Tongue and Lip Motion Patterns of Alaryngeal and Silent Speech

Kristin J. Teplansky[1], Satwik Dutta[2], Beiming Cao[1], Jun Wang[1]
[1]The University of Texas at Austin, [2]The University of Texas at Dallas

Introduction

- A laryngectomy is the surgical removal of the larynx due to oral or laryngeal cancer [1]. Silent speech interface (SSI) technology allows for communication without the vibration of the vocal folds by converting articulatory motion data to an acoustic output [2]. A better understanding of the kinematic patterns of alaryngeal speech will be helpful in the development of this technology.
- Prior work shows that articulatory patterns differ depending on laryngeal activity and the loss of auditory feedback [3, 4]. Previous research has largely overlooked the degree to which alaryngeal articulatory patterns resemble healthy speech in different speaking modes. A better understanding of the patient population is necessary to incorporate SSI technology into practice.
- **The aim of this study** was to characterize articulatory movements during the production of phrases obtained from healthy (voiced, silent) speakers and laryngectomees.

Methods

Participants and speech tasks

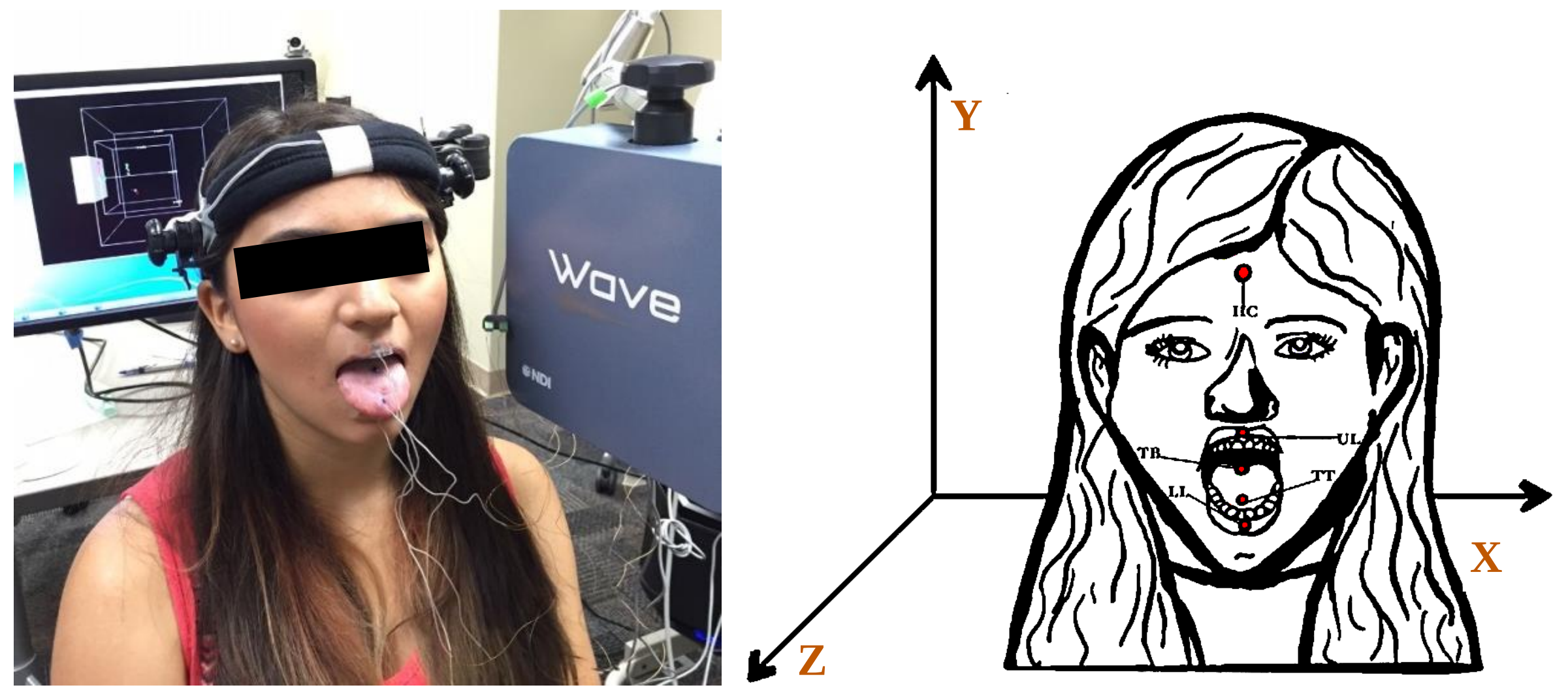
- 6 speakers (3 male/3 female) aged 22 – 52 years (M_{age} = 28.66)
- 2 healthy voiced speakers, 2 healthy silent speakers, 2 TEP alaryngeal speakers.
- Each participant produced 60 phrases one time.

