|  |
| --- |
| MEETING AGENDA 10/8 |

|  |  |
| --- | --- |
| Location: | CDL |
| Date: | 10/8/2024, Tuesday |
| Time: | 12:30PM - 1:30PM |

**Overall Project Goals**

1. **Develop a Multilingual ASR System**:

• Build a functional automatic speech recognition (ASR) system capable of processing and recognizing speech in both English and Mandarin.

• Ensure the system can accurately recognize and differentiate between adult (caregiver) and infant speech (speaker recognition).

2. **Preprocess and Annotate Data**:

• Collect and preprocess high-quality audio recordings of infant-caregiver interactions.

• Annotate the recordings with accurate labels for different speech segments, including caregiver instructions, infant vocalizations, and interaction events.

3. **Train and Optimize the ASR Model**:

• Train the ASR model using annotated audio data to achieve high recognition accuracy.

• Optimize the model’s performance by adjusting parameters and incorporating feedback from initial test runs.

4. **Integrate Multilingual Capabilities**:

• Develop and integrate ASR capabilities for Mandarin, ensuring that the model functions effectively across both languages and maintains a high level of accuracy.

5. **Test and Validate the ASR Model**:

• Conduct thorough testing of the ASR model using diverse datasets to evaluate its performance in recognizing speech from both languages and different speaker types (infant vs. caregiver).

• Fine-tune the model based on test results to improve precision and reliability.

6. **Prepare for Deployment**:

• Document the model development process, results, and any challenges encountered during the project.

• Create a scalable ASR system that can be deployed in future studies and applied to larger datasets or different infant-caregiver settings.

* Test

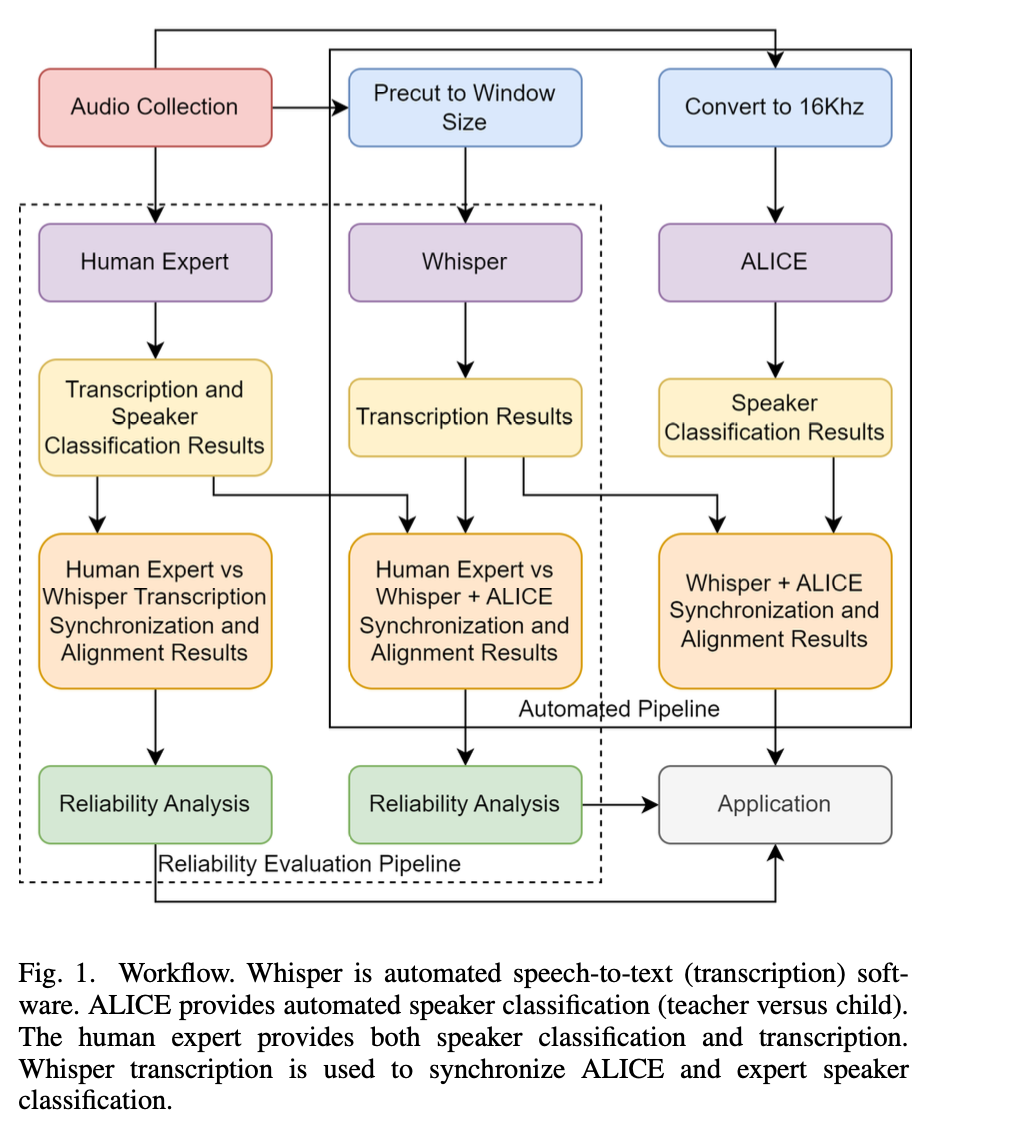
|  |  |  |
| --- | --- | --- |
| Week | Goals | Description |
| 2 | **Team Orientation & Goal Setting** | Introduce team members, define roles and responsibilities, review the pipeline from Sun et al., and set the overall project goals. Establish a clear timeline and division of tasks. |
| 3 | **Data Collection & Preparation** | Begin collecting data for ASR model development. Ensure diverse data inputs (English and Mandarin) from infant-caregiver interactions. Organize and preprocess raw audio files, converting formats and normalizing audio as needed. |
| 4 | **Annotation and Labeling** | Annotate collected data, focusing on infant-caregiver interaction segments. Use existing tools or develop a basic annotation framework for consistency. |
| 5 | **Initial ASR Model Setup** | Set up the ASR framework using Whisper or an alternative ASR system. Test with small data samples to verify model setup and functionality. |
| 6 | **Model Training Phase 1** | Train the ASR model on a subset of the annotated dataset. Run initial training experiments and document outcomes. Identify and troubleshoot issues. |
| 7 | **Model Evaluation & Improvement** | Evaluate the ASR model’s performance using accuracy metrics and other benchmarks. Begin model refinement, adjusting parameters, and optimizing for clarity and precision. |
| 8 | **Integration of Mandarin Data** | Incorporate Mandarin audio data into the model training process. Ensure the model is compatible and efficient with multilingual inputs. |
| 9 | **Testing & Fine-tuning** | Conduct extensive testing using both English and Mandarin datasets. Fine-tune the ASR model based on feedback and accuracy rates. Prepare for scalability tests. |
| 10 | **Final Evaluation & Reporting** | Evaluate the final model’s performance, comparing it to initial benchmarks. Compile a report documenting the methodology, results, and challenges faced. Plan next steps for deployment or further development. |

