MOVING TOWARDS REAL-TIME IMAGINED LANGUAGE CLASSIFICATION

Joseph Zonghi

Computer Engineering M.S. RIT

情報工学修士金沢工業大学

Dr. Cory Merkel

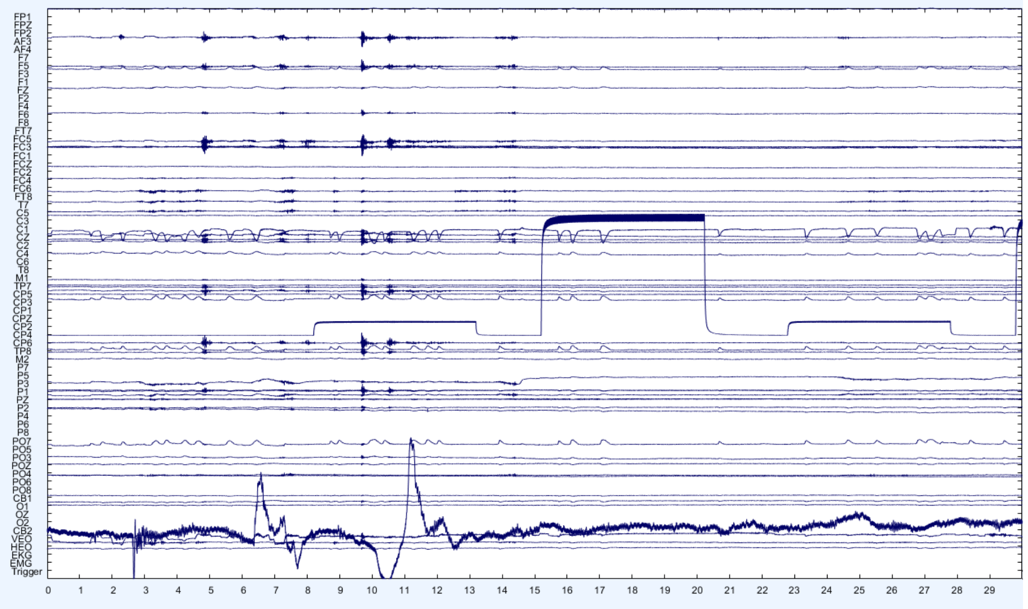
Dr. Minoru Nakazawa

OUTLINE

- Introduction & Goals
- Public Data Set Initial Findings
- Quantization and FPGA
- English/Japanese Dataset Creation
- English/Japanese Results
- Conclusion & Future Work

WHAT IS EEG?

- Electroencephalography
  - Recordings of the electrical activity at the scalp produced by the brain’s normal functions
  - We generate electrical signals from our brains 24/7
- Are these signals useful?
  - Seizure predictions/recordings
  - Sleep studies
  - Language prediction?
- Limitations
  - Signal is very weak at small distances
  - Need special devices for recording
  - Very noisy



SIMILAR WORK

Word classification:

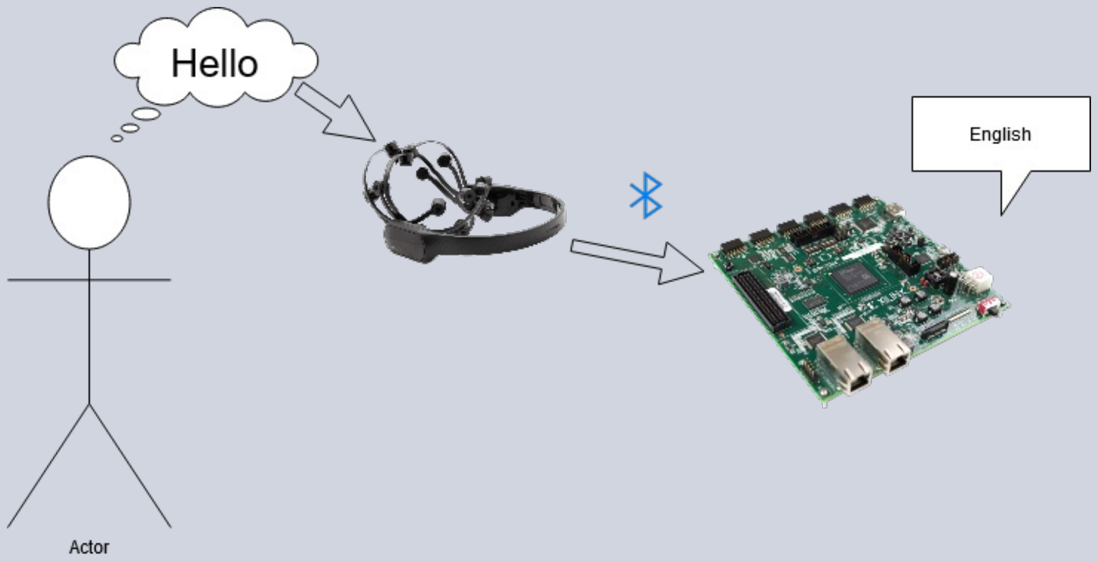
- Generally low success (difficult to get above random guessing) for multi-class
- [1] Torres-Garcia et al.: Support Vector Machine for 5 classes = **20-35% accuracy**
  - Random Forests = **40% accuracy**
- [2] Zhao et al.: Deep Belief Network for binary classification of sounds = **90% accuracy**
  - Publicly available dataset: Kara One

Language Classification:

- [3] Balaji et al.: Artificial Neural Network for yes/no classification = **92% accuracy**
- Not much else… what about whole sentences?

GOALS

- Brainwave Language Prediction
  - Differentiate between imagined English and Japanese
  - Assist with anarthria and dysarthria
  - Assist in multilingual learning environments
- Real-time using Neural Network
  - Preprocess the incoming Bluetooth data
  - Calculate output over a given time window using a neural network



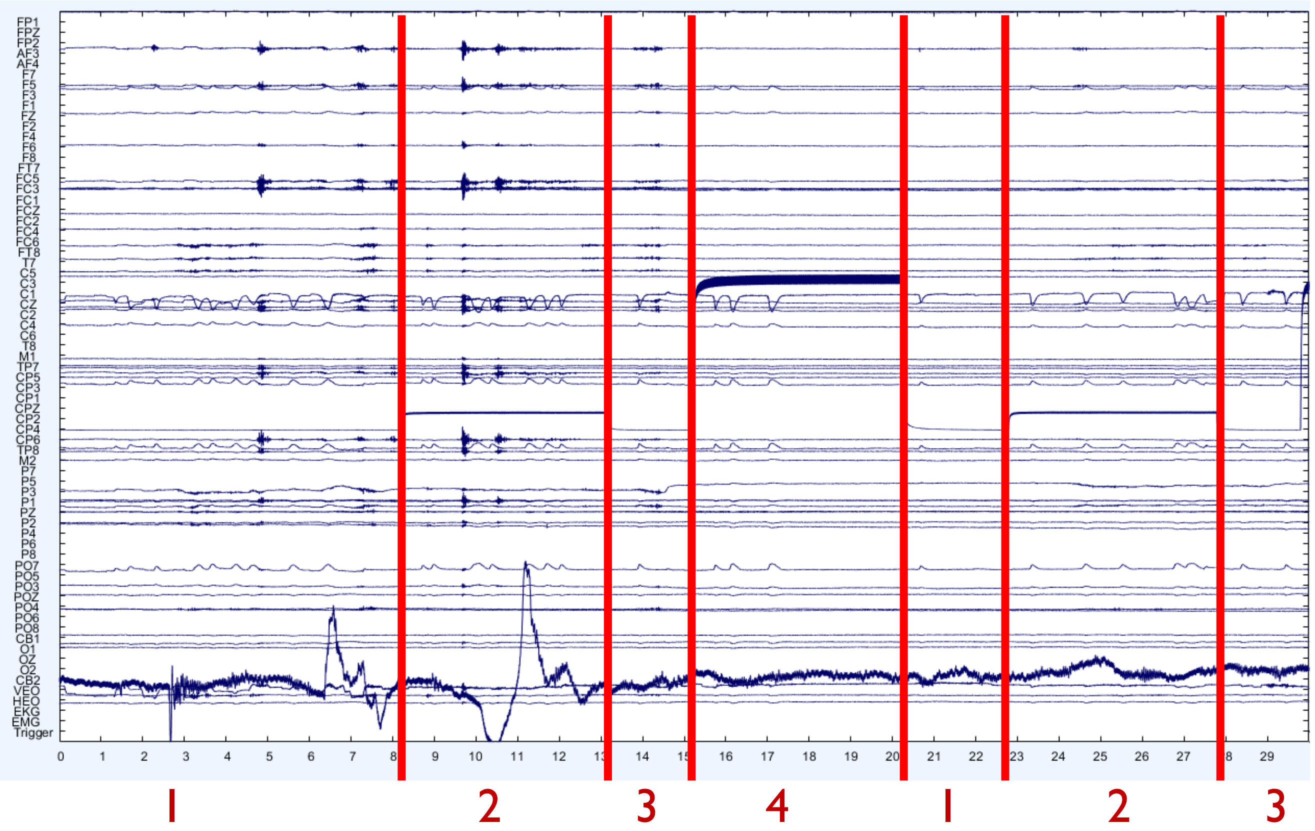
OUTLINE

- Introduction & Goals
- Public Data Set Initial Findings
- Quantization and FPGA
- English/Japanese Dataset Creation
- English/Japanese Results
- Conclusion & Future Work

