Design and Development of sEMG-based Silent Speech Recognition

DELSYS®
WEARABLE SENSORS FOR MOVEMENT SCIENCES

Serge H. Roy[1], Geoffrey S. Meltzner[2], James T. Heaton[3], Yunbin Deng[4], Bhawna Shiwani[1], Gianluca De Luca[1], Joshua C. Kline[1]
[1]Delsys, Inc and Altec Inc, Natick, USA, [2]VocaliD, Inc. Belmont, USA, [3]Harvard Medical School Department of Surgery, MGH, Boston, USA, [4]BAE Systems, Burlington, USA,

Background

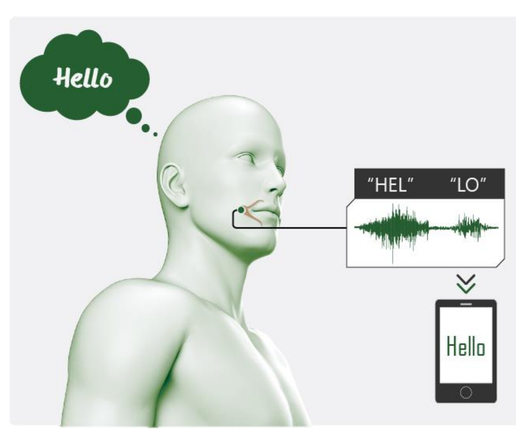
Speech provides an attractive modality for human-machine interface (HMI) through automatic speech recognition (ASR). But ASR suffers from three primary limitations:

1) Performance degradation in presence of ambient noise
2) Limited ability for privacy/secrecy
3) Poor accessibility for those with speech disorders.

These limitations have motivated the need for alternative non-acoustic modalities of subvocal or silent speech recognition (SSR).

Objective

We set out to design and develop a SSR system based on recordings of the surface electromyographic (sEMG) signal from articulator muscles of the face and neck during silently mouthed (subvocal) speech.



Methods

1) Experiment Setup

- **Subjects –** n=18 healthy (11 males, 8 females, age range 18-42 y.o.)
- **sEMG Sensors**: 8 DE 2.1 sensors and Trigno™ Mini sensors (Delsys, Natick, USA)
- **Protocol** – Subjects silently mouthed words while sEMG activity was recorded from muscles of face and neck.

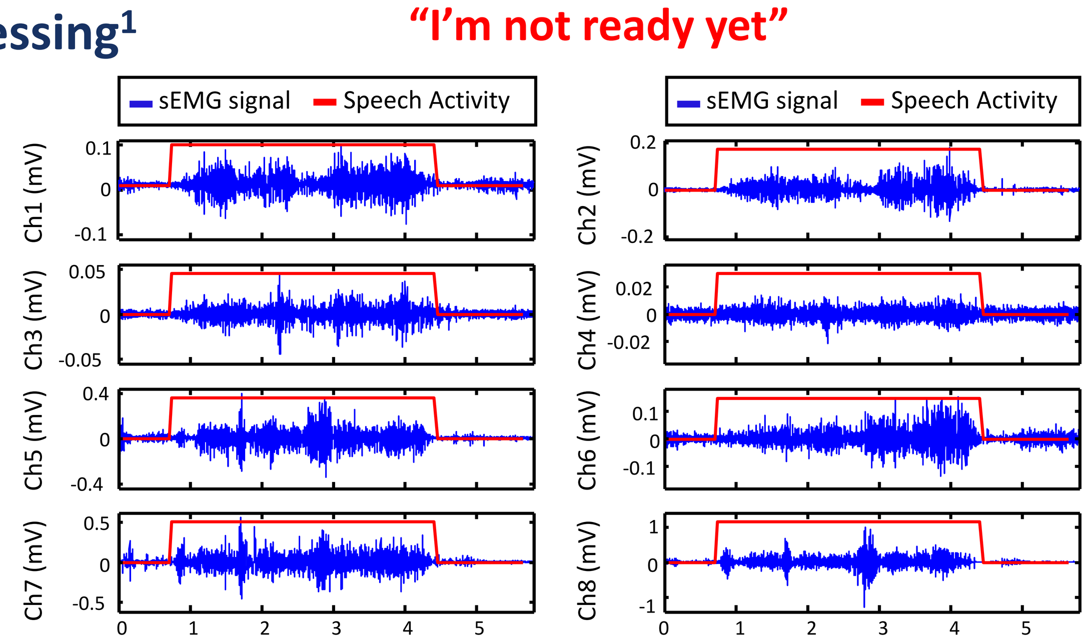
2) Data Collection

Data Corpus	Subjects	Vocabulary/Phrases
Isolated Words	Controls (n=9)	65 words and digits
Sequences of Words	Controls (n=4)	202 words, 1,200 sequences
Continuous Speech	Controls (n=5)	2,200 words, 1,200 phrases

3) Initial Data Processing[1]

- Separating speech from non-speech sEMG activity
- Finite multi-channel state machine
- Robust against noise