Tongue and lip motion tracking

- The Wave System was used to derive synchronized acoustic and tongue and lip motion data (Figure 1a) with a spatial accuracy of 0.5mm.
- Sampling rate is 100Hz.
- Four sensors (Figure 1b). attached to the tongue tip (TT), tongue back (TB), upper lip (UL), lower lip (LL).

Data processing

- Head-independent data
- Low-pass filter with a cutoff frequency of 20Hz was applied to kinematic data.


[redacted]

(a) Wave System

(b) Sensor Locations

Figure 1. *Data collection setup*

Data Analysis

1. Duration (sec.): measured from phrase onset to offset.

2. Average speed (mm/s): average of instant speed, calculated as the change in displacement over time.

3. Range (mm): maximum position subtracted from the minimum position.

- ➢ **Statistical Analysis:** one-way ANOVAs.

4. Machine learning classification:

- ➢ **Support vector machine (SVM):** is a soft margin machine learning classifier that implements a linear hyperplane to separate classes.
- Model performance was evaluated using a 5-fold cross-validation method.
- **Features:** average speed and range across *y* and *z*-dimensions.

Results & Discussion

Kinematic measures

- **Duration (sec.)**: Alaryngeal tongue and lip movements show a significantly longer duration of movement during the production of phrases than voiced and silent speech (Figure 2).

