

Team Communication Analytics Using Automated Speech Recognition

Randall Spain¹, Wookhee Min¹, Jason Saville¹, Bradford Mott¹,
Keith Brawner², Joan Johnston², Gregory Goodwin², & James Lester¹

¹North Carolina State University

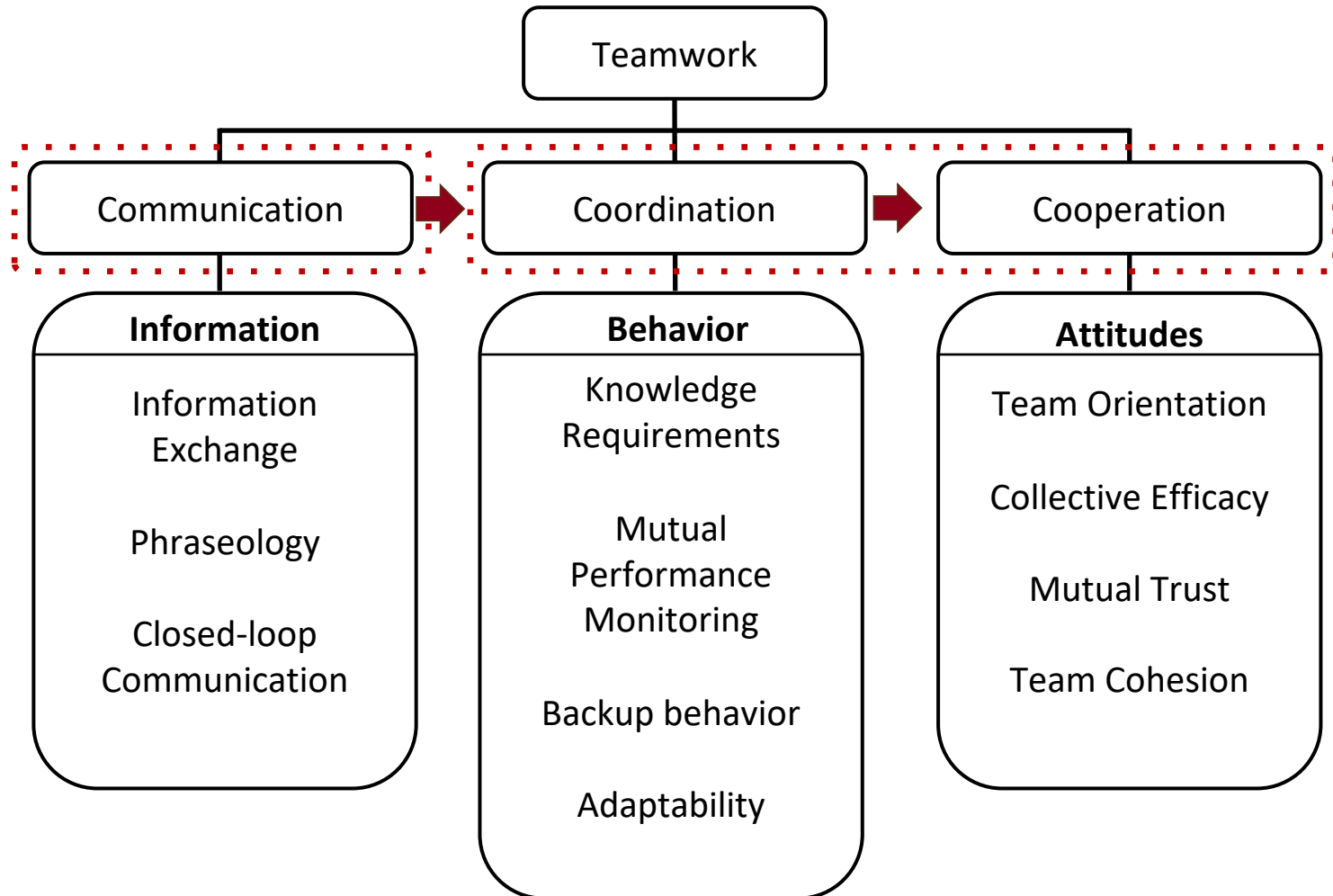
²US Army Combat Capabilities Development Command
Soldier Center, STTC

