NIH COLLABORATION TO IMPROVE AI-READINESS (NOT-OD-21-094) CLOSING OUT MEETING

Detecting Speech Articulation Patterns Following Laryngeal Cancer Treatment Using AI and Machine Learning

(Parent R01: Wearable Silent Speech Processor to facilitate oral communication)

PI: JUN WANG, PHD
Associate Professor
Department of Speech, Language, Hearing Sciences
Department of Neurology
University of Texas at Austin

TEXAS
The University of Texas at Austin

Georgia Tech

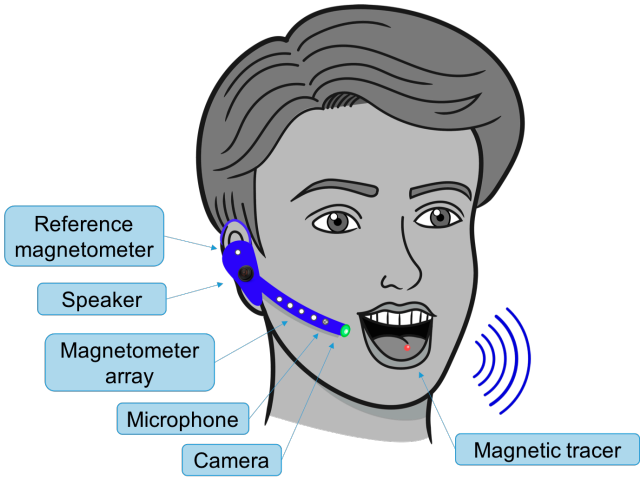
UT Southwestern
Medical Center

THE UNIVERSITY OF ARIZONA

OCT 31, 2022

Project Goals

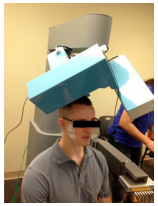
- Wearable silent speech processor to facilitate the oral communication for individuals after laryngectomy (surgical removal of larynx due to the treatment of cancer)
  - Wearable hardware for tongue motion tracking
  - Articulation-to-speech mapping using **AI**, **machine learning**, **automatic speech recognition (ASR)**, and **speech synthesis** techniques
  - Better understand the articulation patterns of alaryngeal speech



Conceptual design

Highlights

- Hardware – wearable device
- Algorithm – up to 90% accuracy
- Articulatory patterns of alaryngeal speech



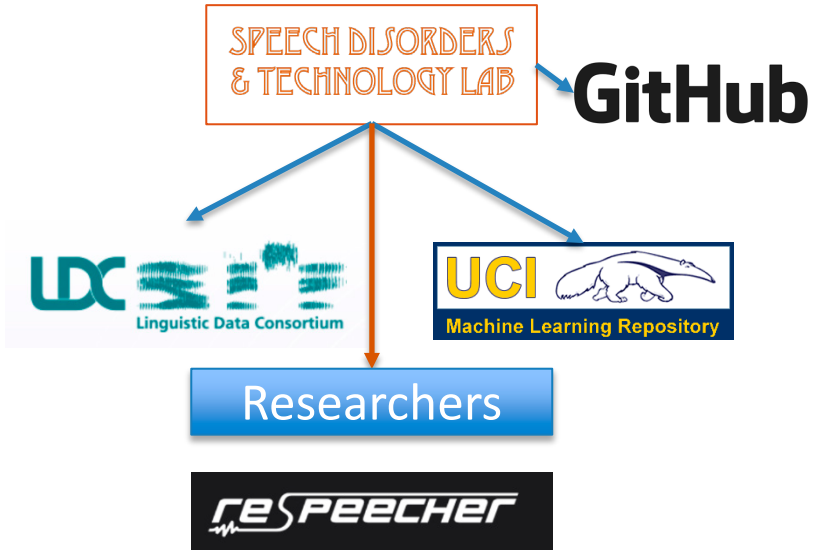
Commercial EMA
(Carstens, Germany)