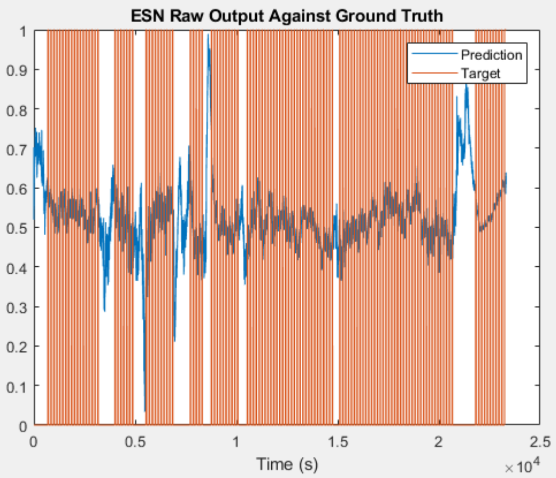
KARA ONE DATASET

- Provided by Zhao et al. [2]
  - Tried to classify presence of sounds in words: nasal word, vowel-only word, etc.
- Included 4 states per word
  1. Resting
  2. Stimuli
  3. Preparing
  4. Speaking

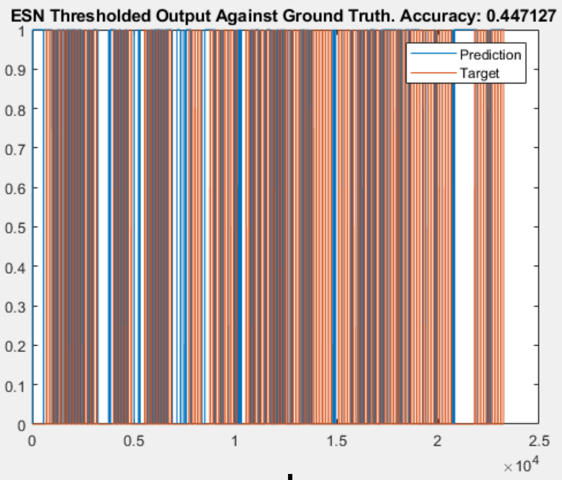
Classify?



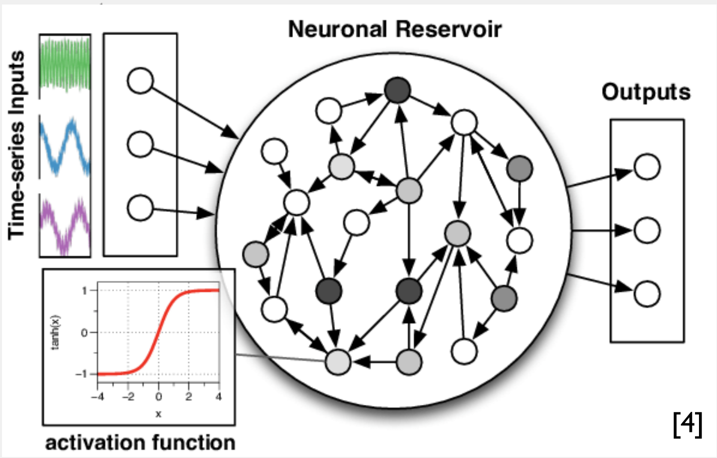
PRELIMINARY TESTING



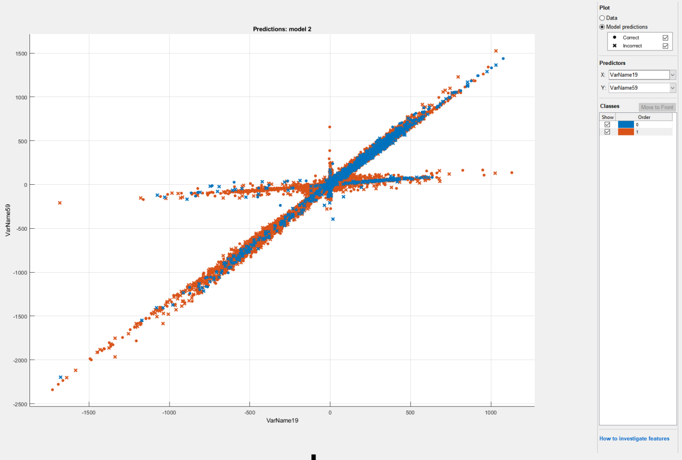
a.



b.



c.



d.

- Echo State Network
  - Unsatisfactory results
  - Difficult to differentiate between classes with high changing frequency (a.)
  - Difficulty finding reasonable threshold outputs (b.)
  - Many various hyperparameters tested (c.)
  - Raw data not inherently easily differentiable (d.)

SWITCH TO WINDOW-BASED

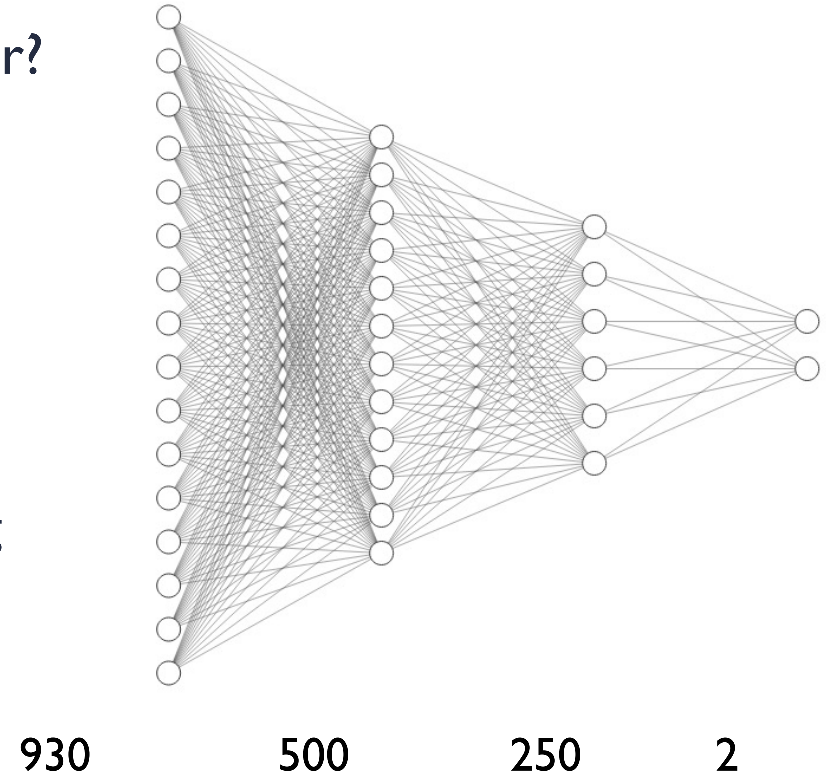
- Following results achieved by Zhao et al. [2]
- Preprocess data by extracting features over a window
  - Mean
  - Median
  - Min
  - Max
  - Standard Deviation
  - Variance
  - Kurtosis
  - Skewness
  - Etc.

1	Tree	Last change: Fine Tree	Accuracy: 70.7% — 2790/2790 features
2	SVM	Last change: Linear SVM	Accuracy: **94.3%** — 2790/2790 features
3	SVM	Last change: Quadratic SVM	Accuracy: 91.0% — 2790/2790 features
4	SVM	Last change: Cubic SVM	Accuracy: 87.4% — 2790/2790 features

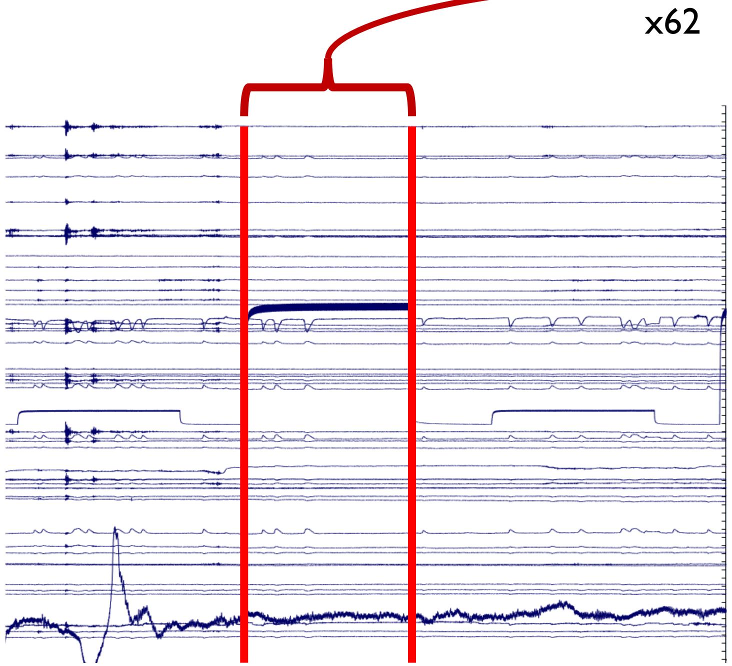
15 features x 62 channels = 930 input features