Military Teams



- Foundational building blocks of the military
- Units train and conduct missions as teams
- Effective teams coordinate actions, share information, and assist one another
- Training Soldiers to operate as effective team members is critical for readiness

Team Communication and Teamwork



(Wilson, Salas, & Priest, 2007)

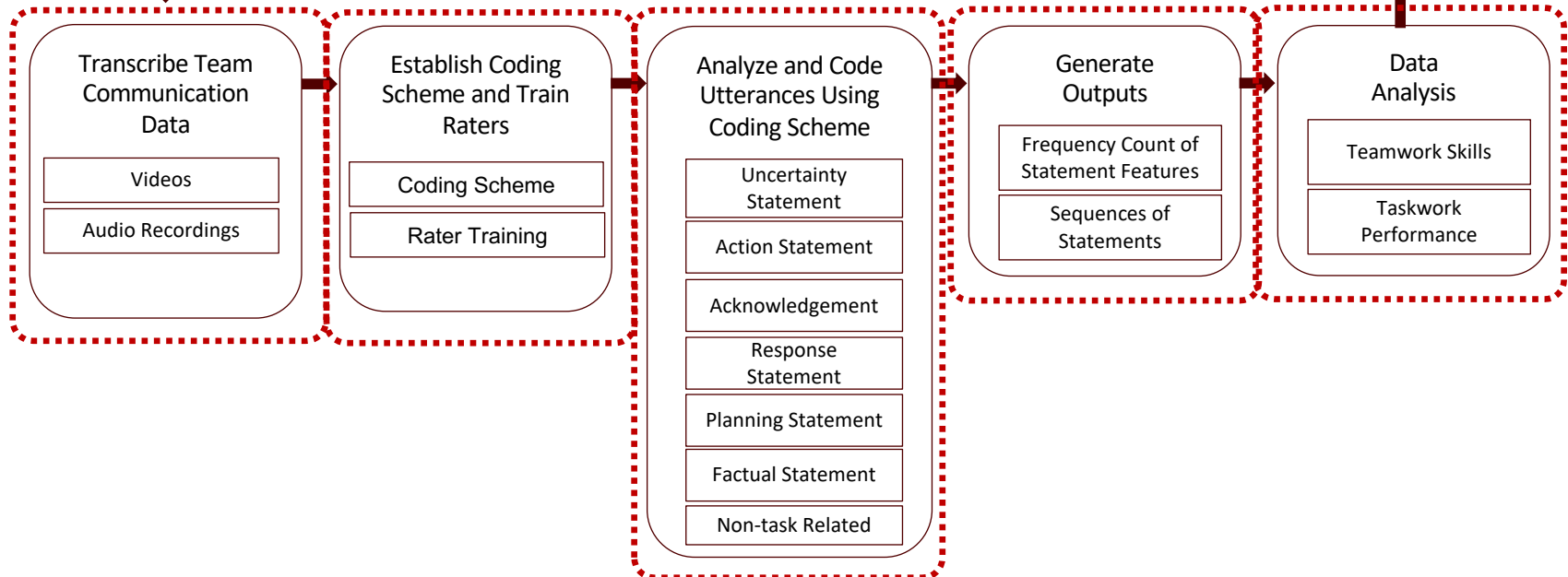
Team Communication and Team Outcomes

- Communication **quality** is a better predictor of team performance than communication **quantity** (Marlow et al., 2018)
- Communication provides a unique lens for studying the dynamics of **team cognition** and shared team mental models (Grand et al., 2016)
- Communication plays a central role in **team learning** (Sottolare et al., 2018)
- Team communication can provide insights into **team cohesion** and **collective efficacy** (Norman, 2019)