Our device: MagTrack

- Cao, B., Ravi, S., Sebkhi, N., Bhavsar, A., Inan, O. T., Xu, W., & **Wang, J.** (2022). MagTrack: A wearable tongue motion tracking system for silent speech interfaces, *Journal of Speech, Language, and Hearing Research* (under review)
- Cao, B., Wisler, A., & **Wang, J.** (2022). Speaker adaptation on articulation and acoustics for articulation-to-speech synthesis, *Sensors*, 22, 6056, 1-16.
- Cao, B., Teplansky, K., Sebkhi, N., Bhavsar, A., Inan, O., Samlan, R., Mau, T., & **Wang, J.** (2022). Data augmentation for end-to-end silent speech recognition for laryngectomees, *Proc. Interspeech*, pp. 3653 - 3657.
- Cao, B., Sebkhi, N., Bhavsar, A., Inan, O. T., Samlan, R., Mau, T., & **Wang, J.** (2021). Investigating speech reconstruction for laryngectomees for silent speech interfaces, Proc. Interspeech, pp. 651 - 655.
- Sebkhi, N., Bhavsar, A., Anderson, D. V., **Wang, J.**, & Inan, O. T. (2021). Inertial measurements for tongue motion tracking based on magnetic localization with orientation compensation, *IEEE Sensors Journal*, 21(6), 7964-7971.
- Dinh, T., Kain, A., Samlan, R., Cao, B., & **Wang, J.** (2020). Increasing the intelligibility and naturalness of alaryngeal speech using voice conversion and synthetic fundamental frequency, *Proc. Interspeech*, pp. 4781 - 4785.
- Teplansky, K., Wisler, A., Cao, B., Liang, W., Whited, C. W., Mau, T., & Wang, J. (2020). Tongue and lip motion patterns in alaryngeal speech, *Proc. Interspeech*, pp. 4576 - 4580.
- Teplansky, K., Tsang, B., & **Wang, J.** (2019). Tongue and lip motion patterns in voiced, whispered, and silent vowel production, *Proc. International Congress of Phonetic Sciences*, no. 832, pp. 1-5.
- Cao, B., Tsang, B., & **Wang, J.** (2019). Comparing the performance of individual articulatory flesh points for articulation-to-speech synthesis, *Proc. International Congress of Phonetic Sciences*, no. 761, pp. 1-5.
- Cao, B., Wisler, A., & **Wang, J.** (2019). Is articulation-to-speech synthesis language independent? A pilot study, *Proc. International Congress of Phonetic Sciences*, no. 759, pp. 1-5.
- Cao, B., Sebkhi, N., Mau, T., Inan, O. T., & **Wang, J.** (2019). Permanent magnetic articulograph (PMA) vs electromagnetic articulograph (EMA) in articulation-to-speech synthesis for silent speech interface, *Workshop on Speech and Language Processing for Assistive Technologies* , pp. 17-23.

Highlights – Data Sharing

- Positional (x, y, z coordinates), magnetic, and acoustic data
- Obtained IRB approval for sharing
- Hired a new, external member to check the AI readiness of the data from an outsider perspective
- Data sharing (to public repositories and researchers)
  - Lab website as a portal
  - Github
  - UC Irvine Machine Learning Repository
  - Linguistic Data Consortium (U Penn)
  - Directly to researchers via UT Box (e.g., Respeecher)



Challenges & Future Work

modality.ai

Challenges and Solutions

- Participant recruitment, particularly due to COVID19
  - Solution: remote data collection use a web-based tool (Modality.ai) to collect audio and video (facial movement) data during speech tasks
- PhD student and postdoc recruitment
  - Recruited two postdocs, two PhD students, one research engineer, and a number of undergraduate students so far.
  - No new PhD student in Fall 2022 at UT Austin - Working to hire a new postdoc

Future Work

- Collect more kinematic and acoustic data, as the university is fully open.
- Share to more researchers or communities
- Actively maintain the shared data, code, and documentation

Acknowledgments

Team
Jun Wang, PhD, UT Austin
Omer Inan, PhD, Georgia Tech
Nordine Sebkhi, PhD, Georgia Tech
Arpan Bhavsar, Georgia Tech
Cem O. Yaldiz, Georgia Tech
Beiming Cao, UT Austin
Kristin Teplansky, UT Austin
Cissy Cheng, UT Austin
Carl Sigmond, UT Austin
Ted Mau, MD, PhD, UT Southwestern Medical Center
Robin Samlan, PhD, CCC-SLP, University of Arizona

Thanks to all research participants!

Funding
- R01DC016621, 2019-2024 (PI: Wang)
- 3R01DC016621-03S1, 2021-2024 (NCE) (PI: Wang)

NIH National Institute on Deafness and Other Communication Disorders