PRELIMINARY NEURAL NETWORK TESTING

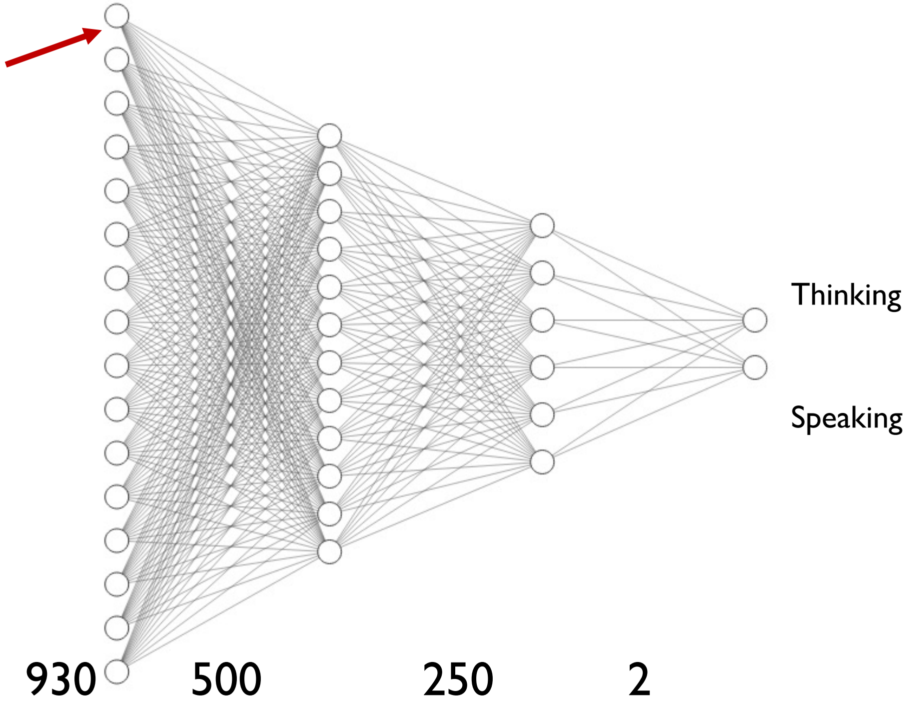
- Can the accuracy be increased further?
- NN Properties:
  - Normalizing input layer
  - Fully-connected internal layer(s)
  - ReLU activation layers
  - Softmax output activation layer
  - Classify between thinking and speaking states



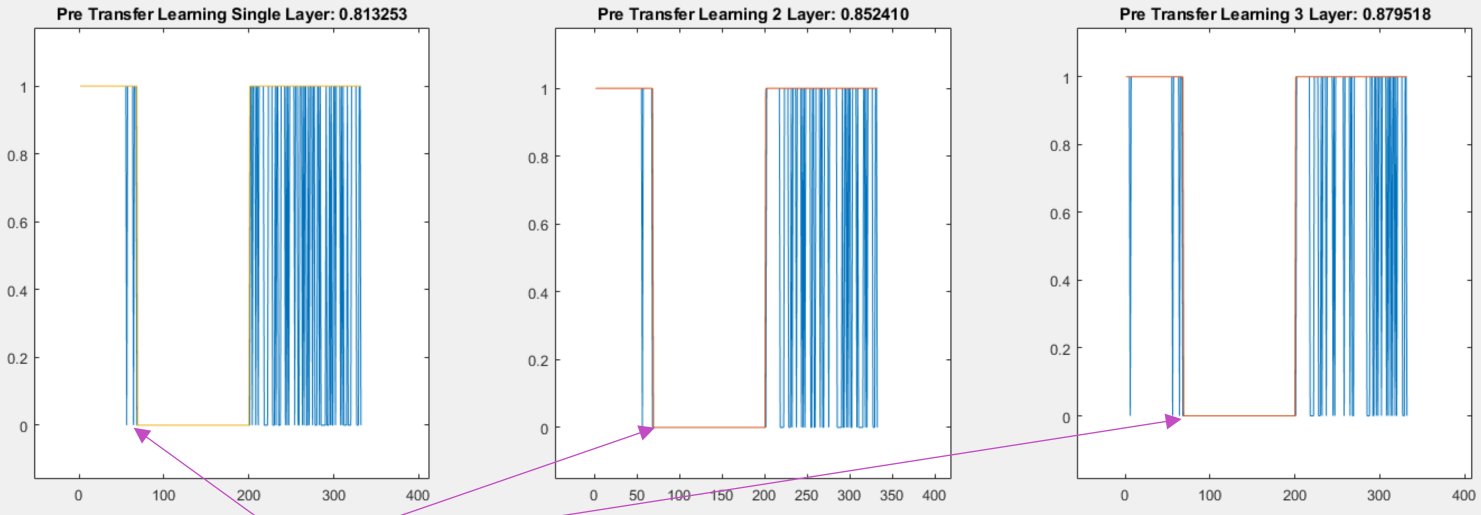
PRELIMINARY NEURAL NETWORK TESTING



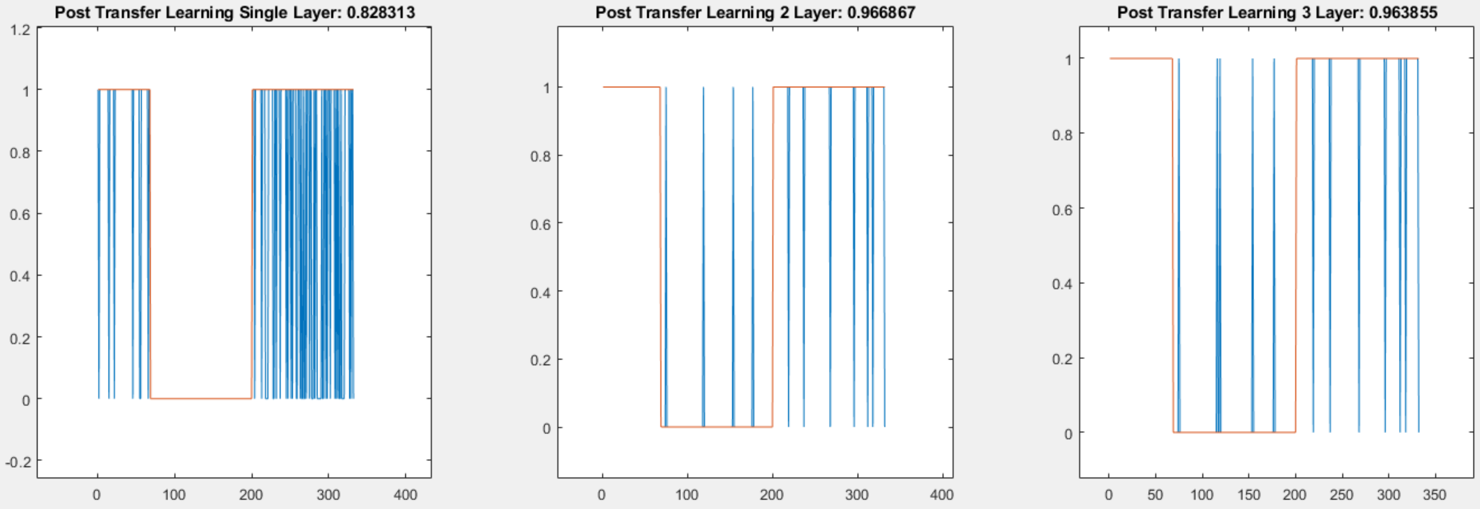
Mean
Absolute Mean
Maximum
Minimum
Absolute Maximum
Absolute Minimum
Sum of Max and Min
Delta of Max and Min
6th Mean
Standard Deviation
Variance
Kurtosis
Skewness
Sum
Median



PRELIMINARY NN TESTING



From 75 onward, a completely new person is tested upon



Here they are now trained upon with transfer learning

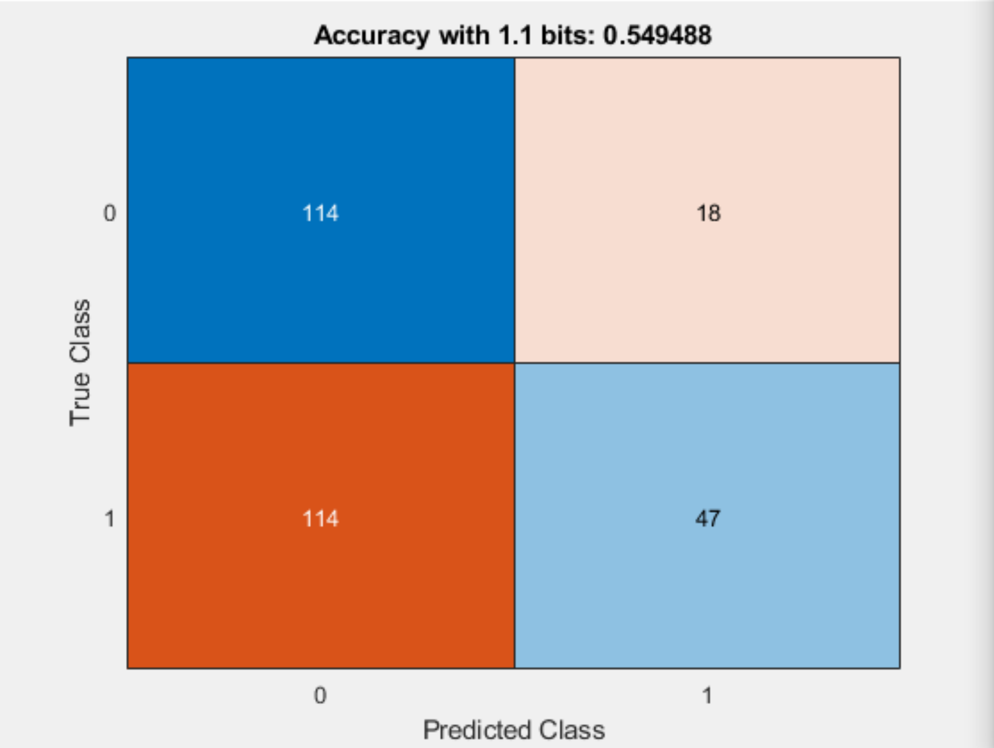
- 0 = Thinking
- 1 = Speaking
- Orange = ground truth
- Blue = network guess

What happens if we test on a brand-new person?