Week 2:

1. Google drive folder: <https://drive.google.com/drive/u/1/folders/1JlJQ_T1MDaIKLW52I8FG1eXFJlzXAfmq>
2. Set up python
3. Set up remote desktop: <https://docs.google.com/document/d/1K8dnO_2ADpaDUskeHzm-6C7b7YpWk6pfIoqWuUG1d_o/edit>
4. Connect to Anzhe Sun to see if he could provide open source code (done, will be made available by the end of this month)
5. Google collab
   1. Tai
6. Github
   1. Create github account and share with Yueyan (todo)
   2. Start a repository (todo)
7. Data collection, Preprocessing, Training, Evaluation
8. Share audio file (28\_05\_test)
   1. Access through 02 (todo)
9. Download pytorch and whisper:
   1. pip install git+https://github.com/openai/whisper.git
   2. pip install torch
10. <https://medium.com/gimz/how-to-install-whisper-on-mac-openais-speech-to-text-recognition-system-1f6709db6010>

TODO LIST:

1. Read on pre processing audio files & preprocess the audio files
   1. Use 28\_05\_test audio file, make sure knows how to
      1. Convert audio files to the required frequency (e.g., 16kHz for ALICE) to align with the input requirements of your ASR models.
      2. Segment long recordings into shorter epochs (e.g., 2-minute segments) before transcription to optimize model performance and reduce errors.
      3. Whisper runs smoothly on those files.
2. Convert home videos to audios (English) 5 files (4, 6, 8, 10, 12 months) + download open-source audio files in Mandarin (CHILDES, download) : <https://childes.talkbank.org/> -check if ALICE could work with Mandarin
   1. Data collection pipeline
   2. Documentation - Tai
3. Set up github repository, download pytorch, whisper (alice) -individual

****