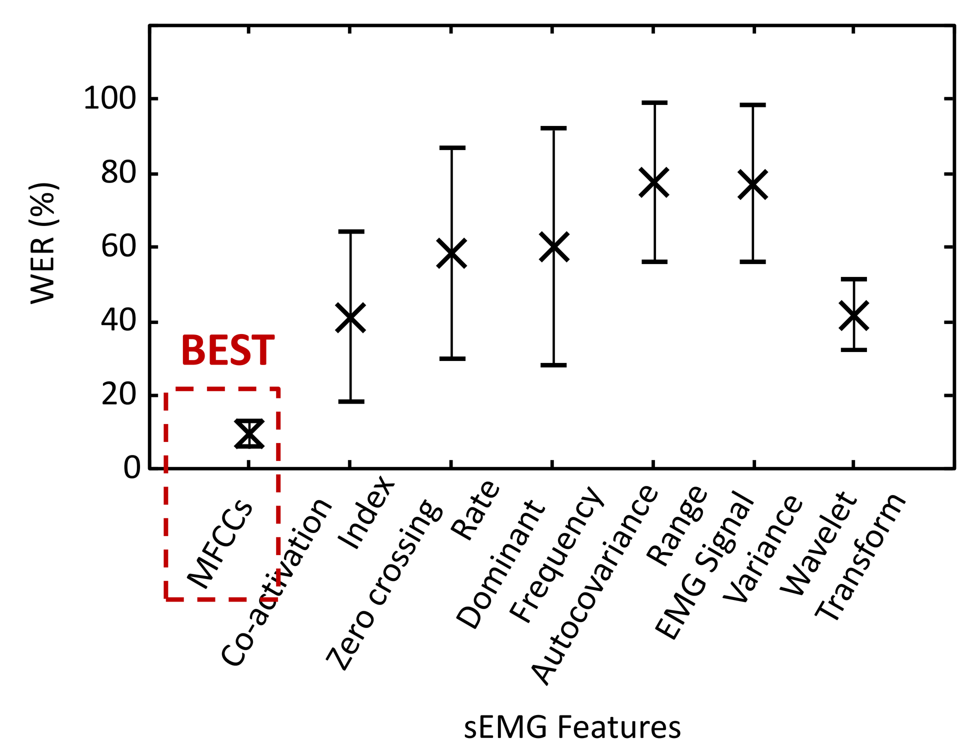
"I'm not ready yet"



Algorithm Development – Strategic evolution of Hidden Markov Models (HMMs) for SSR

Challenge 1

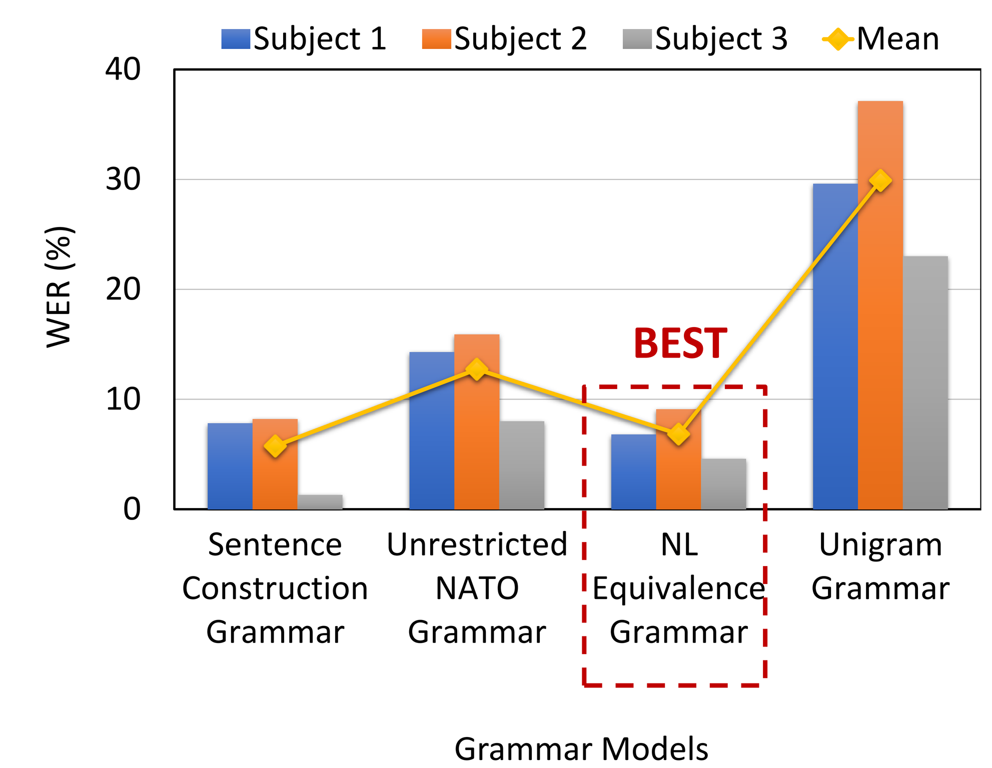
Discriminating isolated words using sEMG-based speech related features



- Mel Frequency Cepstral Coefficients (MFCCs) provided the lowest word error rate (WER) of 9.6% across of all combinations of features tested.