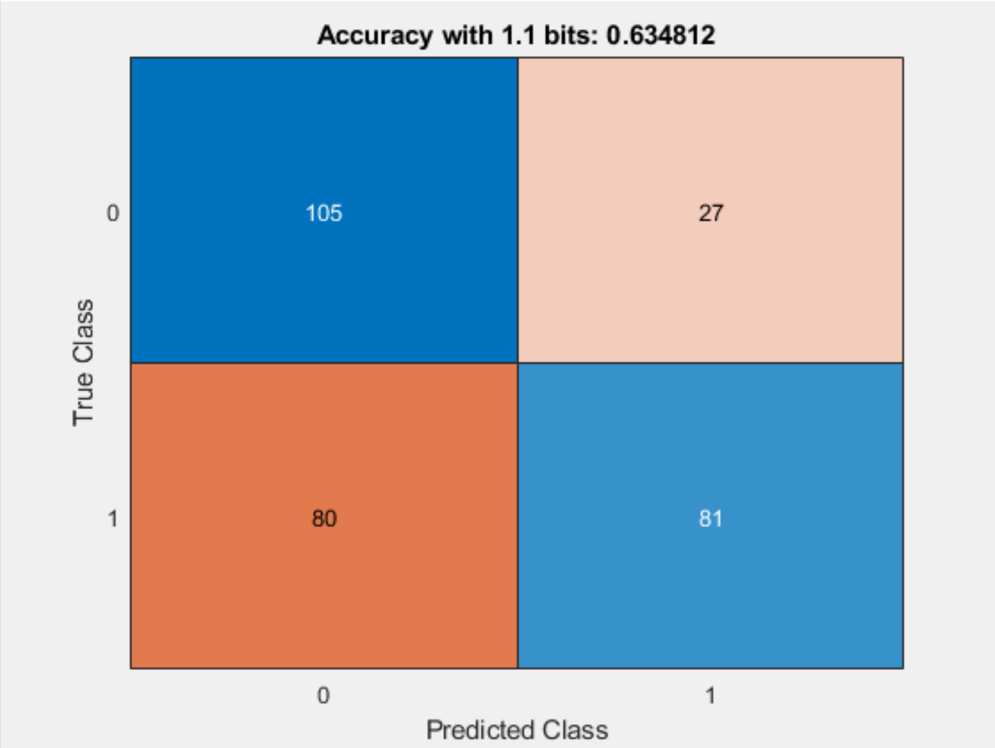
Important Takeaways:

- Stimuli heavily affects a person's EEG response
- Lack of stimuli is easy to train to
- EEG is heavily personalized

MORE TRAINING VS. TARGETED TRAINING



2000 training samples on various people



1500 training samples on less people overall but same people as test data

Key Takeaways:

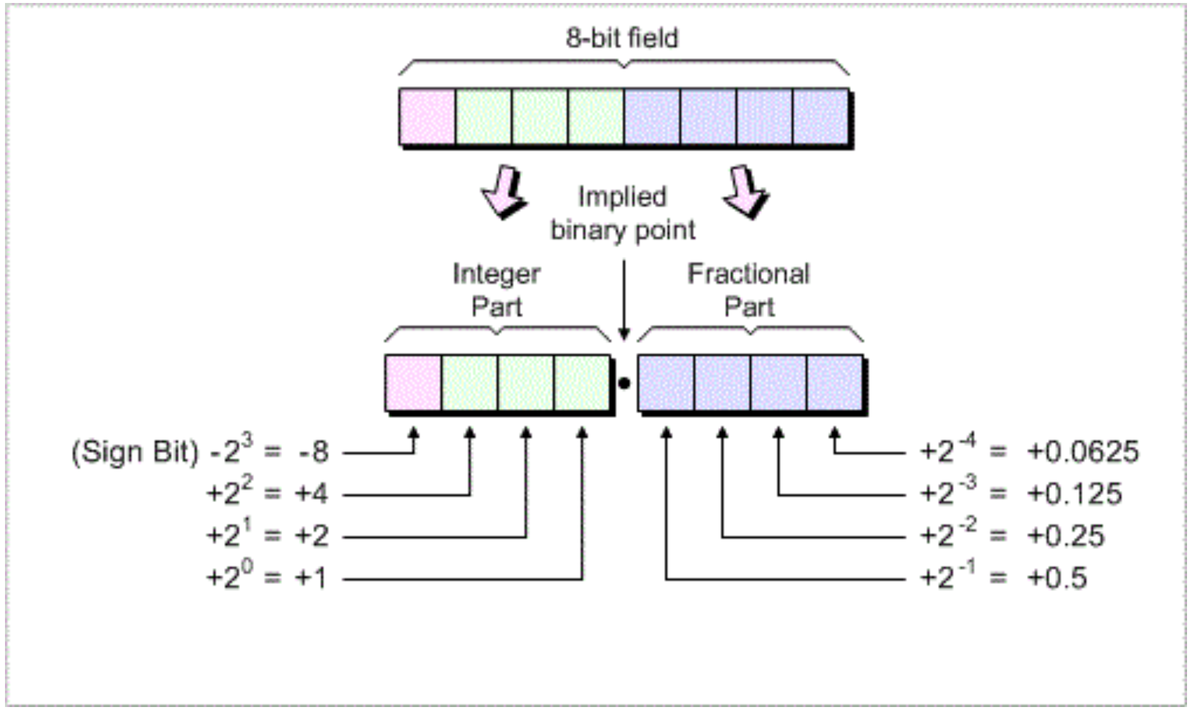
- EEG is heavily personalized!
- It might be better to have less training data but include the people you want to test on.

OUTLINE

- Introduction & Goals
- Public Data Set Initial Findings
- Quantization and FPGA
- English/Japanese Dataset Creation
- English/Japanese Results
- Conclusion & Future Work

QUANTIZATION

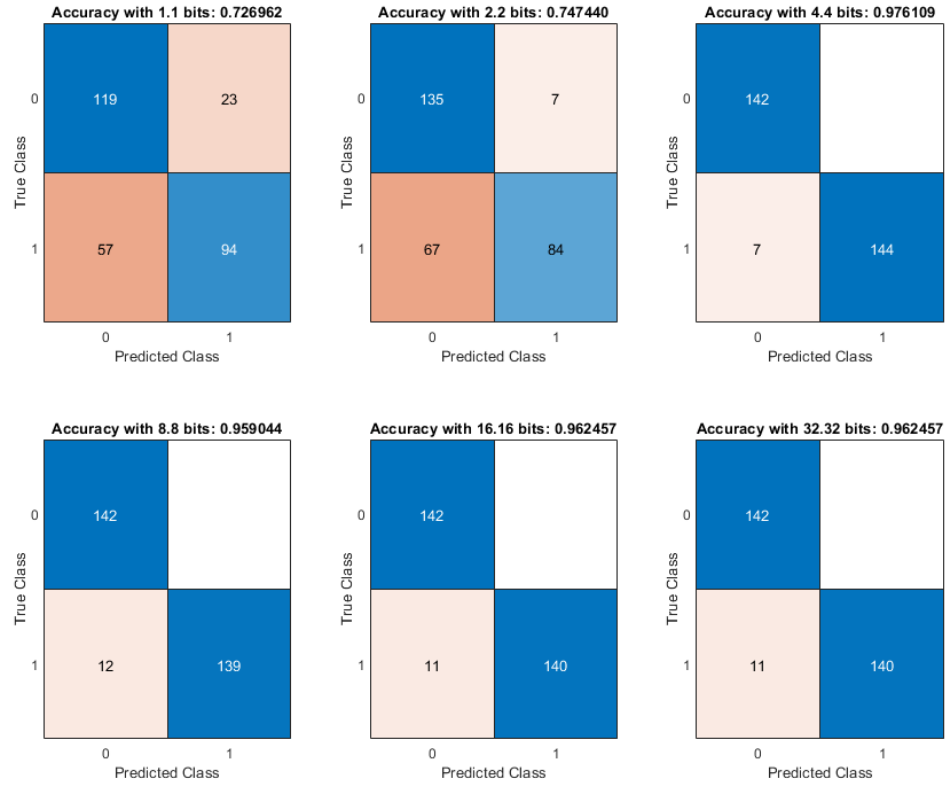
- Converting previously full precision (32 or 64 bit floating points for MATLAB) numbers to fixed point
- MATLAB usually uses 64 bits (double), but the DeepNetworkDesigner uses 32 bits for the weights
- Tensorflow has Quantization-Aware Training



Bits	Fixed Point
1	0.5
2	0.25
3	0.125
4	0.0625
5	0.03125
6	0.015625
7	0.0078125
8	0.00390625
9	0.001953125
10	0.000976563
11	0.000488281
12	0.000244141
13	0.00012207
14	6.10352E-05
15	3.05176E-05
16	1.52588E-05

