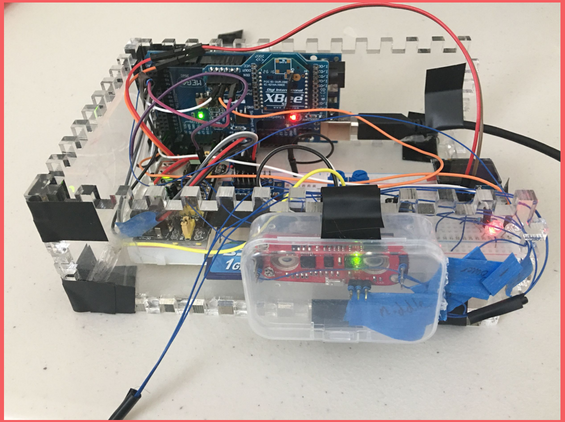
SpeakUp: A Machine Learning Based Speech Aid to Enable Real-Time Silent Communication for the Paralyzed by Translating Neuromuscular EMG signals to Speech

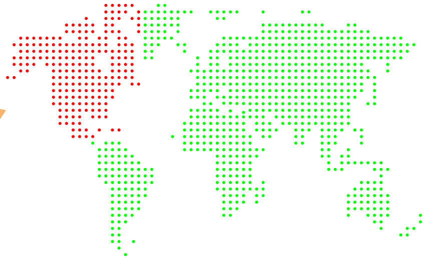
SpeakUp
- A ML Based Speech Aid to Enable Silent Communication



Varun Chandrashekhar

Introduction to Speech Interfaces

Stroke / ALS / Cerebral Palsy



7.5 Million

Conventional Speech Interfaces (CSI) - Eye/Cheek Tracking

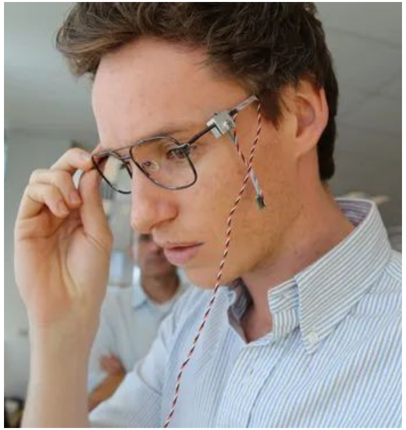


Image shows cheek tracker used by Stephen Hawking to communicate

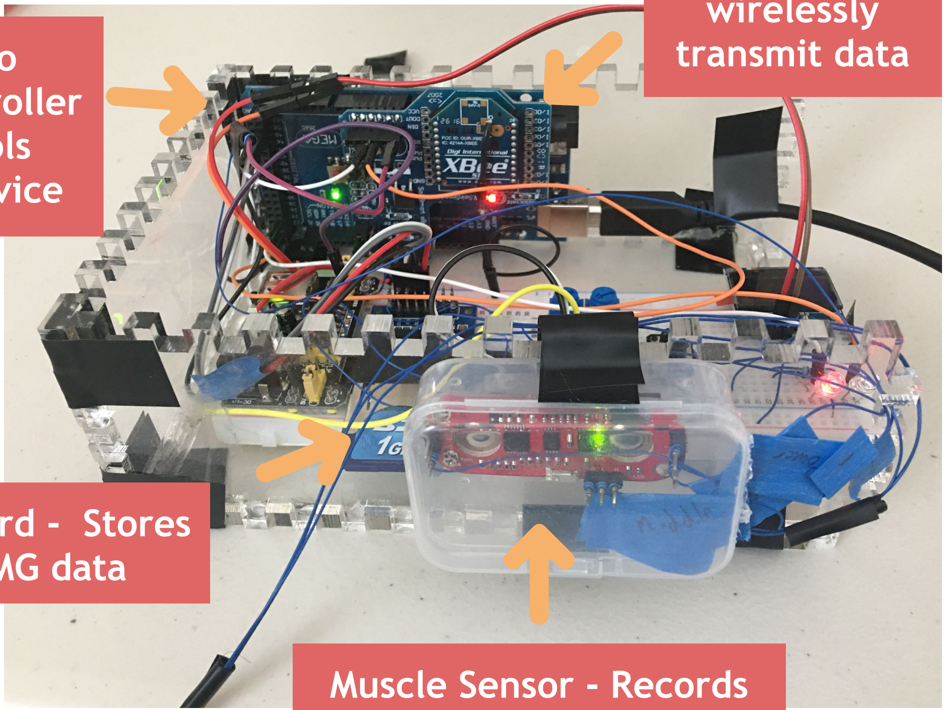
Fatiguing

Slow / Inaccurate

SpeakUp Hardware

Arduino Microcontroller - controls whole device

Radio to wirelessly transmit data

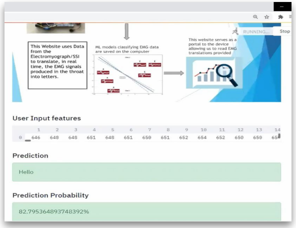


SD Card - Stores EMG data

Muscle Sensor - Records EMG Signals

EMG Signals Picked Up by Device and wirelessly sent to Computer

Computer Automatically Translates EMG Signal



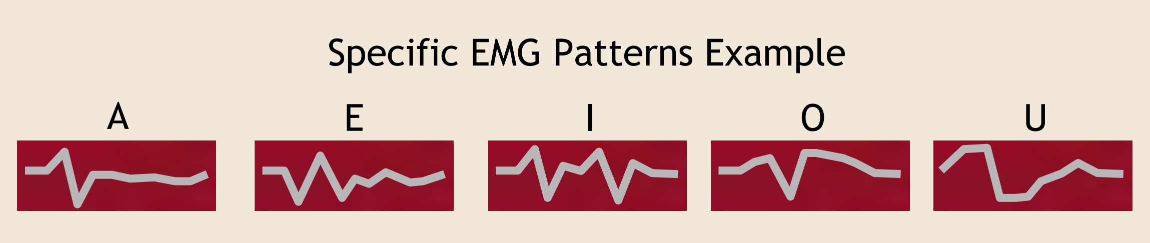
Translation is displayed on Python Based Web Application