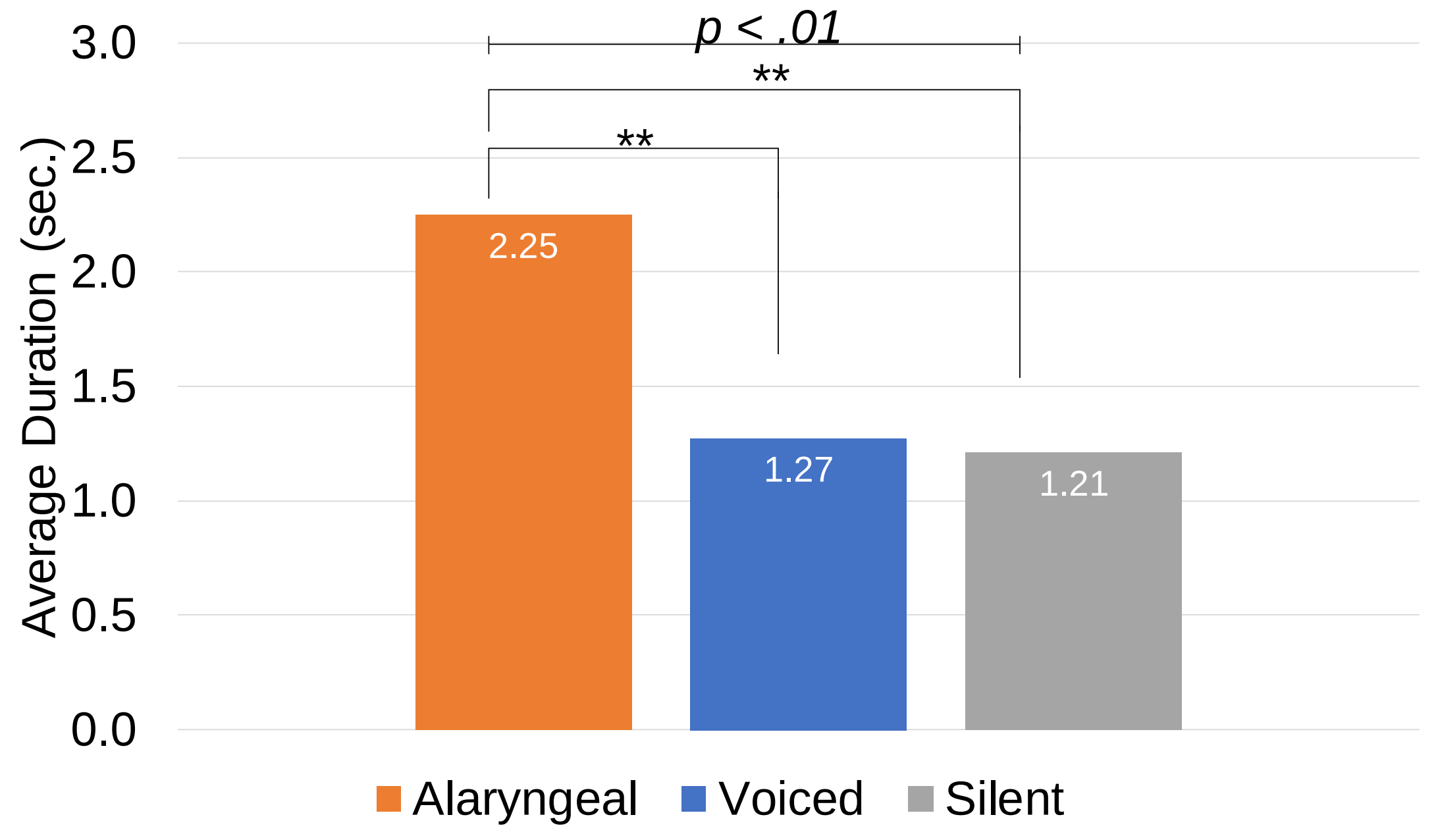


Figure 2. *Duration of articulatory movement for three types of speech.*

- **Speed (mm/s)**: Silent tongue movements were significantly slower than voiced speech for all sensors. This finding is consistent with prior research using phrases and vowels [3, 4]. However, the UL and LL were faster than the voiced condition (Figure 3).

