EMG of Facial Muscles for Subvocal Speech Recognition

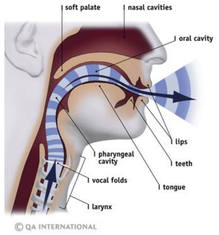
DELSYS®
SENSORS FOR MOVEMENT SCIENCES

Geoffrey S. Meltzner[2], James T. Heaton[4], Yunbin Deng[3], Bhawna Shiwani[1], Gianluca De Luca[1], Serge H. Roy[1], Joshua C. Kline[1]

[1]Delsys, Inc and Altec Inc, Natick, USA, [2]VocaliD, Inc. Belmont, USA, [3]BAE Systems, Burlington, USA, [4]Massachusetts General Hospital Voice Center, Boston, USA

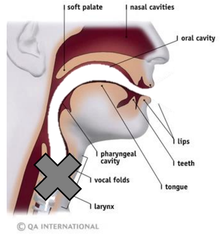
Background

Natural Speech Vocalization



- Air pressure from lungs
- Frequency (pitch) variation from larynx
- Auditory filtering from tongue, cheeks, lips, etc.

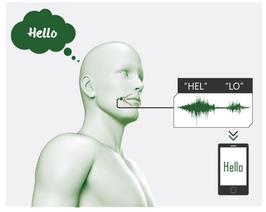
Impaired Speech Vocalization



- 60,000 cases of cancer in mouth, throat and larynx each year in US
- Most result in vocal impairment
- In such cases, speech musculature still remain intact and can provide an Alternative Speech source

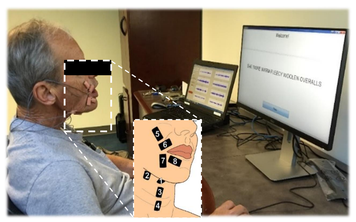
Objective

To provide sEMG-based Subvocal Speech Recognition platform that utilizes silently mouthed (subvocal) speech for Augmentative and Alternative Communication (AAC).



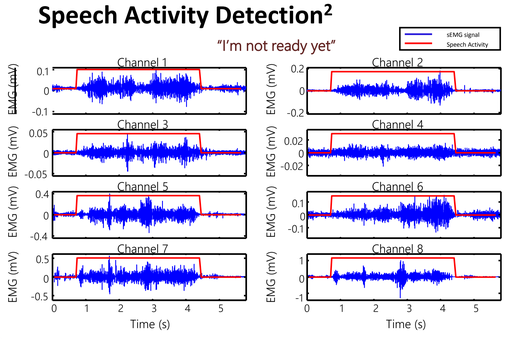
Methods

Experiment Setup[1]



Data Collection	
Subject Population	n = 7 Total Laryngectomy
Sensor Type	Trigno wireless sEMG (Delsys, Inc)
Sensor Location	
1,2	Neck – submental
3,4	Neck – ventromedial
5,6	Face - supralabial
7,8	Face - infralabial
Data Corpus	2500 words/ 980 continuous sentences

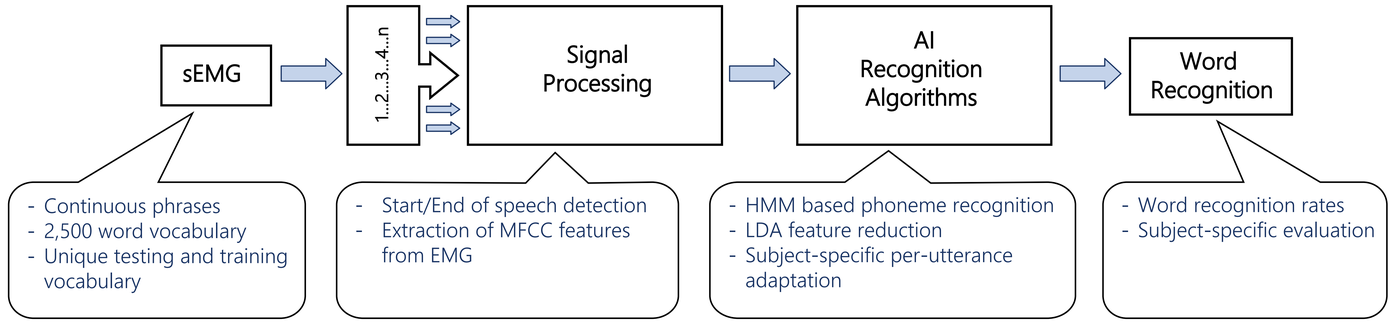
Speech Activity Detection[2]



- Unique EMG activation across muscles
- Separated speech from non-speech activity by identifying simultaneous multi-channel activation
- Robust against single channel noise

Subvocal Algorithm Architecture

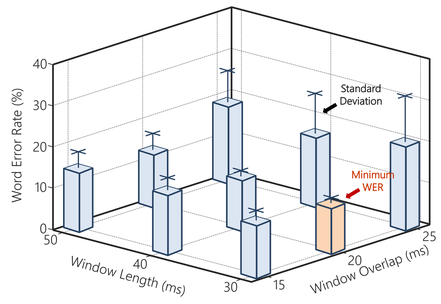
Front-end Signal Processing + Back-end Machine Learning



sEMG Speech Performance Findings

Key Finding 1

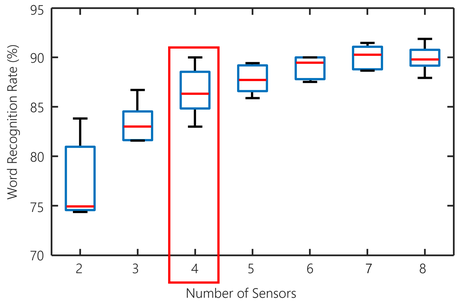
Average WER as a function of Window Length, Overlap (ms)



- Word Error Rate (WER) is highly influenced by window lengths and overlaps selected
- Most effective Window/Overlap pair: 30ms/20ms
- Mean WER at 30ms/20ms: 11.4%

Key Finding 2

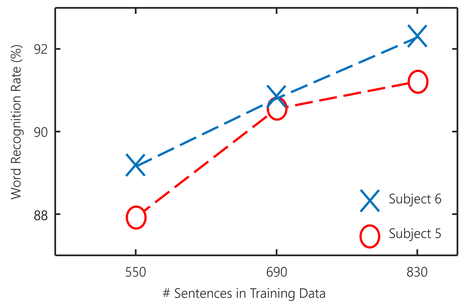
Word Recognition as a function of the Number Sensors



- Increase in number of sensors is exponentially related to increase in Word Recognition
- Optimal number of sensors for facially worn interface: 4

Key Finding 3

Word Recognition vs Training Data



- Word Recognition improves with increase in number of sentences used for training
- Training data expansion has potential to improve subvocal speech recognition

Conclusion

- Proof of concept of first sEMG –based subvocal speech recognition system that provides 85-90% WERs for continuous speech from 2500 word vocabulary
- Restores sense of embodiment & intuitive speech by augmenting natural voice musculature
- With larger data corpus and optimal parameters guided by key findings, this technology has the potential to provide AAC that approximates natural speech vocalization

Acknowledgements

- VocalID, Inc. Belmont, USA
- MGH Voice Center, Boston, USA
- BAE Systems, Inc. Burlington, USA

Support

DE LUCA FOUNDATION

NIH National Institute on Deafness and Other Communication Disorders
(R44DC014870)

References

1. Meltzner et. al. Transactions on Audio, Speech and Language Processing,2016
2. Meltzner et. al. Annual International Conference of the IEEE EMBS, 2011