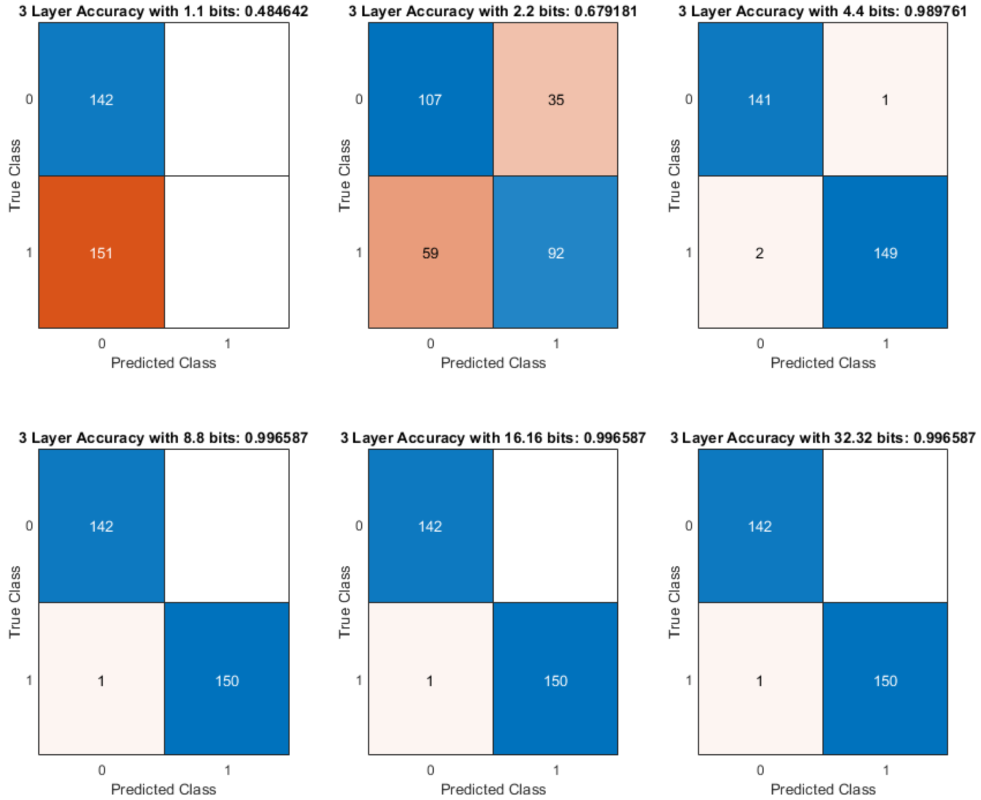
Quantization Results (Rounding after Training)

Single Layer Network



- Smaller networks are better with less resolution

Three Layer Network

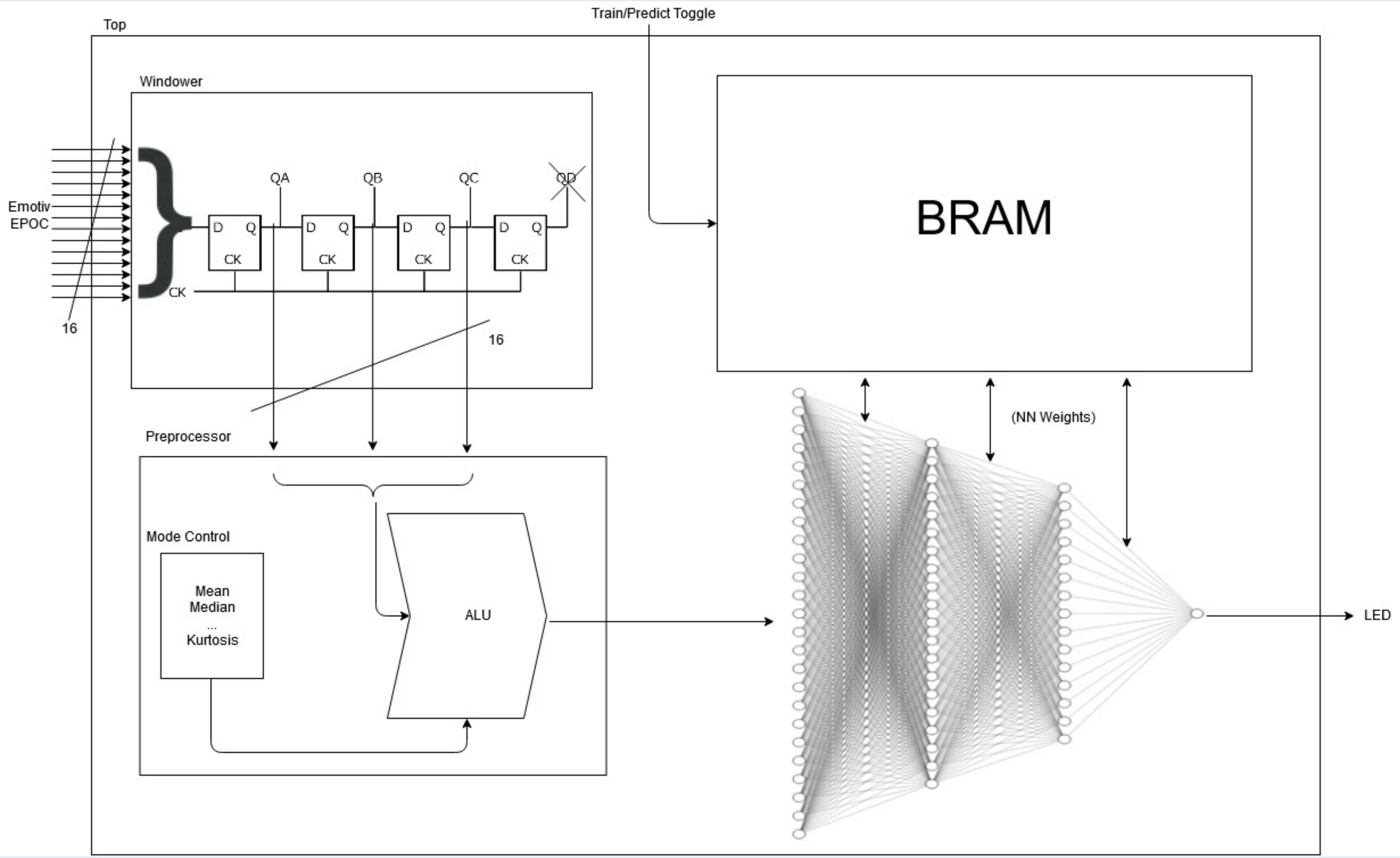


- Bigger networks propagate error more with less resolution
- Perform better at higher resolution

Quantization-Aware Training

- Three methodologies:
  - Base model training
    - Normal Tensorflow Training
  - 1-bit training
    - Train with awareness of 1-bit inputs
  - i-bit training
    - Train with awareness equal to the bit size of the final quantized weights
- Train with the respective methodologies, round afterwards
- Low precision networks use many weights, and high precision have few weights
- For low precision, use quantization aware training, but normal training is recommended for high precision

ORIGINAL FPGA DESIGN

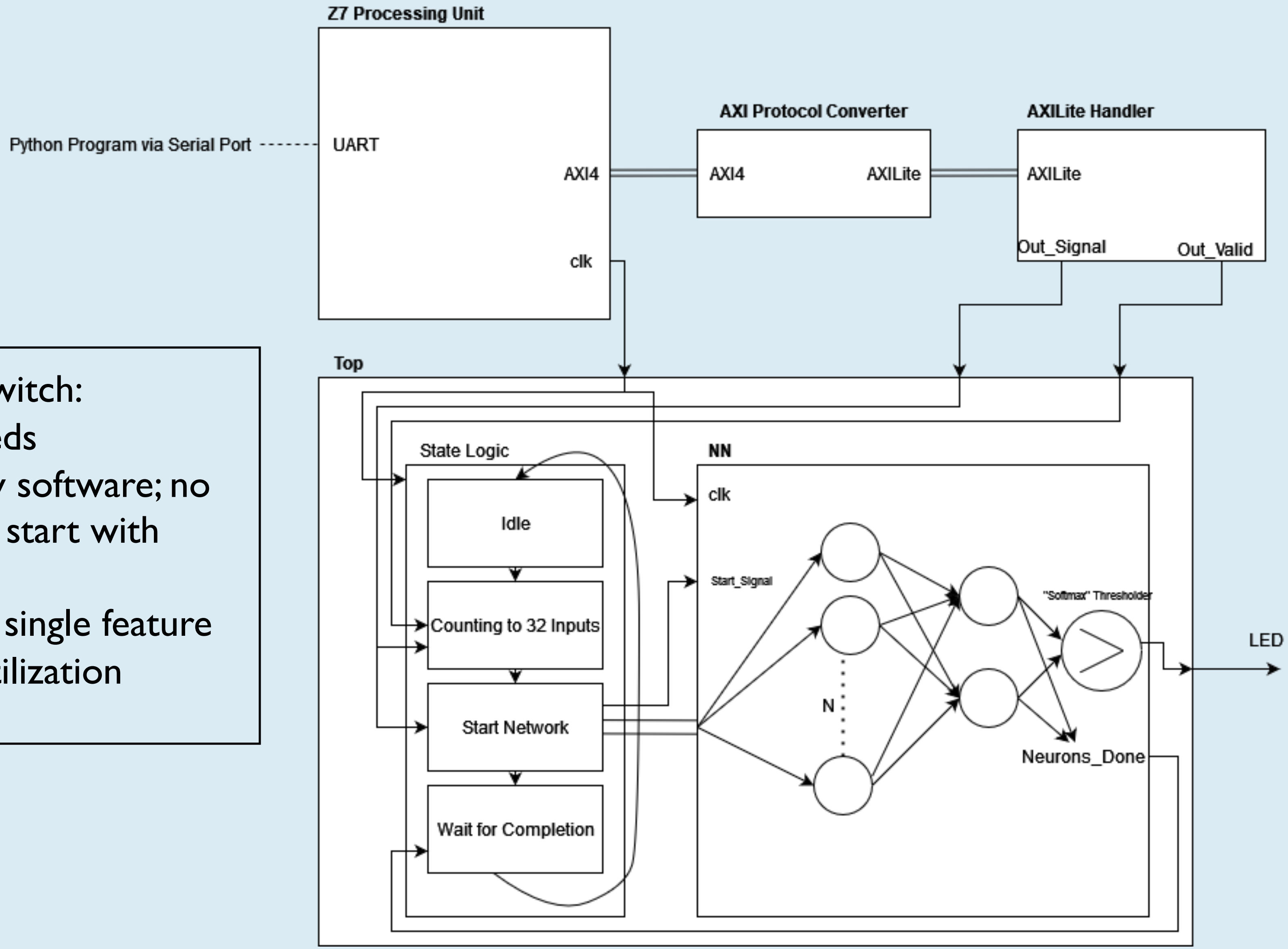


BEHAVIORAL NEURAL NETWORK IN VHDL

- Neuron State Machine:
  - Idle
    - Wait for start signal (from parent neural network component)
  - Inputs
    - Get input signal(s) as bus array
    - Set sum equal to bias
  - Multiplication
    - Mult <= weight(i) * input(i)
    - Go to sum if i != 0, else go to activation
  - Sum
    - Add mult result to current sum value
    - Decrement i