Team Communication Analysis Pipeline

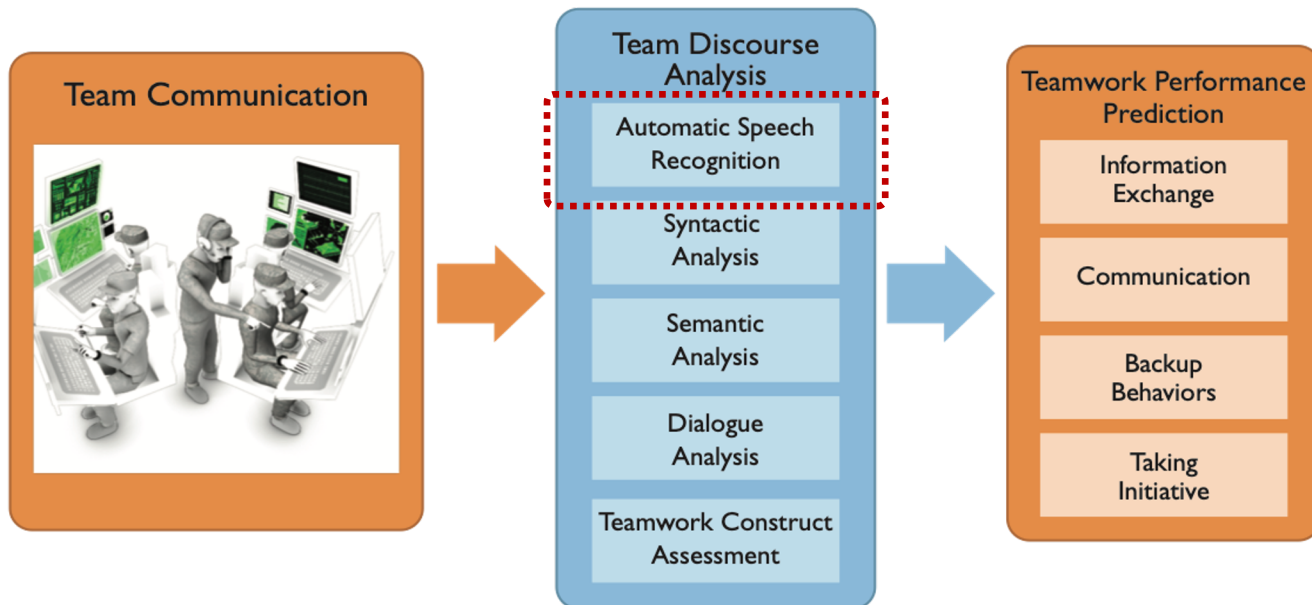


← Teamwork Training

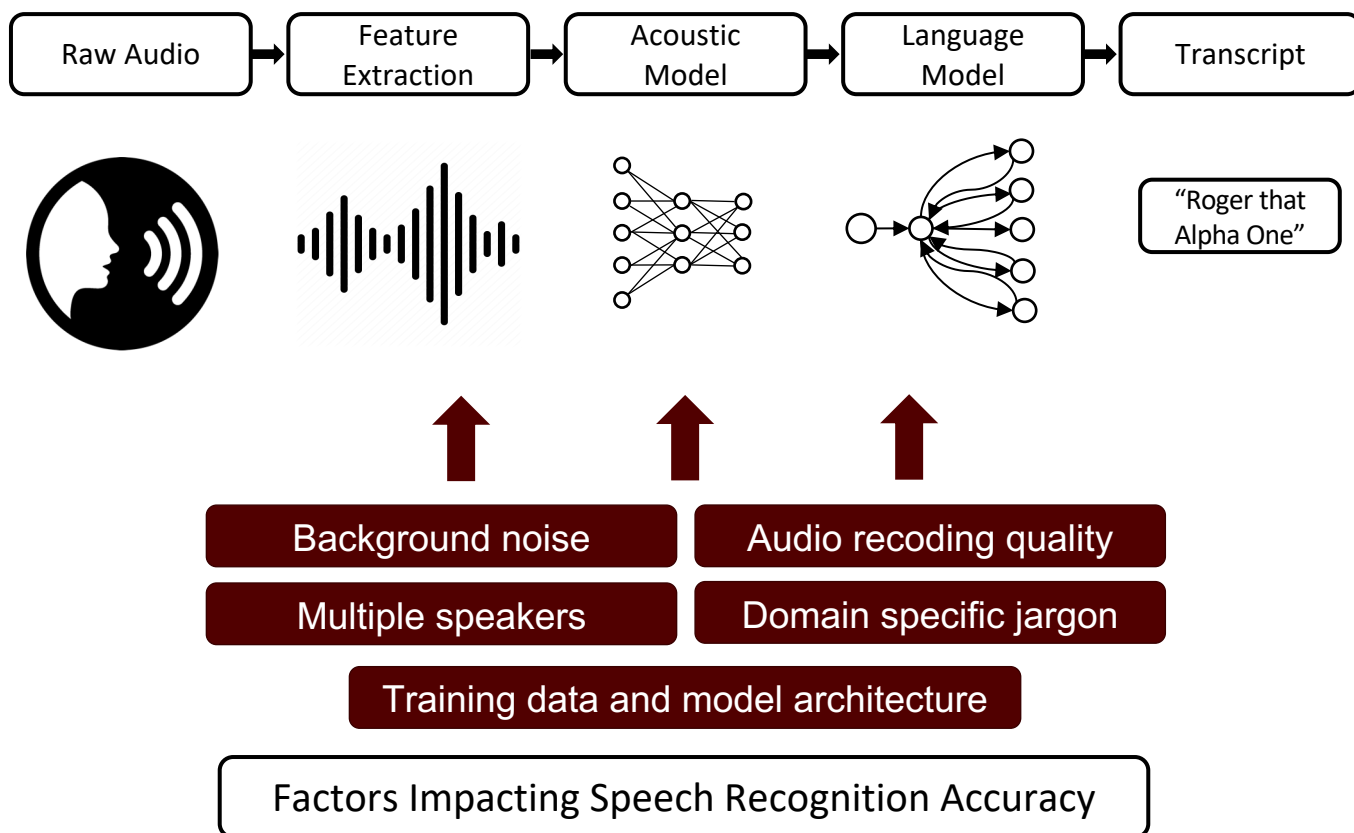


Team Communication Analysis Toolkit

Develop a **deep learning-based team communication analysis** toolkit that can perform real-time end-to-end natural language analysis on team members' spoken dialogue and generate team dialogue analytics that drive adaptive scaffolding.



Automatic Speech Recognition



ASR Performance

- **Speech in noisy classrooms**
 - Blanchard et al. (2015) analyzed recordings of questions posed by teachers in noisy classrooms
 - Results showed Google Speech (WER 44%) and MS Azure (WER 48%) produced the most accurate transcripts
- **Speech in a training setting**
 - Kim et al. (2019) found YouTube (WER 28%) and Google Cloud (WER 35%) performed the best when analyzing videos of simulated patient interviews
- **Speech in different domains with virtual characters**
 - Georgila et al. (2020) compared ASR performance of Amazon, Apple, Google, IBM, Microsoft, and Kaldi using data collected from human interactions with virtual agents
 - Google Speech performed the best across datasets
 - WER for VR training recordings (35 - 80%) was considerably higher compared to interactions with virtual characters (8 - 18%)