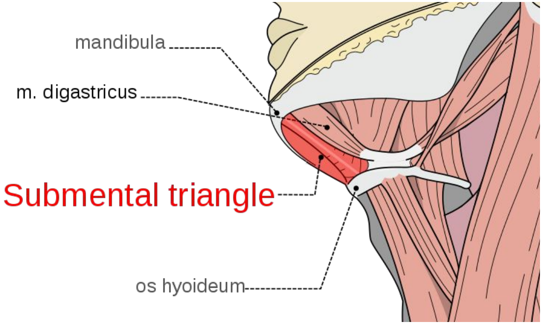
Silent Speech

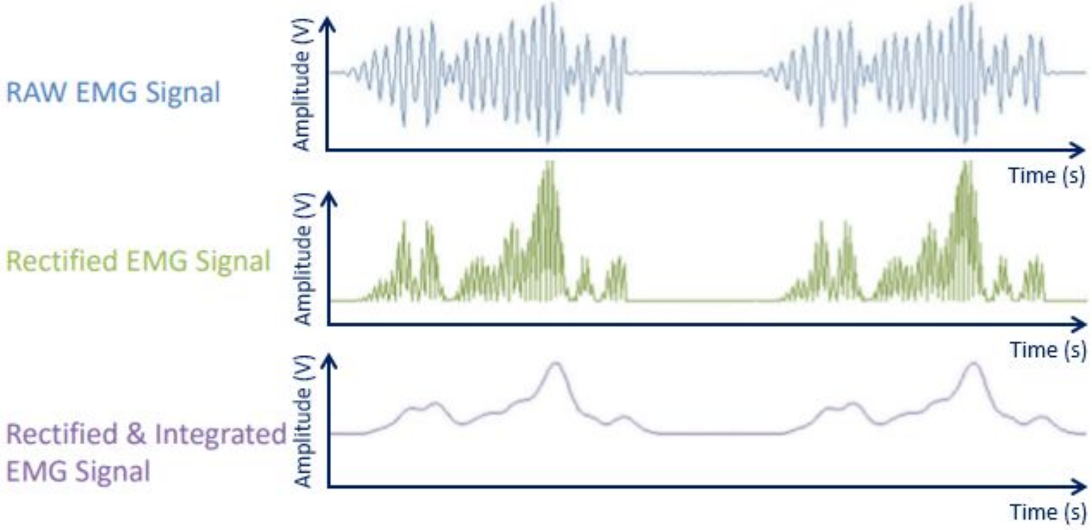
Silent Speech: minimally or internally articulating words without producing sounds or moving the mouth

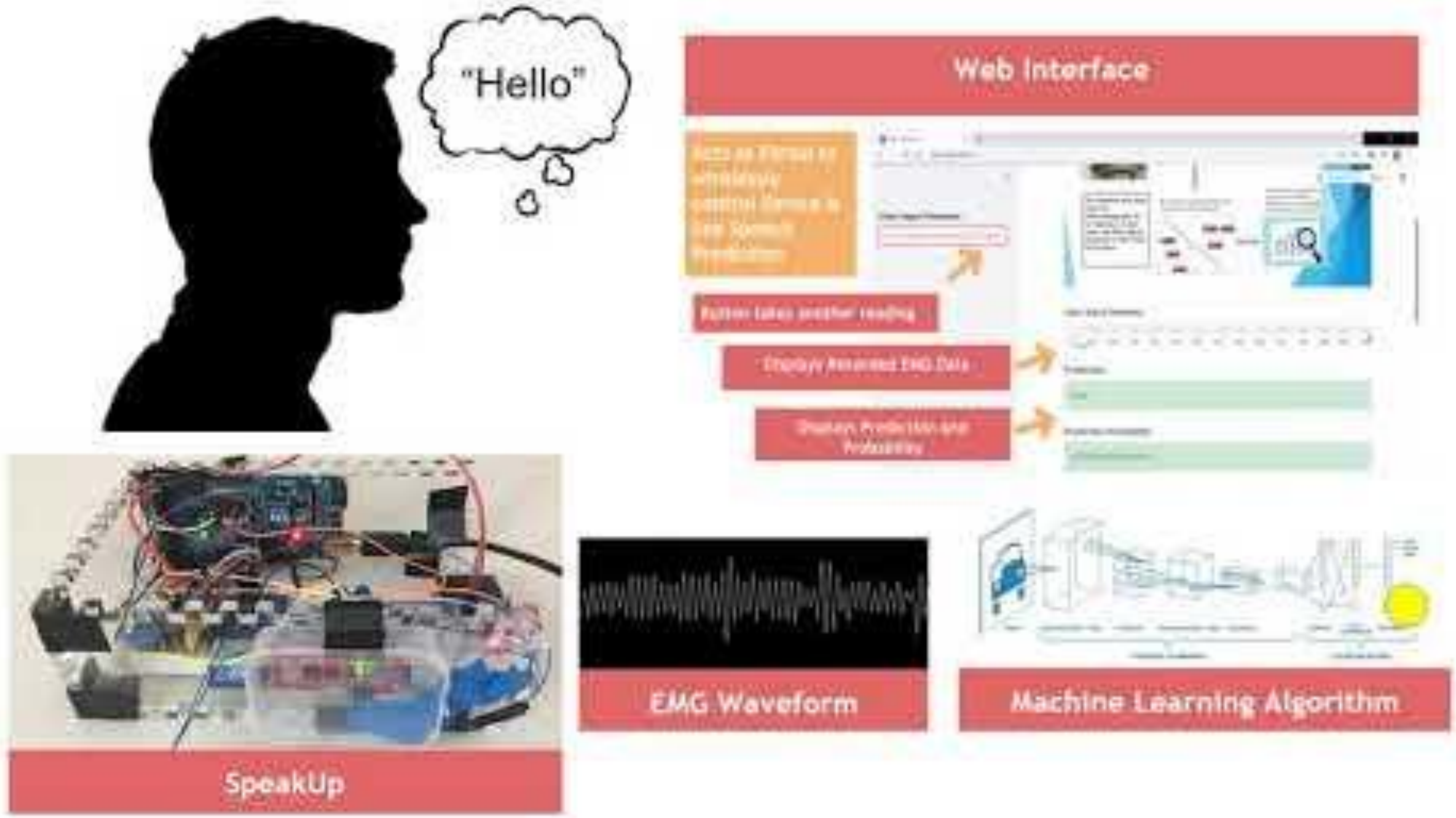
ElectroMyoGraph (EMG) signals: electrical signals generated by muscle contractions

sEMG signals: EMG signals collected from surface electrodes into real time speech



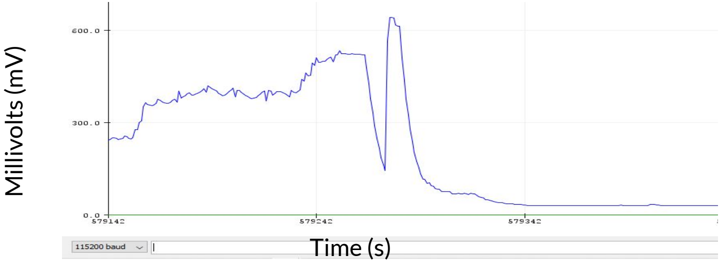




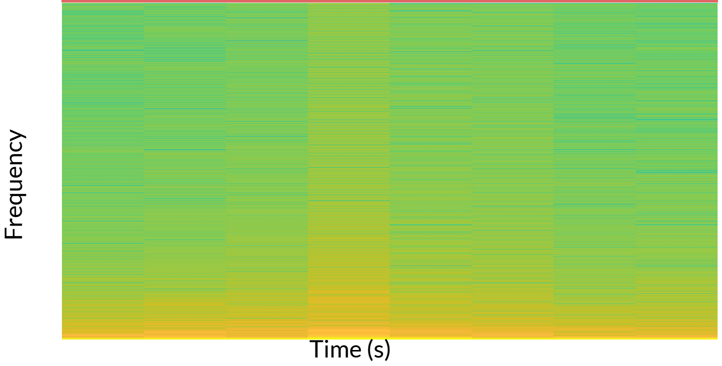

"Hello"
Web Interface
Button takes another reading
EMG Waveform
Machine Learning Algorithm
SpeakUp