- Activation:
  - Send output to activation function component and done signal to 0 (active-low)
  - ReLU
    - If input > 0
      - Ouput <= input
    - Else
      - Output <= 0
  - Softmax
    - If input1 > input2
      - Output <= input1
    - Else
      - Output <= input2

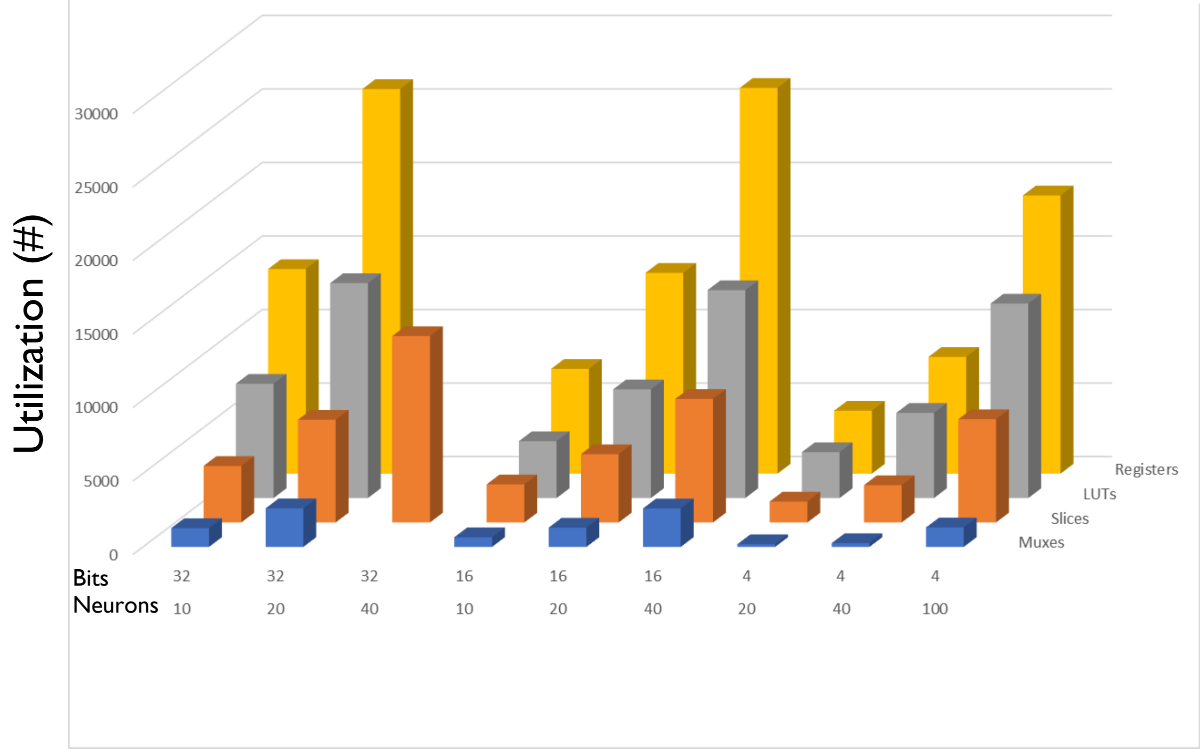
UPDATED FPGA DESIGN

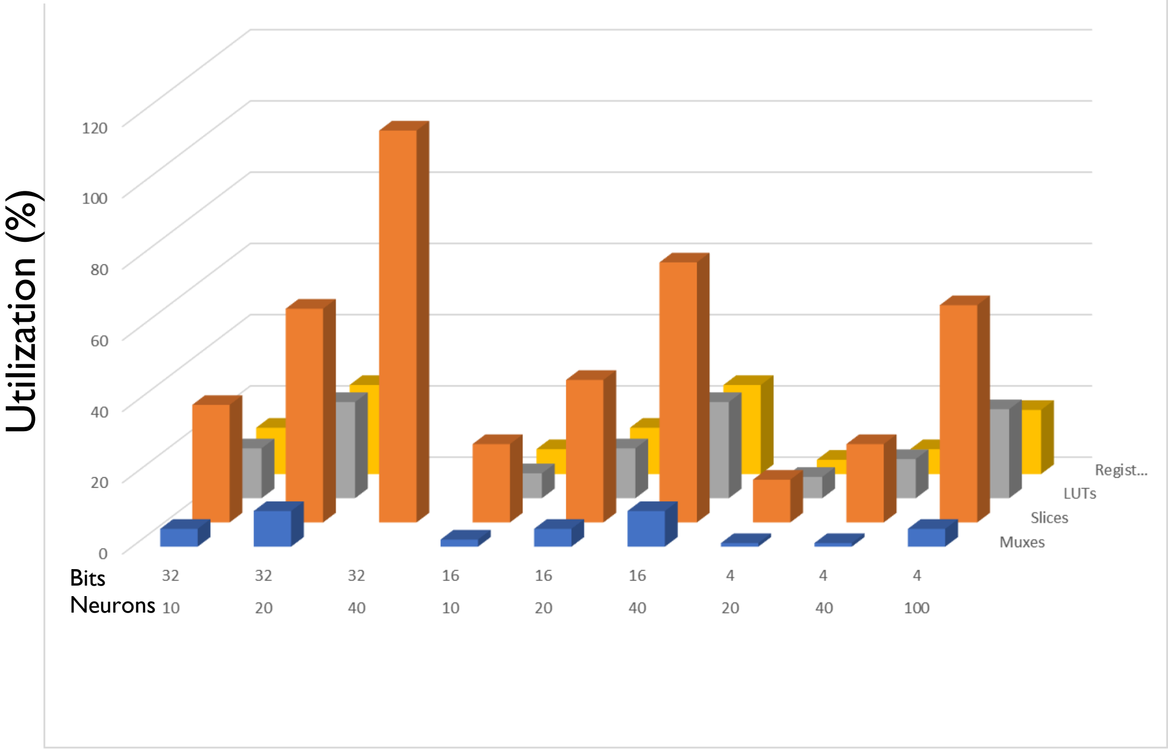


Reasons for switch:

- Emotiv needs proprietary software; no way to not start with software
- Only use a single feature
- Offloads utilization

UTILIZATION RESULTS





- Relative linear scaling with the total number of bits present (bits * neurons)
  - 20 neurons * 32 bits is about the same utilization as 40 neurons * 16 bits
- Pick combination based on goals
- Only small networks can fit!
- More weights leads to slower networks
- Less precision leads to less accuracy

OUTLINE

- Introduction & Goals
- Public Data Set Initial Findings
- Quantization and FPGA
- English/Japanese Dataset Creation
- English/Japanese Results
- Conclusion & Future Work

DATASET

- 5 subjects: 4 native Japanese, 1 native English
- Read English or Japanese sentence combinations displayed on screen
- 60 prompt combinations per person (3 sets of 20)
- Example prompt combination:
  - Today is very hot, but it seems like it will rain next week. + The supermarket sells bananas, but they don't have blueberries.
  - 今日はとても暑いけど、来週は雨が降りそう。+ スーパーはバナナを売っているけど、ブルーベリーがない。
- Random, unscripted imagined speech included as well