Goals

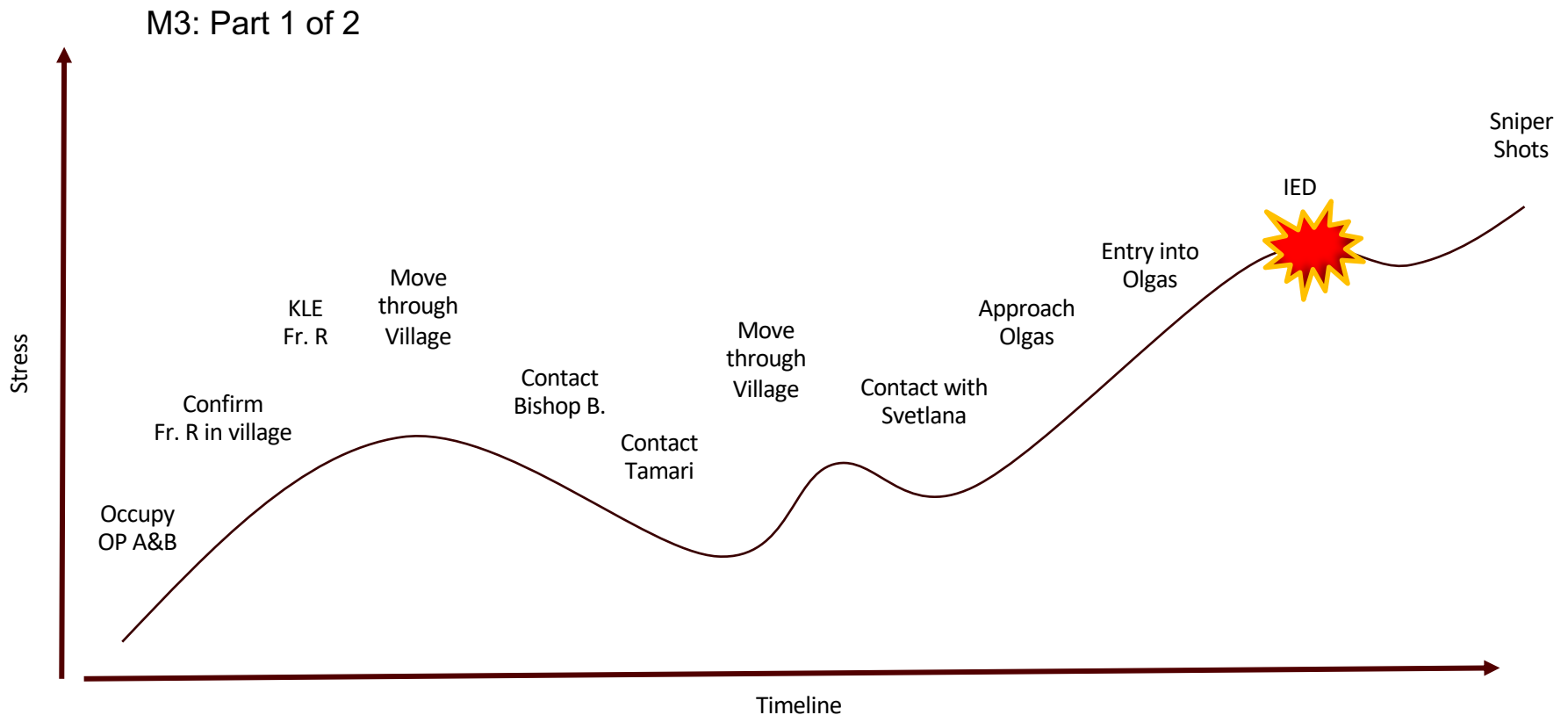
Extend previous research and compare the speech recognition performance of contemporary ASR services using two team communication datasets to evaluate their accuracy for supporting team training analytics.

Team Speech Recordings



- Squad Overmatch Study
- Completed virtual and live training events
 - VBS3 scenarios
 - Live training missions
 - AAR session
- Team communication recordings
 - Multiple speakers
 - Environmental noise
 - Degraded recording quality

Sequences of Events



Squad Mission: Conduct a zone reconnaissance in order to conduct KLE; exploit intelligence; confirm location of a suspected arms cache; and, exploit the site, if able

Data Preparation Process



- **Live Training**
 - 45 mins of audio
 - Noisy environment
- **AAR**
 - 40 mins of video
 - Sterile environment