Raw EMG Signals

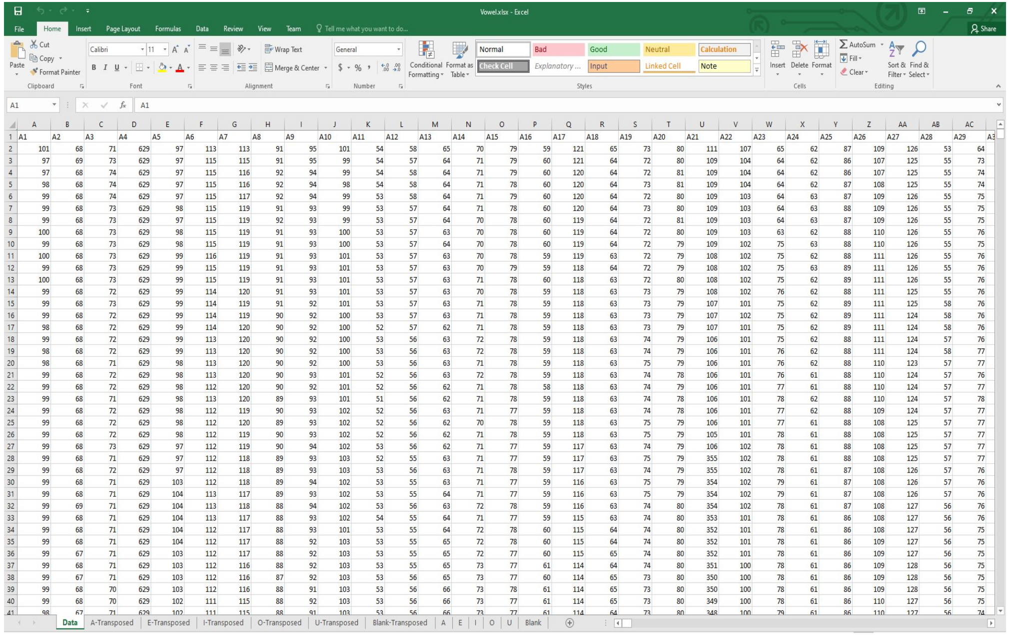
Graphed



Spectrogram



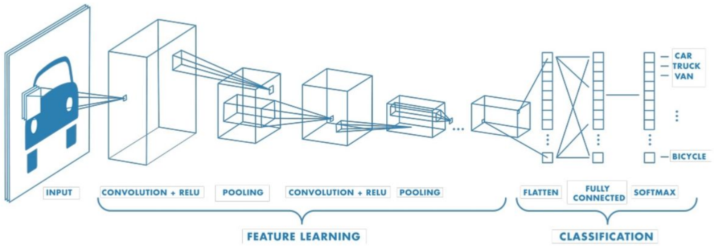
Excel - Dataset



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23	A24	A25	A26	A27	A28	A29
2	101	68	71	629	97	113	113	91	95	101	54	58	65	70	79	59	121	65	73	80	111	107	65	62	87	109	126	53	64
3	97	69	73	629	97	115	115	91	95	99	54	57	64	71	79	60	121	64	72	80	109	104	64	62	86	107	125	55	73
4	97	68	74	629	97	115	116	92	94	99	54	58	64	71	79	60	120	64	72	81	109	104	64	62	86	107	125	55	74
5	98	68	74	629	97	115	116	92	94	98	54	58	64	71	78	60	120	64	73	81	109	104	64	62	87	108	125	55	74
6	99	68	74	629	97	115	117	92	94	99	53	58	64	71	79	60	120	64	72	80	109	103	64	63	87	109	126	55	75
7	99	68	73	629	98	115	119	91	93	99	53	57	64	71	78	60	120	64	73	80	109	103	64	63	88	109	126	55	75
8	99	68	73	629	97	115	119	92	93	99	53	57	64	70	78	60	119	64	72	81	109	103	64	63	87	109	126	55	75
9	100	68	73	629	98	115	119	91	93	100	53	57	63	70	78	60	119	64	72	80	109	103	63	62	88	110	126	55	76
10	99	68	73	629	98	115	119	91	93	100	53	57	64	70	78	60	119	64	72	79	109	102	75	63	88	110	126	55	75
11	100	68	73	629	99	116	119	91	93	101	53	57	63	70	78	59	119	63	72	79	108	102	75	62	88	111	126	55	76
12	99	68	73	629	99	115	119	91	93	101	53	57	63	70	79	59	118	64	72	79	108	102	75	63	89	111	126	55	76
13	100	68	73	629	99	115	119	91	93	101	53	57	63	71	78	60	118	63	72	80	108	102	75	62	89	111	126	55	76
14	99	68	72	629	99	114	120	91	93	101	53	57	63	70	78	59	118	63	73	79	108	102	76	62	88	111	125	55	76
15	99	68	73	629	99	114	119	91	92	101	53	57	63	71	78	59	118	63	73	79	107	101	75	62	89	111	125	58	76
16	99	68	72	629	99	114	119	90	92	100	53	57	63	71	78	59	118	63	73	79	107	102	75	62	89	111	124	58	76
17	98	68	72	629	99	114	120	90	92	100	52	57	62	71	78	59	118	63	73	79	107	101	75	62	89	111	124	58	76
18	99	68	72	629	99	113	120	90	92	100	53	56	63	72	78	59	118	63	74	79	106	101	75	62	88	111	124	57	76
19	98	68	72	629	99	113	120	90	92	100	53	56	63	72	78	59	118	63	74	79	106	101	76	62	88	111	124	58	77
20	98	68	71	629	98	113	120	90	92	100	53	56	63	71	78	59	118	63	75	79	106	101	76	62	88	110	123	57	77
21	99	68	72	629	98	113	120	90	93	101	52	56	63	72	78	59	118	63	74	78	106	101	76	61	88	110	124	57	76
22	99	68	72	629	98	112	120	90	92	101	52	56	62	71	78	58	118	63	74	79	106	101	77	61	88	110	124	57	77
23	99	68	71	629	98	113	120	89	93	101	51	56	62	71	78	59	118	63	74	78	106	101	78	62	88	110	124	57	78
24	99	68	72	629	98	112	119	90	93	102	52	56	63	71	77	59	118	63	74	78	106	101	77	62	88	109	124	57	77
25	99	68	72	629	98	112	120	89	93	102	52	56	62	70	78	59	118	63	75	79	106	101	77	61	88	108	125	57	77
26	99	68	72	629	98	112	119	90	93	102	52	56	62	71	78	59	118	63	75	79	105	101	78	61	88	108	125	57	77
27	99	68	73	629	97	112	119	90	94	102	53	56	62	71	77	59	117	63	74	79	106	102	78	61	88	108	125	57	77
28	99	68	71	629	97	112	118	89	93	103	52	55	63	71	77	59	117	63	75	79	355	102	78	61	88	108	125	57	77
29	99	68	72	629	97	112	118	89	93	103	53	56	63	71	78	59	117	63	74	79	355	102	78	61	87	108	126	57	76
30	99	68	71	629	103	112	118	89	94	102	53	55	63	71	77	59	116	63	75	79	354	102	79	61	87	108	126	57	76
31	99	68	71	629	104	113	117	89	93	102	53	55	64	71	77	59	116	63	75	79	354	102	79	61	87	108	126	57	76
32	99	69	71	629	104	113	118	88	94	102	53	56	63	72	78	59	116	63	74	80	354	102	78	61	87	108	127	56	76
33	99	68	71	629	104	113	117	88	93	102	54	55	64	71	77	59	115	63	74	80	353	101	78	61	86	108	127	56	76
34	99	68	71	629	104	112	117	88	93	101	53	55	64	72	78	60	115	64	74	80	352	101	78	61	86	108	127	56	75
35	99	68	71	629	104	112	117	88	92	103	53	55	65	72	78	60	115	64	74	80	352	101	78	61	86	109	127	56	75
36	99	67	71	629	103	112	117	88	92	103	53	55	65	72	77	60	115	65	74	80	352	101	78	61	86	109	127	56	75
37	99	68	71	629	103	112	116	88	92	103	53	55	65	73	77	61	114	64	74	80	351	100	78	61	86	109	128	56	75
38	99	67	71	629	103	112	116	87	92	103	53	56	65	73	77	60	114	65	73	80	350	100	78	61	86	109	128	56	75
39	99	68	70	629	103	112	116	88	91	103	53	56	66	73	78	61	114	65	73	80	350	100	78	61	86	109	128	56	75
40	99	68	70	629	102	111	115	88	92	103	53	56	66	73	77	61	114	65	73	80	349	100	78	61	86	110	127	56	75
41	98	67	71	629	102	111	115	88	91	103	53	56	66	73	77	61	114	64	73	80	348	100	79	61	86	110	127	56	74