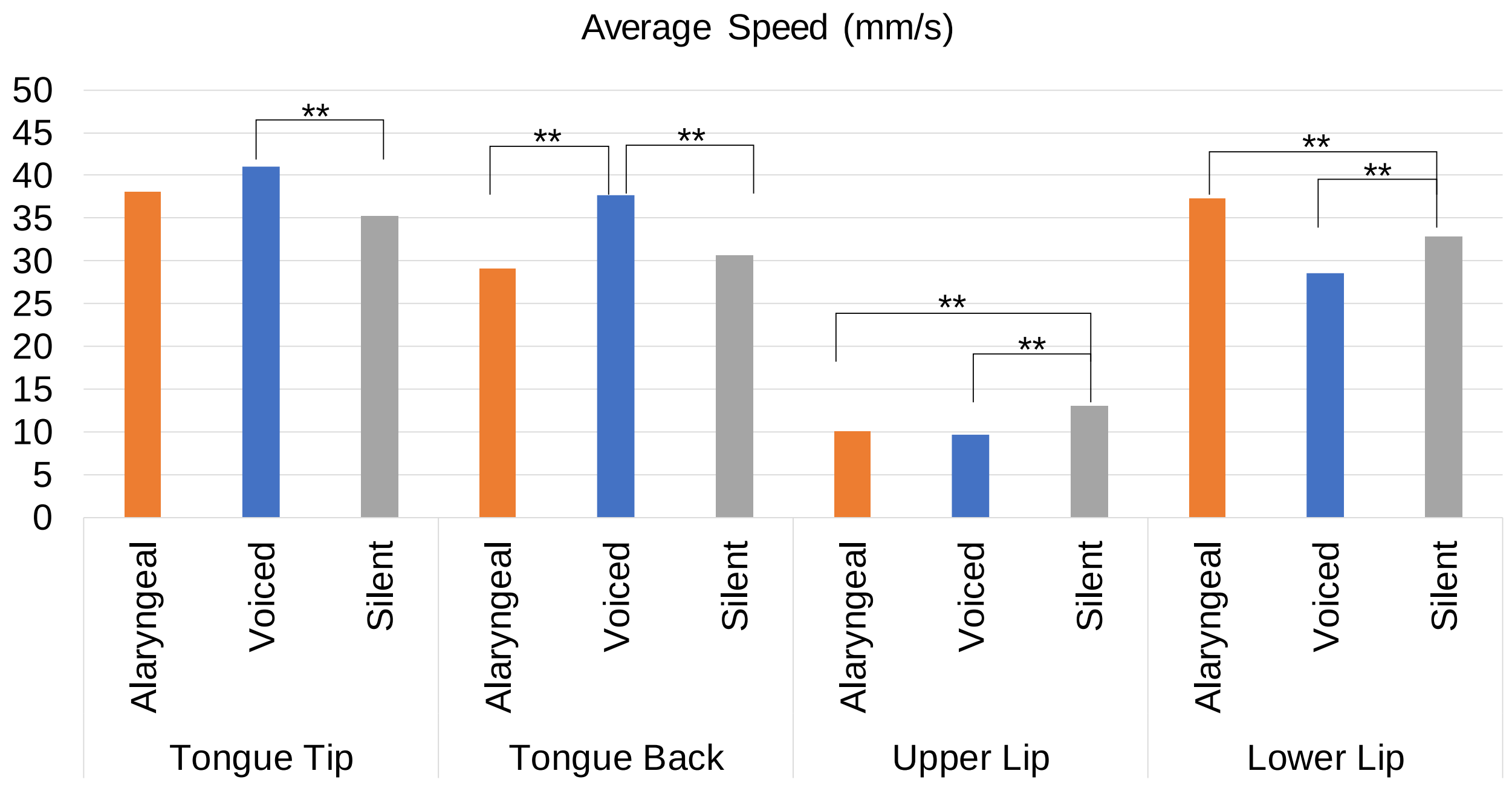


Figure 3. *Average speed of articulatory movement for all sensors.*

Continued

- **Range (mm):** Alaryngeal TT, UL, LL range of movement was significantly larger than voiced and silent speech. The silent speech condition showed more similar sensor movement range to the voiced condition than the alaryngeal speakers (Figure 4).