Keywords

- Google Speech
- Microsoft Azure
- Kaldi

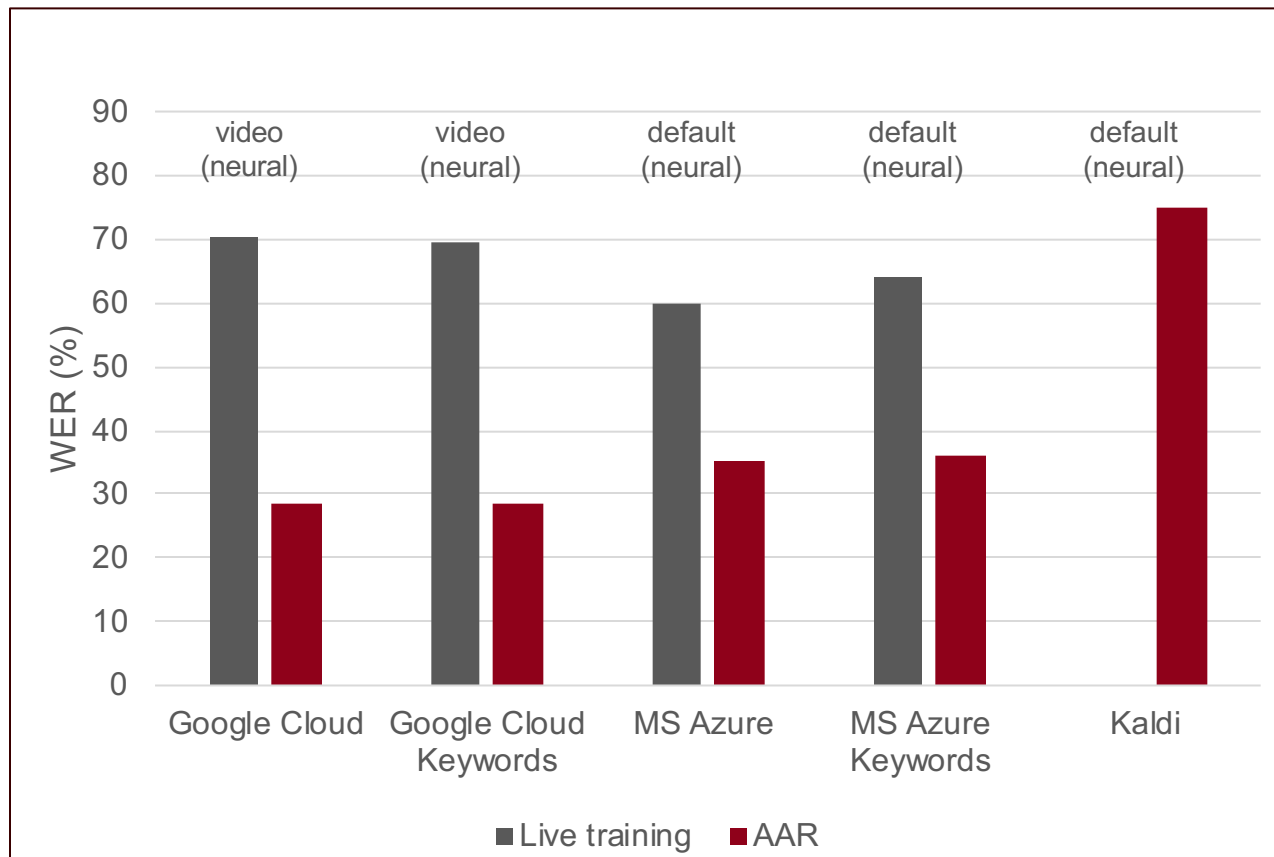
Speech Segments

- **Live Training**
 - SQD Leader
 - 200 speech turns
- **AAR**
 - 321 speech turns