Machine Learning Algorithms used in SSIs

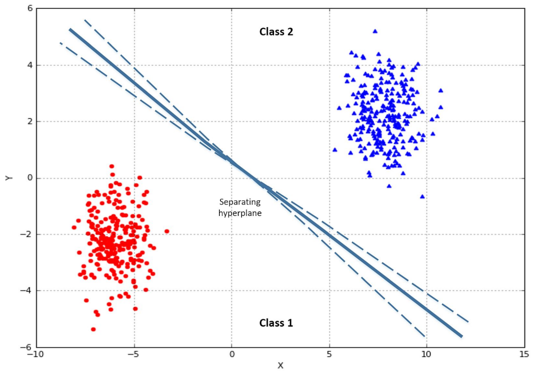
Convolutional Neural Networks (CNN)



CNN passes images through a series of layers that attempt to **learn features and classify the images** into different categories

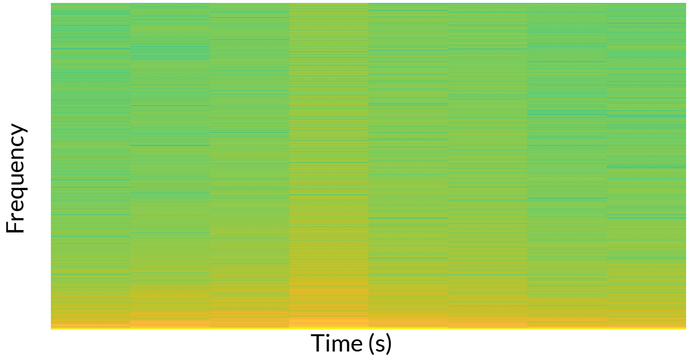
Pattern Recognition

Pattern Recognition algorithms attempt to **develop a division between multiple classes** which are represented in red and blue in this image.

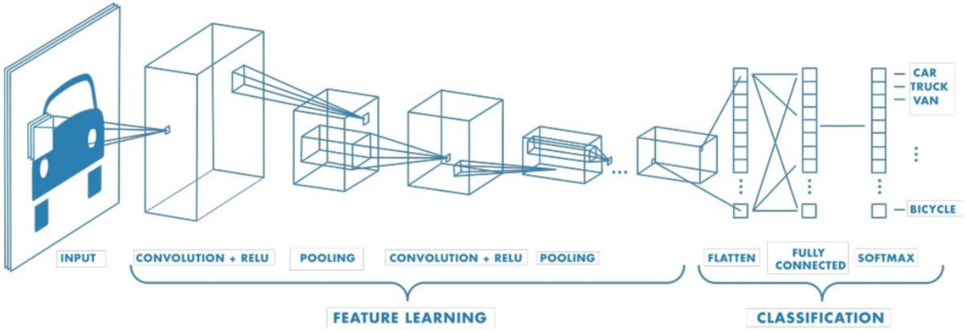


CNN - GoogleNet

- GoogleNet - Most Accurate CNN developed
- Transfer Learning - Making use of an existing CNN for a different application
- Transfer Learning was used to develop a CNN

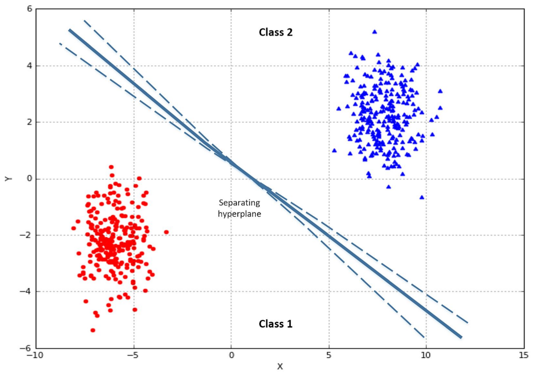


Spectrogram for Letter A



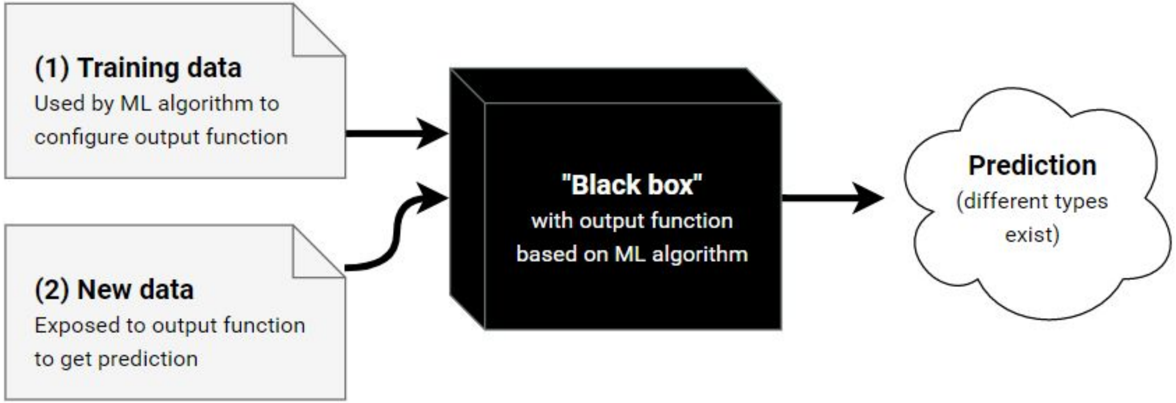
Pattern Recognition

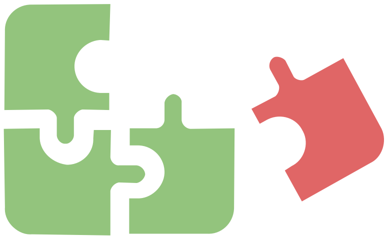
- Various ways to create a division between two classes
- These 7 Pattern Recognition Algorithms were tested in this study



