A Review of Methods to Increase AI Transcription Accuracy in Cybersecurity Research

Christine Lo and David Schuster
San José State University

Introduction

- AI-generated transcripts with automated speech recognition (ASR) tools are widely used in qualitative research due to time and cost efficiency
- Fields like healthcare and law enforcement have explored ASR tools extensively
- However, cybersecurity remains an underexplored domain for ASR tools despite its growing importance in qualitative studies
- Problem: Generic ASR systems struggle with domain-specific technical language, reducing accuracy and interpretability in cybersecurity qualitative research

Current Study & Observation

- Evaluated transcripts from interviews with cybersecurity experts
- Poor recognition of technical terminology:
  - Cybersecurity jargons
  - Different speech patterns
- Common issues observed anecdotally:
  - Omissions of key terms
  - Fabricated or distorted words
  - Incorrect punctuation and grammar, distorting the overall sentence meaning

---

METHODS

We propose recommendations to improve qualitative research in cybersecurity

vectrlab.net

Recommendations

1. Domain-Specific Transcription Models
   - Train or fine-tune ASR models by providing them with a set of technical dictionaries or contextual information
   - E.g., feeding the system a glossary of cybersecurity terms (Parente, Kock & Sonsini, 2004; Sun et al., 2023)
1. Pre-processings Denoising
   - Suppress background noise before transcription to enhance the clarity of speech (Kinoshita et al., 2020)
1. Multi-model Inputs
   - Audio-Visual Speech Recognition (AVSR)
   - Integrates video processing (focusing on the speaker's facial and lip movements) alongside audio transcription for improved human verification (Ryumin, Ivanko & Ryumina, 2023)

Key Findings

- Generic ASR tools are not yet sophisticated enough; our observations in the lab underscore the knowledge gaps in current ASR tools for cybersecurity
- Errors can lead to data misinterpretation, especially when precise language is essential
- Future research should leverage existing advancements and adopt successful practices from other technical domains to enhance cybersecurity qualitative research

References

1. Parente, R., Kock, N., & Sonsini, J. (2004). An analysis of the implementation and impact of speech-recognition technology in the healthcare sector. Perspectives in health information management, 1, 5.
1. Sun, C., Ahmed, Z., Ma, Y., Liu, Z., Kabela, L., Pang, Y., Kalinli, O. (2023). Contextual Biasing of Named-Entities with Large Language Models. Arxiv. https://doi.org/10.48550/arXiv.2309.00723
1. Kinoshita, K., Ochiai, T., Delcroix, M., & Nakatani, T. (2020). Improving noise robust automatic speech recognition with single-channel time-domain enhancement network. NTT Communication Science Laboratories. https://doi.org/10.48550/arXiv.2003.03998
1. Ryumin, D., Ivanko, D., & Ryumina, E. (2023). Audio-Visual Speech and Gesture Recognition by Sensors of Mobile Devices. Sensors (Basel, Switzerland), 23(4), 2284. https://doi.org/10.3390/s23042284

Acknowledgments

vectr lab @ sjsu

Poster layout adapted from Morrison (2019).