Measures

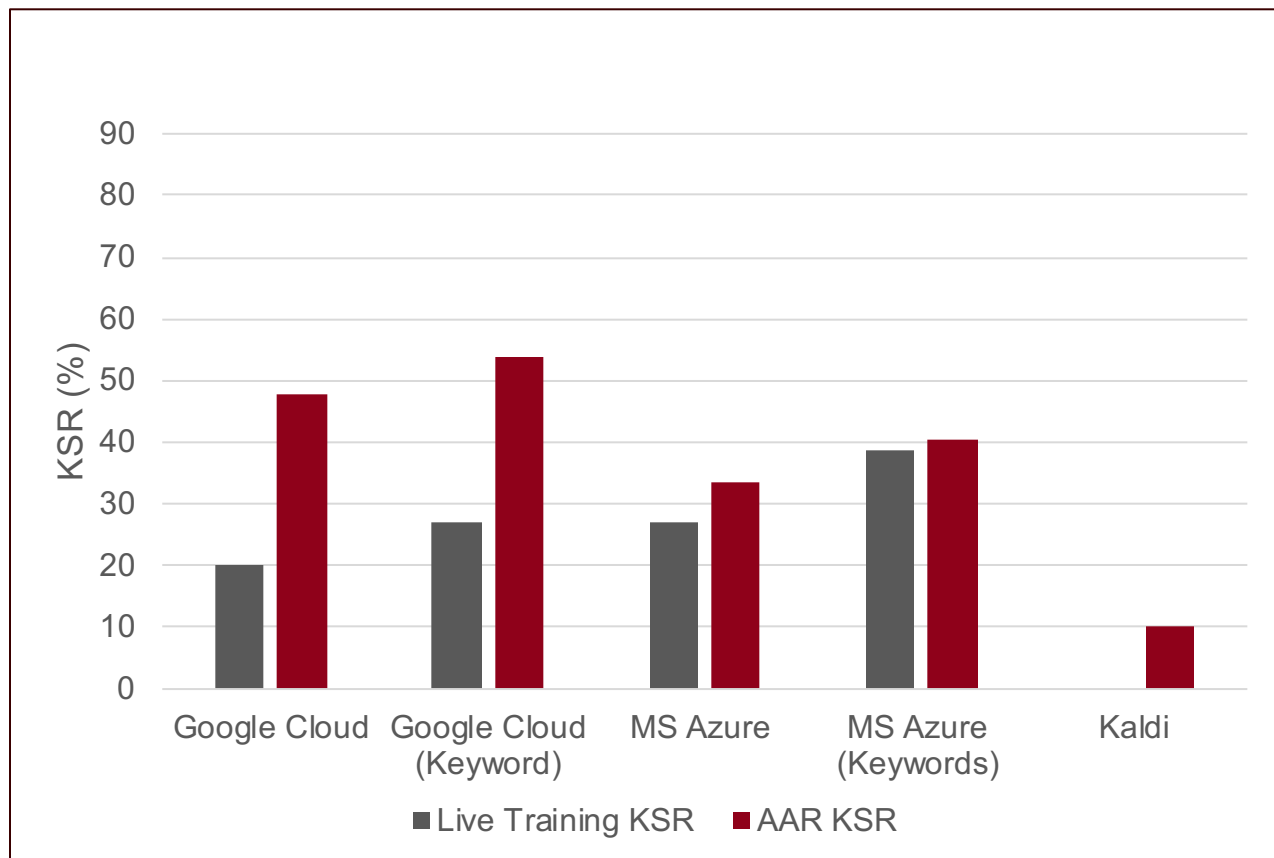
- Word error rate (WER)
- Keyword spotting rate (KSR)
- Transcript omission rate (TOR)