Algorithm		Division/Assumption
Linear Discriminant	→	Linear Line
Quadratic Discriminant	→	Quadratic Line
Naive Bayes	→	Independent Predictors
Tree	→	Split into large sets
K Neighbors	→	Closest Data Point
Ensemble	→	Combine 2+ models
Support Vector Machine (SVM)	→	Line with boundary

Research Gap





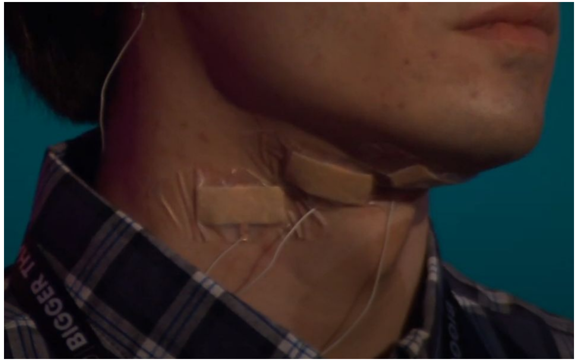
- Machine Learning is a black box
- Cannot identify optimal ML model for a given problem without testing out various models

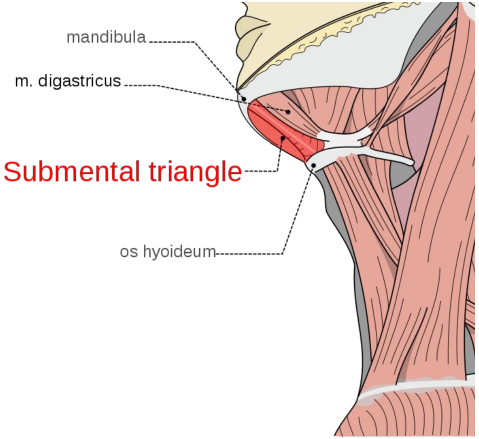
Research Question & Engineering Goal

Research Question:

Which type of Machine Learning Algorithm (CNN or Pattern Recognition) is most accurate at classifying surface ElectroMyoGraph (sEMG) signals from the submental triangle (area under the chin) to develop a Non-Invasive Silent Speech Interface?

Engineering Goal: To Develop a Non-Invasive Silent Speech Interface that can both collect and classify sEMG signals from the submental triangle (area under the chin) with greater than 80% accuracy.





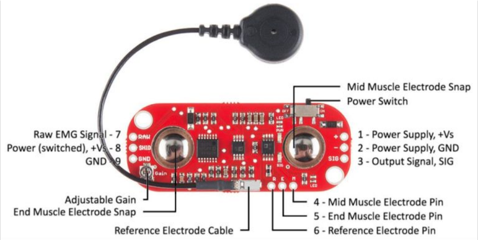
Assumptions and Value

Assumptions

1. It is possible to obtain accurate EMG data from low-cost muscle sensors such as the myoware
2. The Arduino would have a high enough sample rate to generate detailed data for the machine Learning models

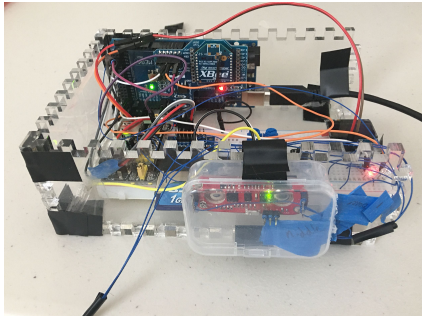
Value

1. Real time speech
2. High Accuracy
3. Helps even those with complete paralysis



EMG Accuracy

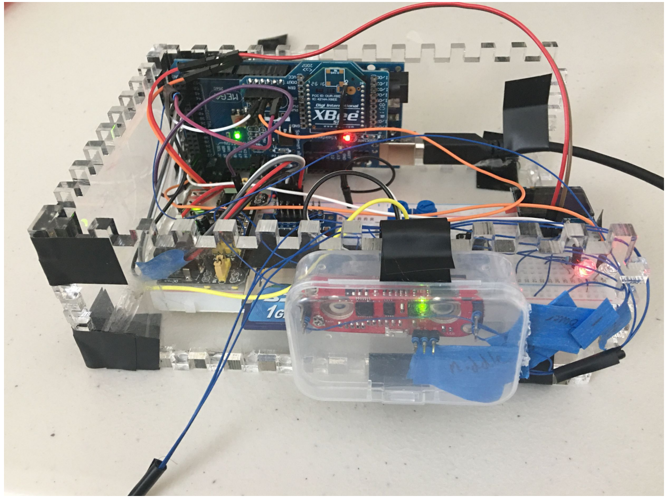
High Sample Rate

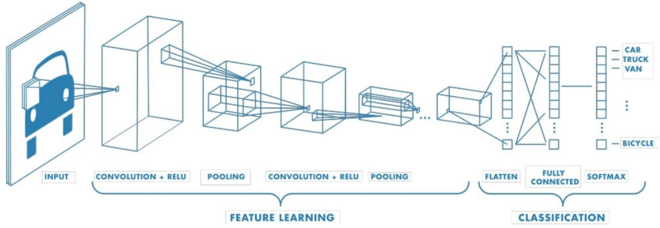


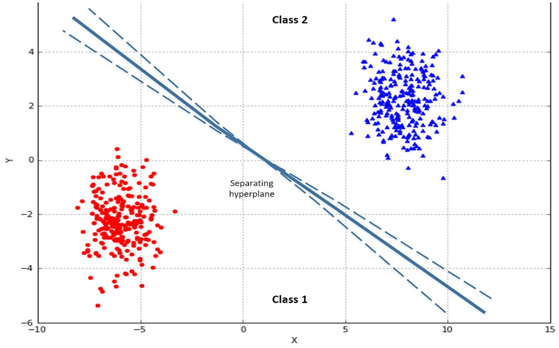
Device Speed & Accuracy

Methods

Engineering	Quantitative True Experimental
Create SSI Device	Evaluate Best ML Model