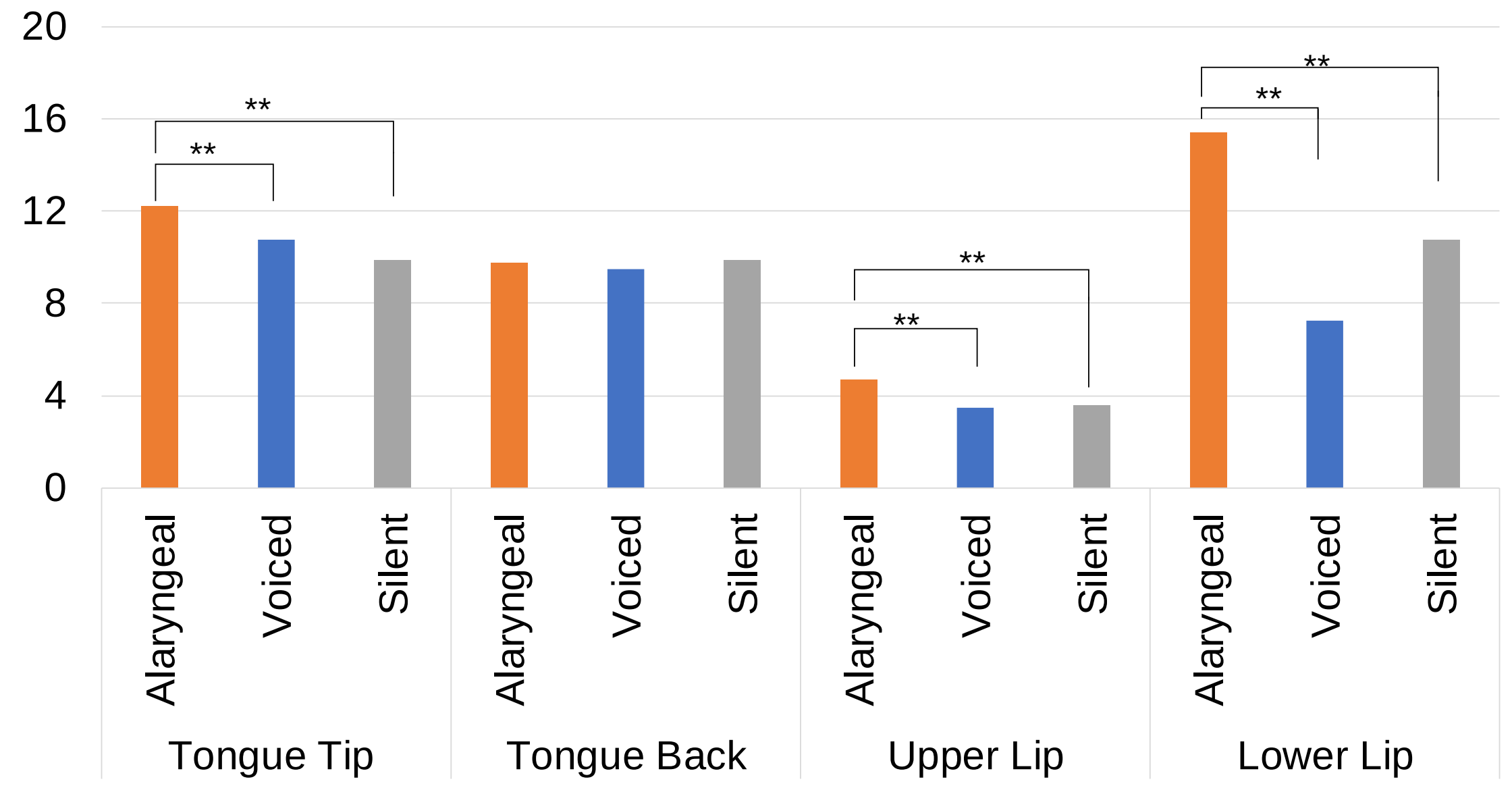


Figure 4. *Range of articulatory movement for all sensors (y-dimension) in three speech conditions.*

SVM

- Overall classification accuracy of the three speech conditions was 99.05% (Table 1).

Table 1. SVM classification matrix.

Truth\Predicted	Alaryngeal	Voiced	Silent	*Total Truth*
Alaryngeal	**77**	2	0	79
Normal	1	**116**	0	117
Silent	0	0	**117**	117
Total Predicted	78	118	117	**313**

Conclusion & Future Work

- The data provide preliminary evidence to suggest that articulatory strategies are impacted by laryngeal activation.
- Alaryngeal speakers show longer tongue and lip duration and a larger TT, UL, LL range of movement than voiced and silent speakers. Silent speech produced by healthy speakers may be more similar to voiced speech than alaryngeal speech. The duration in this study is different than [3, 4], possibly due to the small number of participants.
- Future work will include a larger sample size, additional features, and machine learning algorithms to further investigate alaryngeal speech.

Acknowledgements

This work was supported by the National Institutes of Health under award numbers R03DC013990 and R01DC016621, and by the American Speech-Language-Hearing Foundation through a New Century Scholar Research Grant. We thank Alyssa Shrode and the volunteering participants.

References

[1] Bailey, B. J., Johnson, J. T., & Newlands, S. D. (2006). *Head & neck surgery – otolaryngology.* Philadelphia: Lippincott Williams & Wilkins.
[2] Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J. M., Brumberg, J. S.(2010). Silent speech interfaces. *Speech Communication*. 52(4), 270–287.
[3] Dromey, C., & Black, K. M. (2017). Effects of Laryngeal Activity on Articulation. *IEEE/ACM Transactions on Audio Speech and Language Processing*, *25*(12),
[4] Teplansky, K., Tsang, B., Wang, J. (2019). Tongue and lip motion patterns in voiced, whispered, and silent vowel production. *International Congress of Phonetic Science.*