EMOTIV EPOC X VS. FLEX

Emotiv EPOC X

- 14 Channels
- 14-16 Bit Precision
- 128 or 256 Hz
- 5th order Sinc Filtering

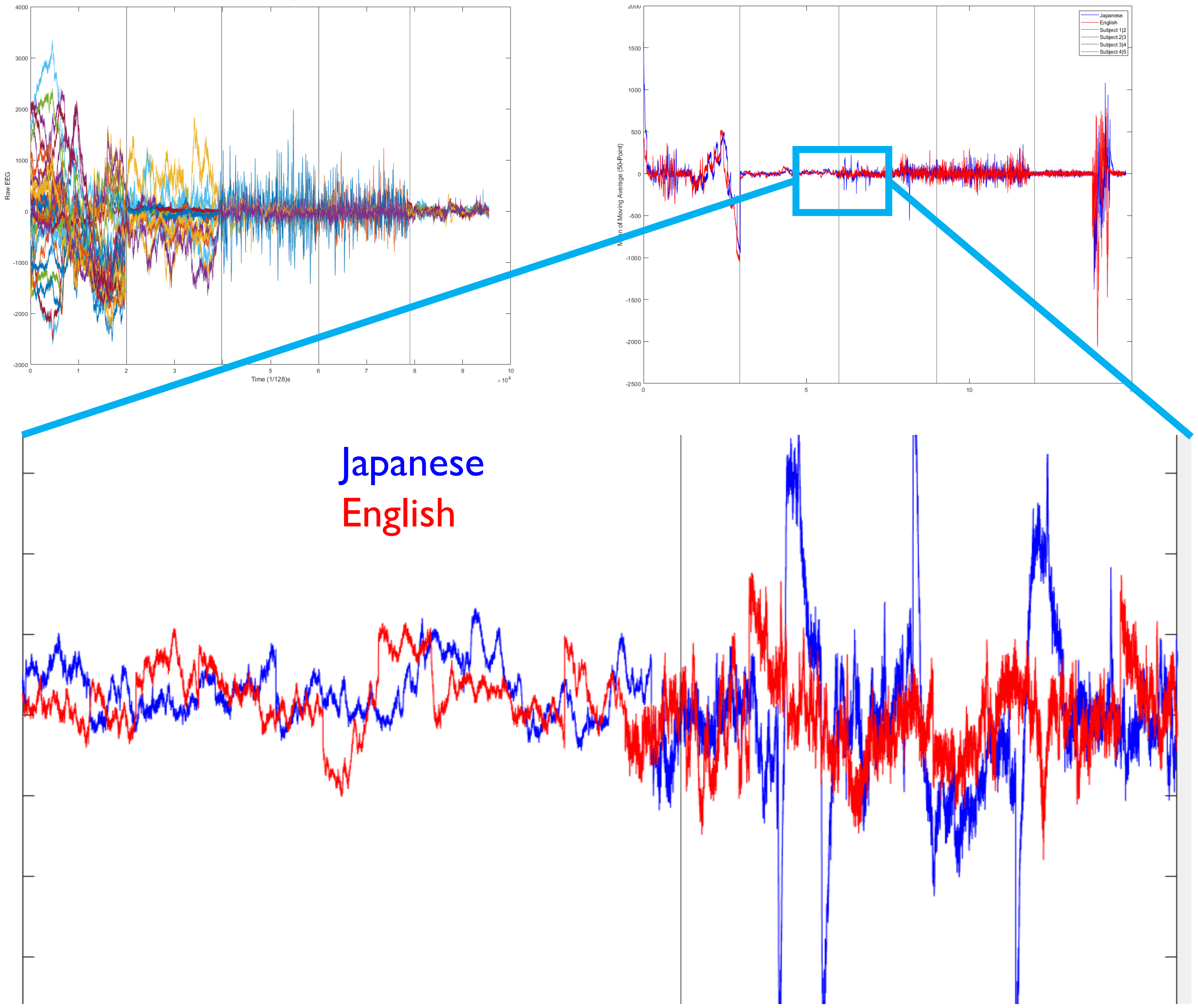
Emotiv EPOC Flex

- 32 Channels
- 14 Bit Precision
- 128 Hz
- 5th Order Sinc Filtering

VIEWING THE DATA

Random Speech

Prompt-Based Speech

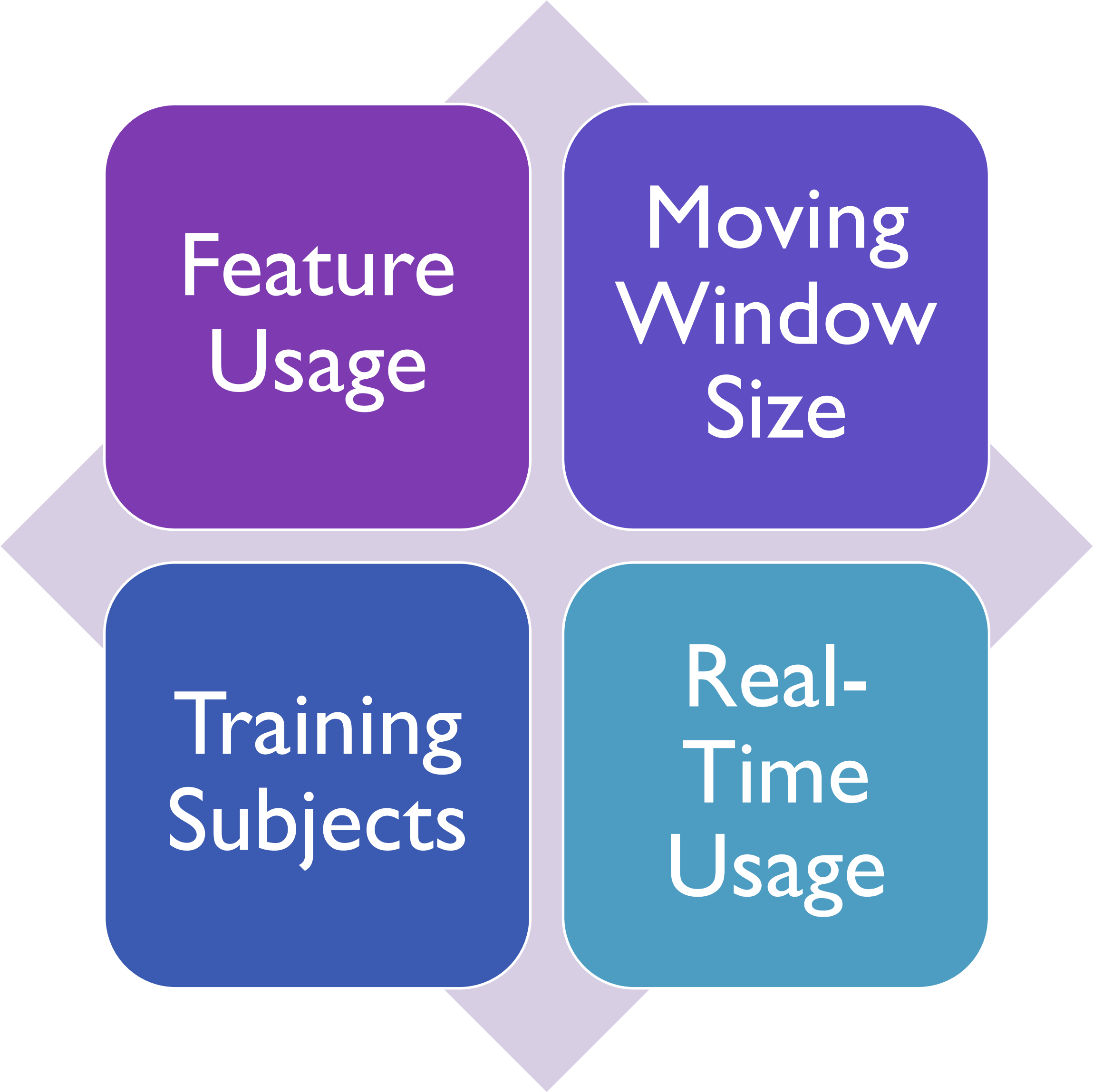


Very Noisy!

OUTLINE

- Introduction & Goals
- Public Data Set Initial Findings
- Quantization and FPGA
- English/Japanese Dataset Creation
- English/Japanese Results
- Conclusion & Future Work

ANALYZING THE DATA

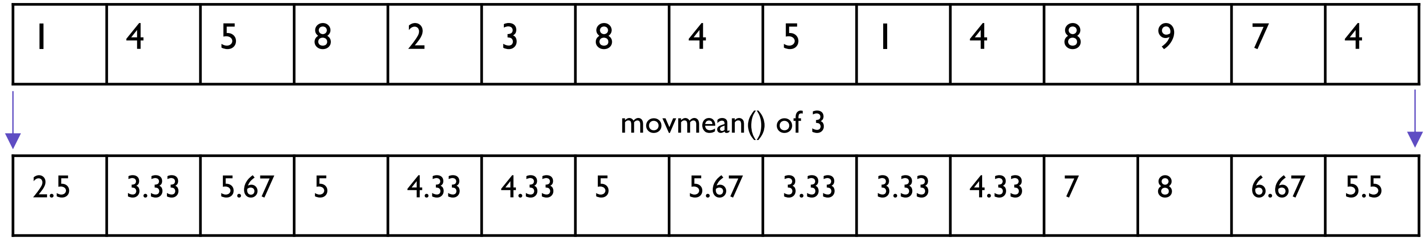


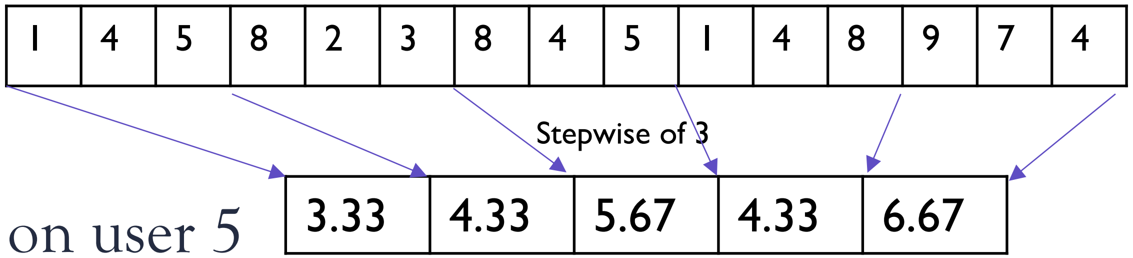
FEATURE SELECTION

- More features = less accuracy?
- Mean alone proves to be the most effective
- Raw EEG is also effective

	EPOC X	EPOC Flex
Mean Only	0.7137 ±0.039	0.9846 ±0.010
Above + Max + Min + Max/Min Related	0.6368 ±0.020	0.9538 ±0.018
Above + Standard Deviation + Variance	0.5897 ±0.022	0.9077 ±0.014
Skewness & Kurtosis	0.5214 ±0.027	0.3692 ±0.076
All 14	0.5940 ±0.019	0.8923 ±0.034
Raw EEG	0.5024 ±0.041	0.9940 ±0.003