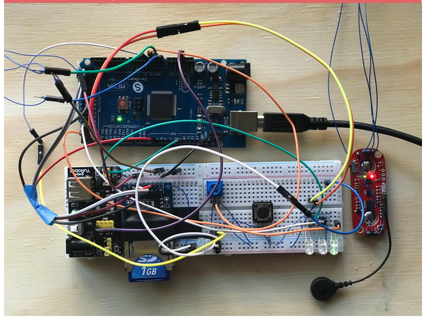




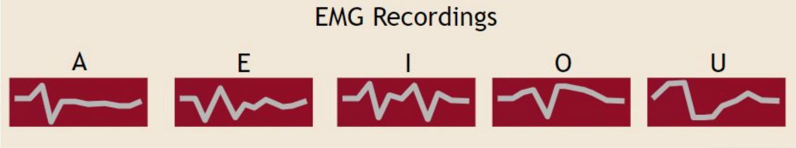


Methodology

Create Electromyograph using the Myoware muscle sensor

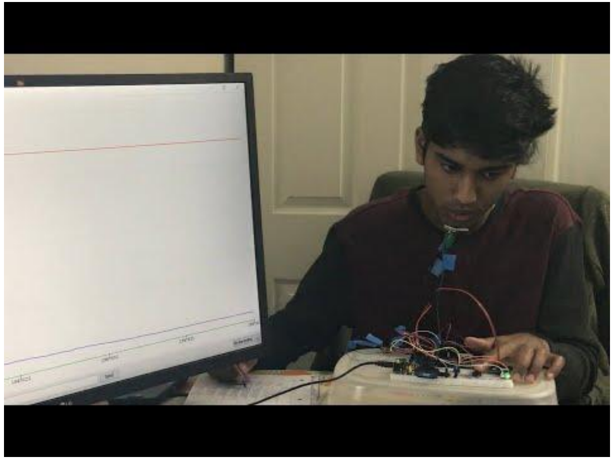


Generate EMG recordings and save it on the laptop/cloud

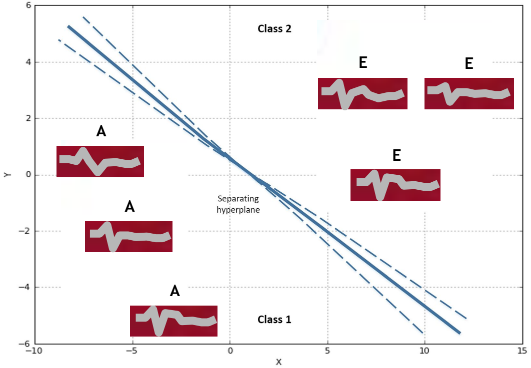


Data generated:
Signals- values ranging from 0-1023 (EMG signal)
Evaluation- Accuracy Percentage & F-scores

Methods:
Engineering - to create device
Quantitative True Experimental - determine effectiveness of SSI

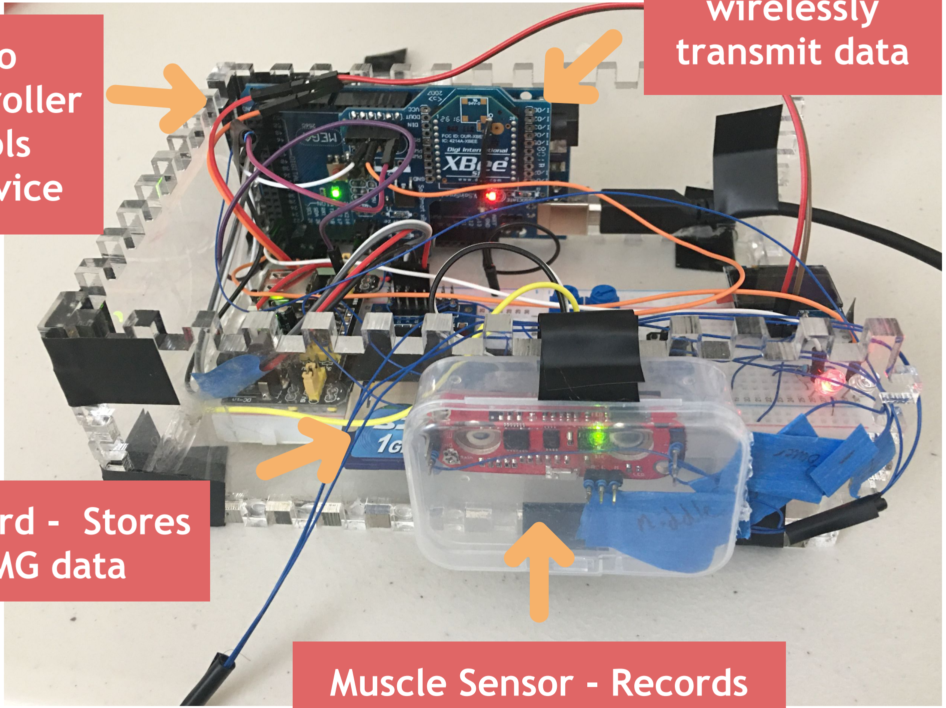


Develop and Deploy Multiple ML models on EMG data



Compare each ML Model accuracy values to identify best EMG Classifies

SpeakUp Hardware



Arduino Microcontroller - controls whole device

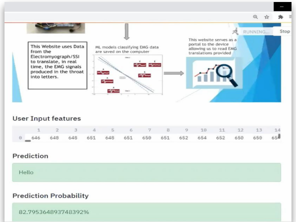
Radio to wirelessly transmit data

SD Card - Stores EMG data

Muscle Sensor - Records EMG Signals

EMG Signals Picked Up by Device and wirelessly sent to Computer

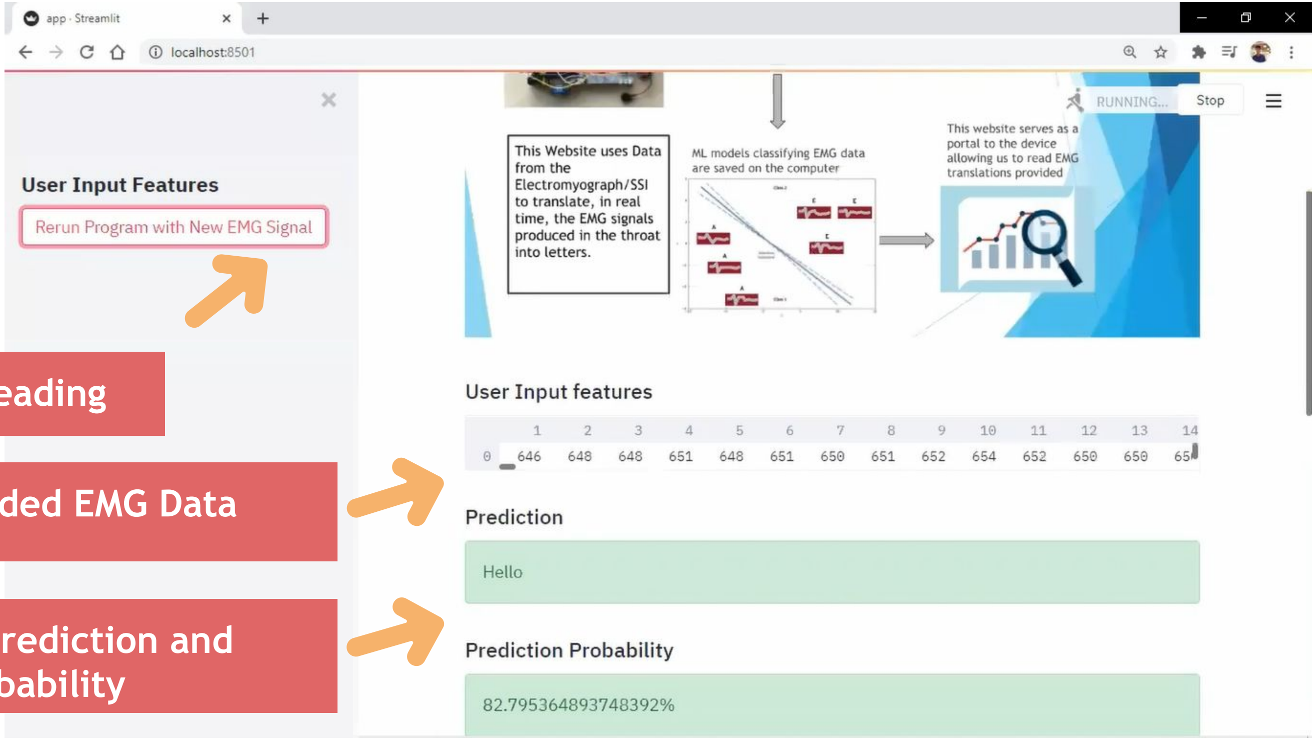
Computer Automatically Translates EMG Signal