Challenge 2

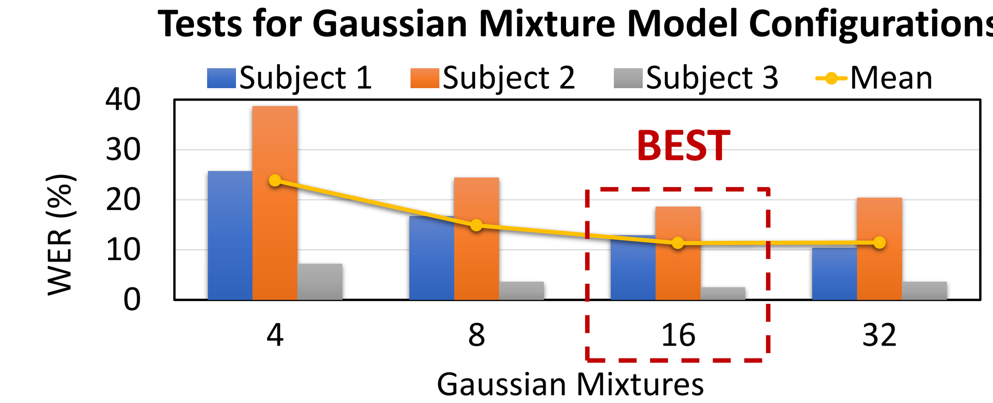
Tracking sequences of words from patterns of sEMG signals using grammatical context



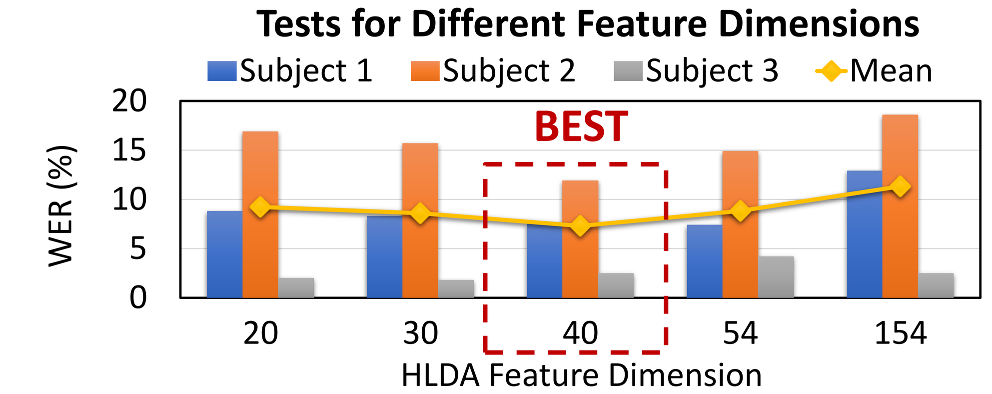
- Natural Language (NL) Equivalence Grammar provided a larger range of linguistically correct English sentences and had the second lowest WER of 6.8%.

Challenge 3

Recognizing a large vocabulary of untrained words using phoneme-based models



- Increasing to 16 Gaussian mixtures to model each HMM state decreased WER to 11.3%



- Reducing the HLDA feature dimension to 40 decreased the WER to 7.3%

Final sEMG SSR System – Algorithms, Sensors and Mobile Deployment

SSR Configuration:

- *Sensors* – sEMG 4 sensor-array (under development) worn on face and neck
- *Features* – MFCCs
- *Grammar* – NL Equivalence
- *Recognition Toolkit* – KALDI
- *Model* – HMM Triphone, HLDA Feature Reduction, maximum likelihood linear regressions (MLLR), subspace Gaussian mixture modelling (SGMM)

SSR Sensors: Trigno™ Quattro Facial Array

- Wireless/Bluetooth communication for mobile use

SSR System Performance (WER):

Subject	Digits	Text Messages	Special Operations	Common Phrases	Mean WER
1	2.7	0.9	2.0	0.0	1.4
2	15.4	15.9	8.0	15.4	13.9
3	20.7	12.1	13.6	5.2	12.9
4	18.2	10.3	10.6	3.7	10.7
5	12.2	5.6	3.4	1.2	5.6
Mean	13.8	9.0	7.5	5.1	**8.9**
SD.	7.0	5.8	4.9	6.1	**5.3**

FINAL

Conclusion

- Our SSR system was able to recognize subvocal speech with 8.9% WER from a 2,200-word vocabulary of 1,200 continuous phrases including previously unseen words.
- The miniaturized sensors provide a robust and unencumbering facial interface that can transmit data via custom wireless or Bluetooth protocol for portable integration with a mobile device.
- These results demonstrate the viability of our SSR system as a silent modality of speech communication that can be developed further for persons with speech impairments (Meltzner et al, 2017), military personnel, or consumer applications.

Acknowledgements

- VocaliD, Inc. Belmont, USA
- MGH, 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. Silent Speech Recognition as an Alternative Communication Device for Persons With Laryngectomy. IEEE Trans. on ASLP, 2017.