Word Error Rate (%) Results



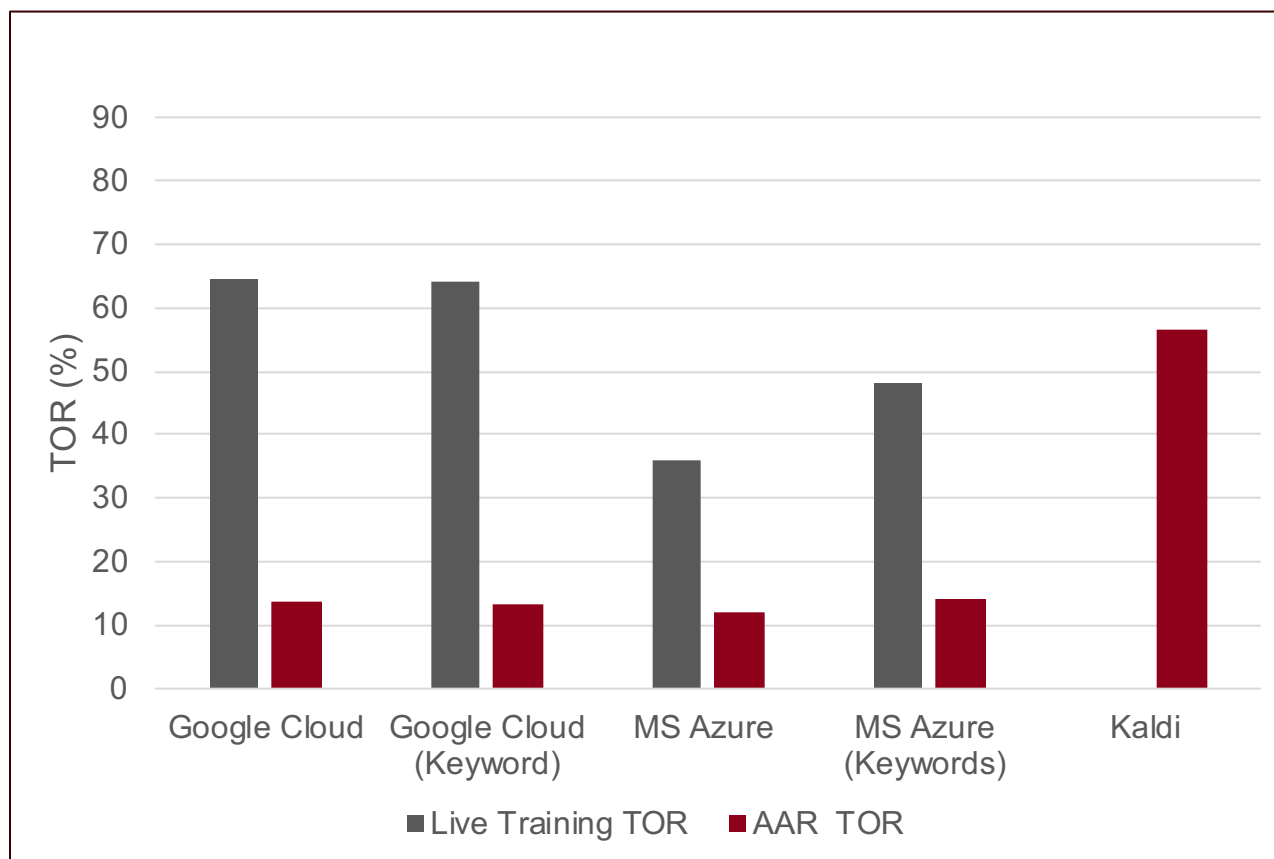
WER: Word Error Rate (Lower is better)

Keyword Spotting Rates (%) Results



KSR: Keyword Spotting Rate (Higher is better)

Transcript Omission Rate (%) Results



TOR: Transcript Omission Rate (Lower is better)

Transcript Examples

Transcript	Google Cloud	Microsoft Azure	Kaldi
“..yeah let's start pushing it up we need to go over towards alpha one this time by where the dude with the mic is alright..”	“..now let's start pushing it up need to go over towards alpha 1 this time I wear the dude from my kids right huh..”	“..Yeah, we'll start pushing up. Need to go over towards A1 this time by where the dude Mike is right..”	N/A
“..seems as though the vendors are arguing on route black but dispersing now hard copy over..”	“..arguing on black..”	“..Seems other vendors arguing on route black, but this person now copy over..”	N/A
“..squad leader talk to me what was the mission and your plan scheme and maneuver..”	“..squalor talk to me what was the mission and your plan scheme maneuver..”	“..squalor talk? Maine was the mission and your plansky maneuver..”	“..as well it taught me was the mission and plants in your gut..”

Conclusion

- Despite many advancements in speech recognition capabilities the transcripts produced by the ASR systems were error prone, particularly for the live training session
- Adding domain specific keywords to the ASR toolkits may have improved keyword spotting performance but it did not increase transcription accuracy
- ASR performance was better for AAR recordings particularly sections where SMEs were providing feedback to squads
- Additional research is needed to identify methods to improve ASR accuracy

Future Directions

- Investigate how techniques for providing contextual information can be used to improve transcription accuracy
- Investigate background noise reduction and voice activity detection techniques
- Bridge the gap between bottom-up and top-down driven NLP team dialogue analysis
- Integrate ASR capabilities and NLP pipeline with GIFT to support team dialogue dictionary and performance assessment

Acknowledgments



Support provided by the U.S. Army Research Laboratory under cooperative agreement W912CG-19-2-0001.