Translation is displayed on Python Based Web Application

Web Interface

Acts as Portal to wirelessly control Device & See Speech Prediction



Button takes another reading

Displays Recorded EMG Data

Displays Prediction and Probability

SSI Device Demonstration

Device Capabilities

- Trained to identify letters
- With more data device could translate words/sentences

Video Shows Demonstration of Web Interface and SSI Device Classifying Words and Letters

Results

- Best Performing Model: SVM - 80.10% accuracy
- Device Accuracy: 80.10%

Engineering Goal Met > 80%

SVM model best classifies sEMG signals in an SSI

CLASSIFICATION ACCURACY AND F-SCORES

MODEL	CLASSIFICATION ACCURACY	F1 SCORES
CONVOLUTIONAL NEURAL NETWORK (CNN) - GOOGLENET	54.90%	0.60
SUPPORT VECTOR MACHINE (SVM) - GAUSSIAN	80.10%	0.81
ENSEMBLE - BAGGED TREES	74.60%	0.73
K-NEAREST NEIGHBORS (KNN) - WEIGHTED	66.70%	0.70
TREE - MEDIUM	59.80%	0.65
NAIVE BAYES - KERNEL	59.30%	0.65
QUADRATIC DISCRIMINANT	55.50%	0.54
LINEAR DISCRIMINANT	49.50%	0.48

Results

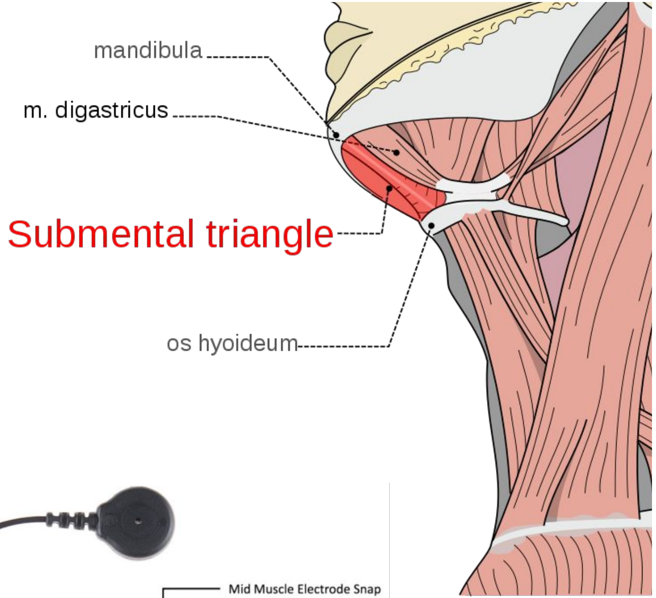
CNN Confusion Matrix

True Class \ Predicted Class	A	B	E	I	O	U
A	20.6%		14.7%	35.3%	14.7%	14.7%
B		100.0%				
E	17.6%		32.4%	23.5%	11.8%	14.7%
I	11.8%		11.8%	38.2%	23.5%	14.7%
O	2.9%		5.9%	14.7%	58.8%	17.6%
U					17.6%	82.4%

SVM Pattern Recognition

True Class \ Predicted Class	A	B	E	I	O	U
A	64.7%		8.8%	14.7%	8.8%	2.9%
B		100.0%				
E	20.6%		58.8%	11.8%	8.8%	
I	8.8%		17.6%	64.7%	8.8%	
O	8.8%		2.9%	14.7%	67.6%	5.9%
U	2.9%				8.8%	88.2%

Limitations / Enhancements

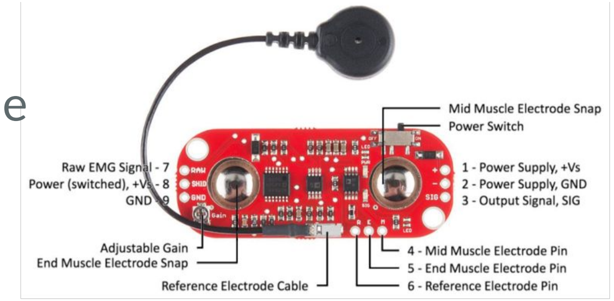


Current Limitations:

- Tested only on healthy individuals
- Records signals from one muscle group
- Accuracy is roughly 80%

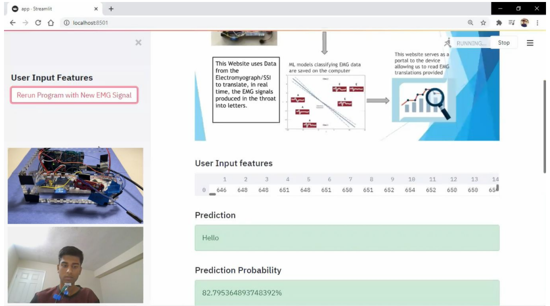
Future Enhancements:

- Test on ALS/paralyzed patients
- Record signals from all speech muscle groups
- Use multiple muscle sensors



