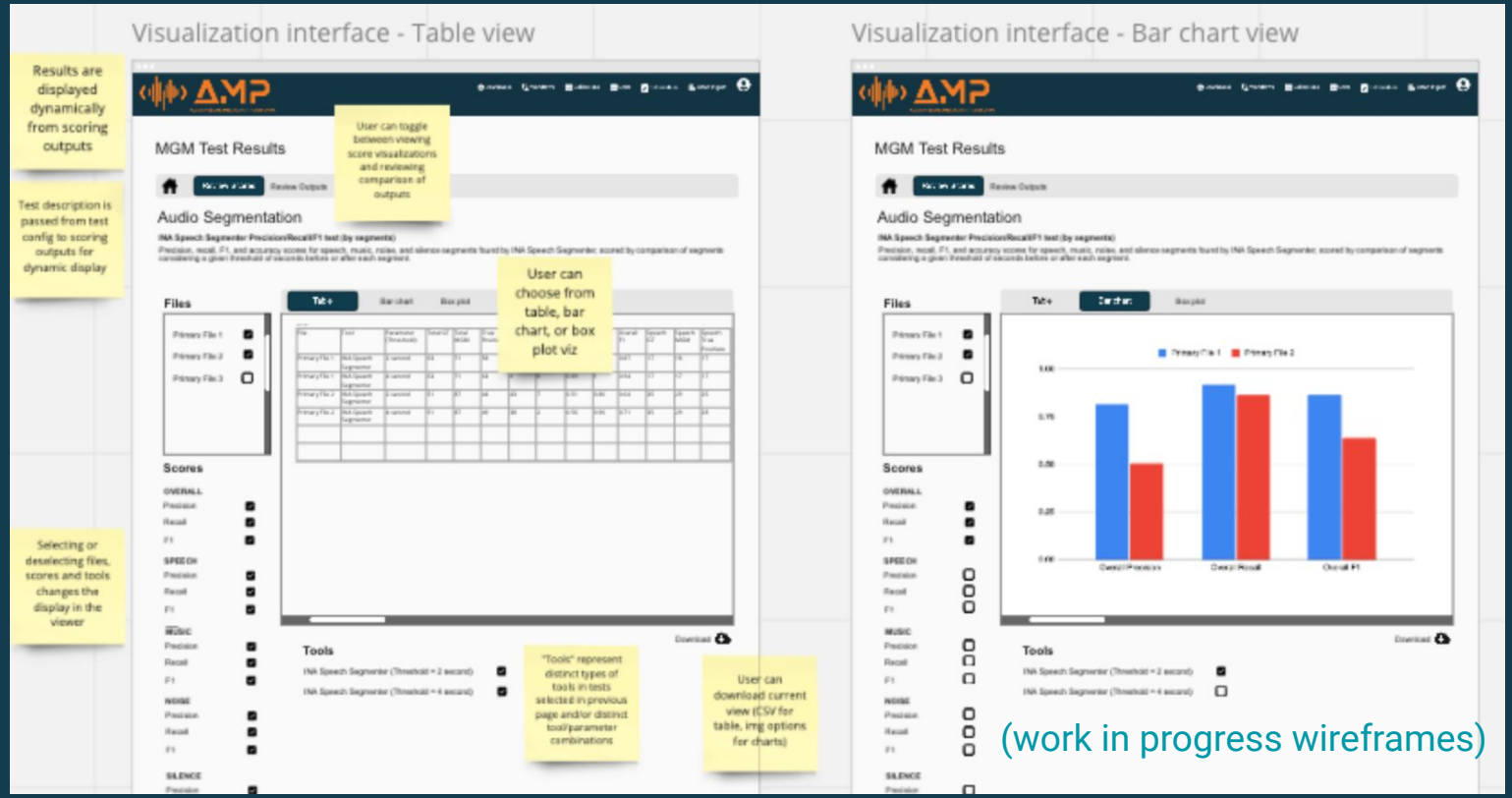
Short Description of the "Ground Truth Template" and requirements, link to more ground truth information.

Download the Ground Truth Template

Parameters

(work in progress wireframes)

Evaluation – Visualization Interface



Qualitative Evaluation

Ground Truth	MGM Output	Error Type
its	its	
my	like	substitution
very	very	
distinct	distinct	
pleasure	pleasure	
to	to	
welcome	welcome	
you	you	
all	all	
to	over	substitution
the	the	
second	second	
patten	**	deletion
lecture	lecture	
of	off	substitution
this	this	

Ground truth comparison view of Amazon Transcribe speech-to-text output

AMP
AUDIOVISUAL METADATA PLATFORM

Packaging

Multi-tier deployment approach

- *Tier 1*: Direct install to OS with installation scripts and detailed documentation
- *Tier 2*: Install component-level Docker containers
- *Tier 3*: Scripts and configurations for orchestration of container deployment

https://github.com/AudiovisualMetadataPlatform (work in progress)

AMP
AUDIOVISUAL METADATA PLATFORM

AMP Future

AMP Future

- Release and adoption of AMP for local and cloud installation
- Integrate additional MGMs and target systems
- Opportunities for training new models
- Support output as IIIF / Web Annotations
- More work around music use cases
- Continue to build community for AI/ML in libraries and cultural heritage (e.g. AI4LAM)

AMP
AUDIOVISUAL METADATA PLATFORM

Thank you!

go.iu.edu/amppd

Jon Dunn: jwd@iu.edu

Shawn Averkamp: shawn@weareavp.com

AMP
AUDIOVISUAL METADATA PLATFORM