Automating Farmer Interview Transcriptions with Speech and Generative AI: A Low-Resource ASR Approach for Cambodian Agriculture

Mr. Soklay Heng
AI lecturer&researcher, CADT

Mr. Soklong Him
AI lecturer&researcher, CADT

Presented by: Mr. Soklay Heng
Institute of Digital Research and Innovation (IDRI),
Cambodia Academy of Digital Technology (CADT)

Cambodia Academy of Digital Technology

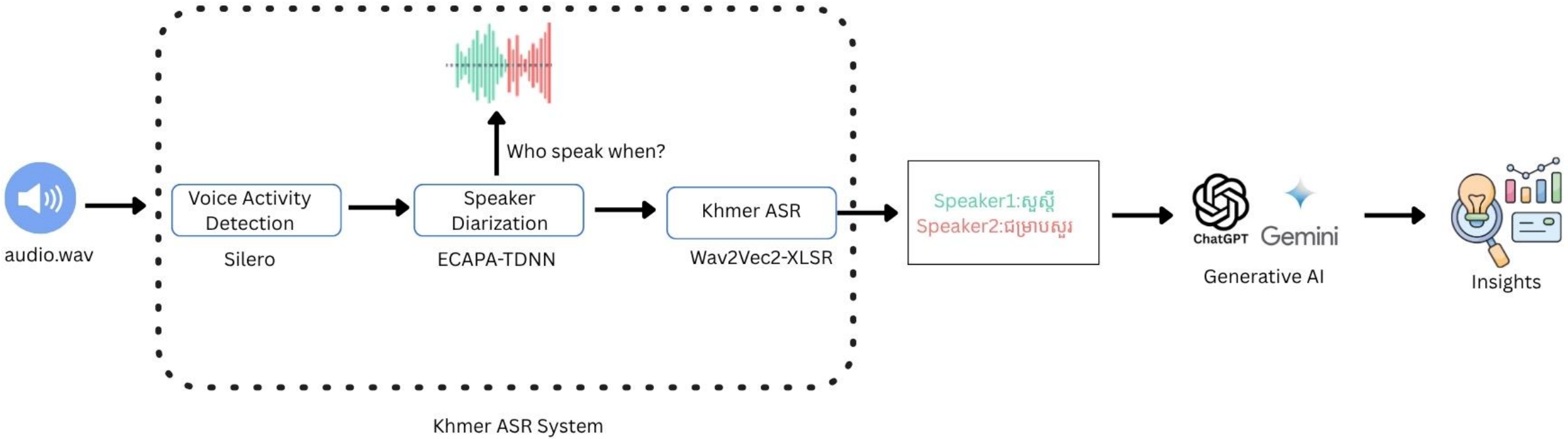
Background

- Agriculture is a key pillar of Cambodia's economy, contributing about 22% of national GDP.
- Around 36.6% of the workforce was employed in agriculture in 2022.
- The government is investing in irrigation, productivity, and agri-tech.
- Government and development partners collect farmer interviews to understand local needs and to design better interventions.
- These interviews are usually recorded in real field conditions, so the audio often contains background noise and overlapped speakers.
- Researchers need to transcribe interviews manually, which is slow, costly, and it delays data analysis and reporting.

Targets

- Develop an automatic system that can transcribe Khmer farmer interviews collected in real field conditions and identify speakers in the recording.
- Utilize generative AI to produce quick insights, analysis, and summaries from long interviews.
- Build a management system tool that helps social science researchers organize, store, and retrieve their interview data.
- Provide a friendly web-based interface so that users can upload audio and directly obtain a structured output.
- Reduce the time from data collection to analysis, so that field data can be used faster to improve agricultural products and services.

Proposed Methods



Architecture Diagram

Project Pipeline

Data Collection

- Collect raw interview and focus group audio from agriculture researchers in the field.

Data Cleaning and pre-procesing

- Perform format and sampling rate conversion, and noise checks.
- Segment long recordings to small audio chunks.
- Annotate audio chunks to use as training and evaluation data.

Model Development

- Select the most suitable AI model for low-resource ASR and diarization.
- Fine-tune pre-trained multilingual models on the collected field data.
- Evaluate the model on real test dataset.

Development of web applicatoin

- Develop a user-friendly web application for social science researchers.
- Integrate the AI models (ASR, diarization, generative AI) into the web backend.
- Implement features for uploading audio and downloading structured transcripts.

Testing & Deployment

- Conduct end-to-end testing with real field users to validate usability and accuracy.
- Refine the models and interface based on feedback.
- Deploy the system for operational use in research projects.

Outcomes

- Develop an ASR pipeline tailored to Khmer farmer interview speech, applying in a low-resource language.
- Deliver a user-friendly web application that researchers can use to upload audio and obtain structured transcripts and insights.
- Create opportunities for joint work between AI researchers and social science researchers in Cambodia and ASEAN.

Impacts

- Enhance food safety and protection.
- Improve work efficiency between field data collection and analysis, allowing researchers to use farmer interview audios more quickly in their studies.
- Provide a practical tool that social science researchers in Cambodia can share and reuse across different projects.
- Serve as a model for extending AI solutions to other local languages and sectors in the region.

Conclusion

- This proposed research innovation reduces the manual workload of social science researchers and improve the efficiency of field data.
- The web application makes advanced AI models accessible to non-technical users.
- The proposed solution can be extended to other domains and ASEAN local languages.
- With more data and partners, this work can become a shared ASEAN resource for low-resource speech technologies.

Thank you!

Institute of Digital Research and Innovation (IDRI)

Cambodia Academy of Digital Technology (CADT)

Mr. Soklay Heng

Email: soklay.heng@cadt.edu.kh