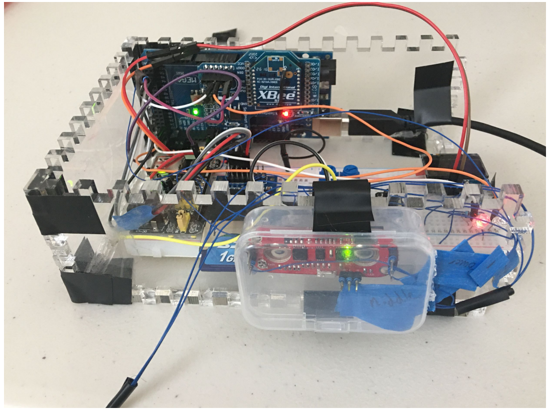
Conclusion

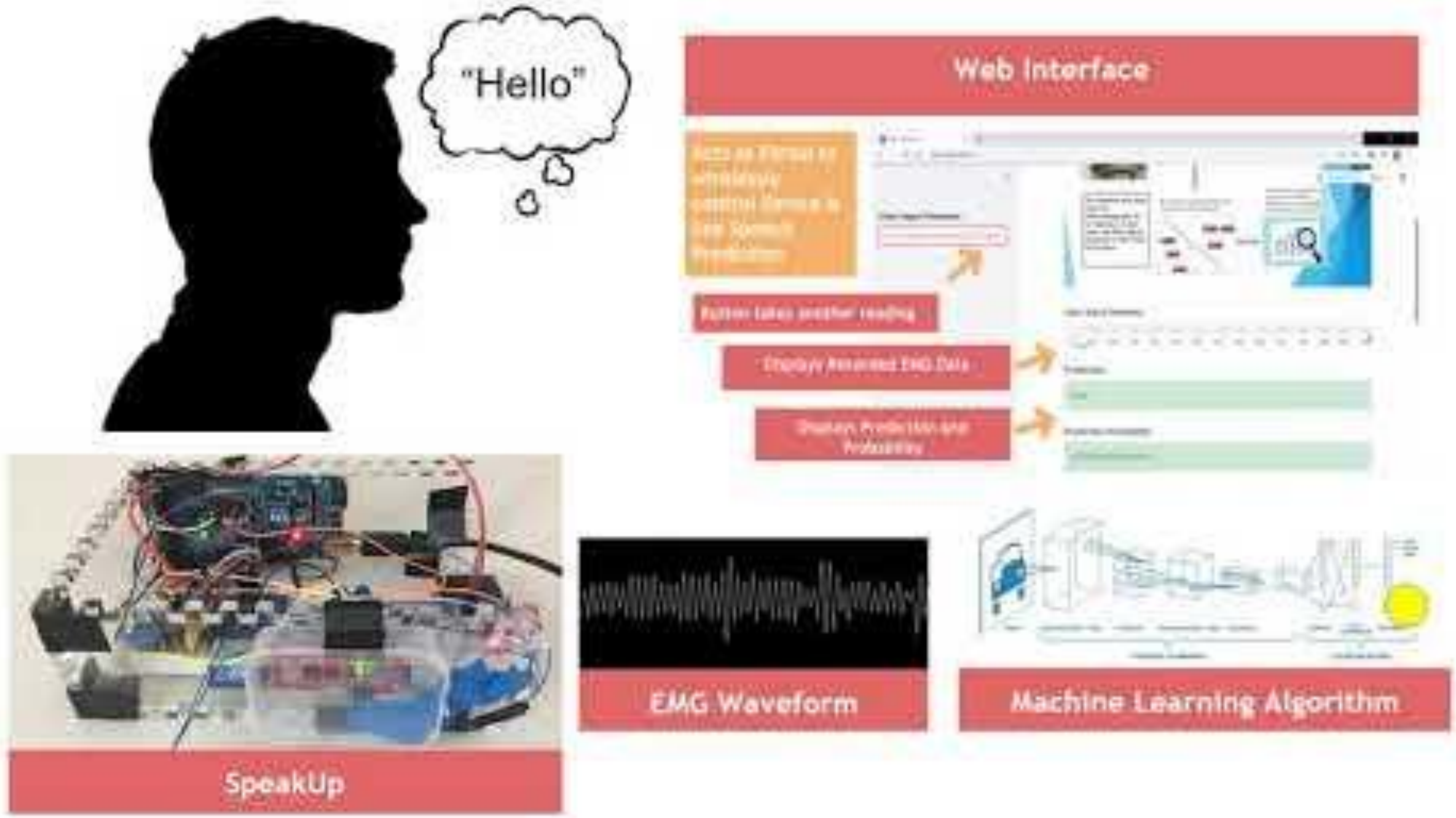
- An SSI was created to help the paralyzed communicate silently
- This study analyzed various Machine Learning Models to determine which ML model best classified EMG data.
- This device has an 80.10% accuracy and has the potential to revolutionize speech and communication aids



SVM model best classifies sEMG signals in an SSI

Engineering Goal Met > 80%




"Hello"
Web Interface
SpeakUp
EMG Waveform
Machine Learning Algorithm

Bibliography I

Alaskar, H. (2018, December). Convolutional Neural Network Application in Biomedical Signals. Retrieved November 24, 2020, from https://pdfs.semanticscholar.org/fd36/d9c6abf405af2c12d1596162421d3f503591.pdf

Arslan, Y. Z. (2006, July). Observations on the Characteristics of EMG Signals Recorded at Different Depths. Retrieved November 24, 2020, from https://www.researchgate.net/figure/Penetration-depth-of-needle-electrode_fig1_241204405

Ayodele, T. (2010, February). Types of Machine Learning Algorithms. Retrieved November 24, 2020, from https://www.researchgate.net/publication/221907660_Types_of_Machine_Learning_Algorithms

Beukelman, D., & Garrett, K. (1988, July 12). Augmentative and alternative communication for adults with acquired severe communication disorders. Retrieved November 24, 2020, from https://www.tandfonline.com/doi/abs/10.1080/07434618812331274687

Bishop, C., & Nasrabadi, N. (1970, January 01). [PDF] Pattern Recognition and Machine Learning: Semantic Scholar. Retrieved November 25, 2020, from https://www.semanticscholar.org/paper/Pattern-Recognition-and-Machine-Learning-Bishop-Nasrabadi/668b1277fbece28c4841eeab1c97e4ebd0079700

Bonaccorso, G. (2017, July 24). Machine Learning Algorithm. Retrieved November 25, 2020, from https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781785889622

Brown, S. (2000, October 01). Swallowing and Speaking Challenges for the MS Patient. Retrieved October 25, 2020, from https://meridian.allenpress.com/ijmsc/article/2/3/4/33107/Swallowing-and-Speaking-Challenges-for-the-MS

Char, D., Chung, T., Mckee, A., & Pai, A. (2018, June). The Human Keyboard. Retrieved August 27, 2020, from https://scholarcommons.scu.edu/cgi/viewcontent.cgi?article=1043&context=idp_senior

Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J., & Brumberg, J. (2011, August 20). Silent Speech Interfaces. Retrieved August 25, 2020, from https://hal.archives-ouvertes.fr/hal-00616227

Bibliography II

Kapur, A., Kapur, S., & Maes, P. (2018, March 5). AlterEgo: A Personalized Wearable Silent Speech Interface. Retrieved October 25, 2020, from https://www.media.mit.edu/publications/alterego-IUI/

Kapur, A., Sarawgi, U., Wadkins, E., & Wu, M. (2019). Non-Invasive Silent Speech Recognition in Multiple Sclerosis with Dysphonia. Retrieved August 24, 2020, from https://ml4health.github.io/2019/pdf/49_ml4h_preprint.pdf

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017, May). ImageNet Classification with Deep Convolutional Neural Networks. Retrieved November 24, 2020, from https://dl.acm.org/doi/10.1145/3065386

Mohanty, S., Hughes, D., & Salathé, M. (2016, September 22). Using Deep Learning for Image-Based Plant Disease Detection. Retrieved November 25, 2020, from https://pubmed.ncbi.nlm.nih.gov/27713752/

Muhammad, U. (2018, August 18). Pre-trained VGGNet Architecture for Remote-Sensing Image Scene Classification. Retrieved November 24, 2020, from https://www.researchgate.net/publication/329315021_Pre-trained_VGGNet_Architecture_for_Remote-Sensing_Image_Scene_Classification

Samuel, A. L. (1959, July). Some Studies in Machine Learning Using the Game of Checkers. Retrieved November 25, 2020, from https://ieeexplore.ieee.org/document/5392560

Schultz, T., & Wand, M. (2009, December 4). Modeling Coarticulation in EMG-based Continuous Speech Recognition Retrieved August 24, 2020, from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.212.6271&rep=rep1&type=pdf

Tang, P., Wang, H., & Kwong, S. (2017, February 15). G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition. Retrieved November 25, 2020, from https://scholars.cityu.edu.hk/en/publications/gms2f(97b29fed-c9f3-4838-a817-c90a321c3fc7).html

Zia ur Rehman, M., Gilani, S. O., Waris, M. A., & Niazi, I. K. (2018, July). Stacked Sparse Autoencoders for EMG-Based Classification of Hand Motions: A Comparative Multi-Day Analyses between Surface and Intramuscular EMG. Retrieved November 24, 2020, from https://www.researchgate.net/publication/326292263_Stacked_Sparse_Autoencoders_for_EMG-Based_Classification_of_Hand_Motions_A_Comparative_Multi_Day_Analyses_between_Surface_and_Intramuscular_EMG