MOVING WINDOW





- For training purposes, two methods were examined on user 5
  - MATLAB's movmean() function
    - Temporally close points are very similar, may lead to overfitting
    - Large amount of training data
  - Stepwise moving average
    - Each group of points are separated by the window size, so the resulting values are means of unique points
    - Less data, but it is more unique

Window Size	Moving Mean	Stepwise
10	0.9942 ± 0.005	0.8995 ± 0.026
20	0.9978 ± 0.002	0.8454 ± 0.030
50	0.9957 ± 0.003	0.6538 ± 0.059
100	0.9964 ± 0.002	0.6410 ± 0.124
200	0.9982 ± 0.001	0.4500 ± 0.265

TRAINING SUBJECTS

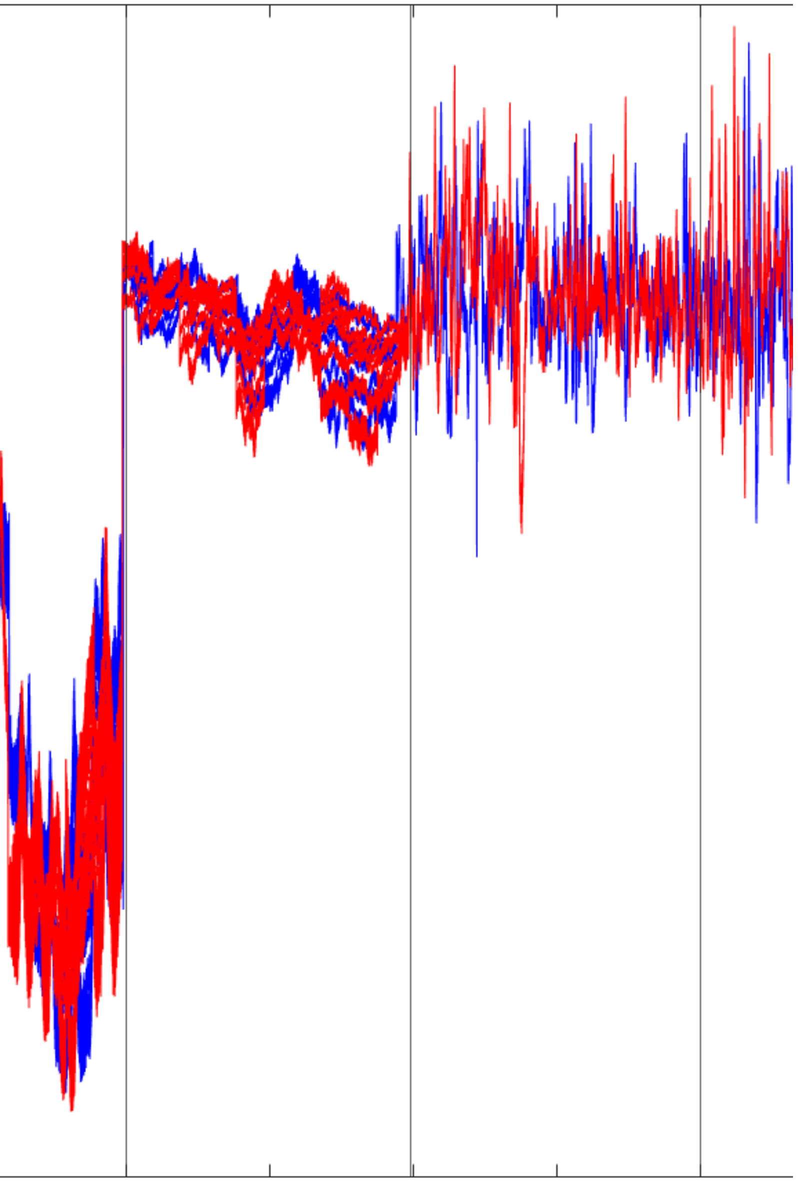
100 Hidden Neuron Network	Test Accuracy	Accuracy for a New Subject
Subjects 1-4	0.7842 $\pm$0.010	0.5111 $\pm$0.047
Subjects 1-3	0.7984 $\pm$0.011	0.5130 $\pm$0.032
Subjects 1-2	0.8570 $\pm$0.017	0.4983 $\pm$0.045
Subjects 2-3	0.8443 $\pm$0.013	0.5083 $\pm$0.032
Subject 2	0.8646 $\pm$0.013	0.5101 $\pm$0.039
Subject 2 (20 hidden neurons)	0.7467 $\pm$0.006	0.5264 $\pm$0.028
Subject 2 (1000 hidden neurons)	0.8919 $\pm$0.005	0.4969 $\pm$0.030

- Would different combinations of subjects as training data work well for testing on a brand-new person to the network?
- New people have such a large variance that even with heavy training regularization, the model can not adapt well.

TRANSFER LEARNING ATTEMPTS

Combination (Neurons)	Test Accuracy on Set	Accuracy for New Subject (% Included)
Subjects 1-5 (1000)	0.8204 $\pm$0.005	N/A
Subjects 1-5 (5000)	0.8141 $\pm$0.003	N/A
Subjects 1-4 (1000)	0.7834 $\pm$0.012	0.6073 $\pm$0.028 (50%)
Subjects 1-4 (1000)	0.7905 $\pm$0.006	0.4463 $\pm$0.032 (25%)
Subjects 1-4 (100)	0.7849 $\pm$0.014	0.5331 $\pm$0.021 (50%)
Subjects 1-4 (100)	0.8167 $\pm$0.008	0.4888 $\pm$0.044 (25%)
Subjects 1-4 (20)	0.6816 $\pm$0.011	0.5331 $\pm$0.051 (50%)
Subjects 1-4 (20)	0.6860 $\pm$0.011	0.4003 $\pm$0.045 (25%)
Subjects 1-3	0.8288 $\pm$0.005	0.4972 $\pm$0.039
Subject 3	0.8714 $\pm$0.010	0.5307 $\pm$0.048

- If new people are very difficult to adequately classify, how about including some of their data when retraining?
- Train first, and then train again using some of the target user's data in the train set
- Better, but they still have too much variance in the rest of their data.



"REAL-TIME" RESULTS

- Using Sets 1+2 for train, 3 for test
- Even with regularization and many different combinations of parameters, data taken at a different time period is too unique for the model to be able to adapt to currently.
- What about trends in the data instead...?

Users	Target	Test Accuracy	"Real-Time" Accuracy	Notes
1-5	1	0.9095 ±0.029	0.5540 ±0.034	Raw EEG
1-5	2	0.8817 ±0.036	0.5502 ±0.019	Raw EEG
1-5	3	0.8792 ±0.031	0.5347 ±0.013	Raw EEG
1-5	4	0.9063 ±0.030	0.5477 ±0.022	Raw EEG
1-5	5	0.9418 ±0.038	0.5049 ±0.027	Raw EEG
1-5	1	0.9136 ±0.024	0.4915 ±0.036	Moving Mean
1-5	2	0.9546 ±0.024	0.5071 ±0.038	Moving Mean
1-5	3	0.9008 ±0.026	0.4998 ±0.048	Moving Mean
1-5	4	0.9126 ±0.028	0.5447 ±0.017	Moving Mean
1-5	5	0.9465 ±0.021	0.4958 ±0.066	Moving Mean
1	1	0.9520 ±0.009	0.5118 ±0.034	Raw EEG
2	2	0.8279 ±0.020	0.5241 ±0.047	Raw EEG
3	3	0.8549 ±0.008	0.5256 ±0.033	Raw EEG
4	4	0.8516 ±0.009	0.5379 ±0.023	Raw EEG
5	5	0.9316 ±0.018	0.4615 ±0.038	Raw EEG
1	1	0.9411 ±0.019	0.4861 ±0.027	Moving Mean
2	2	0.9896 ±0.003	0.6042 ±0.025	Moving Mean
3	3	0.9893 ±0.006	0.5421 ±0.044	Moving Mean
4	4	0.8970 ±0.011	0.5122 ±0.033	Moving Mean
5	5	0.9980 ±0.002	0.4648 ±0.032	Moving Mean

OUTLINE

- Introduction & Goals
- Public Data Set Initial Findings
- Quantization and FPGA
- English/Japanese Dataset Creation
- English/Japanese Results
- Conclusion & Future Work

IN CONCLUSION

Yes, imagined language is differentiable

Post-hoc accuracies reach over 95%!

Real-time accuracies barely surpass 60%...

EEG is heavily temporally dependent and personalized… maybe try an Echo State Network or LSTM?

Emotiv devices need proprietary software – not easily compatible with an FPGA

Quantization-aware training can be useful at low precision (75% at 2 bits to 90% at 4 bits) , but maybe not at high precision (93% from 5 bits onward)

FUTURE WORK

1

Temporal network approach

- Echo State Network, Long Short-Term Memory Network, etc.

2

Increase of subjects for the dataset and/or increase of data per subject

3

EEG recording device reconsideration for usage with FPGA

ACKNOWLEDGEMENTS

- Advisors: Dr. Cory Merkel & Dr. Minoru Nakazawa
- Dr. Andres Kwasinki, dual MS program advisor
- Fujimura-san, Fukami-san, and Kugo-san from KIT study abroad
- Tokida-san, Watanabe-san, Adachi-san, and Shimizu-san from Nakazawa Lab
- Taga-san for the support
- My family in Natick

REFERENCES

- [1] A. Torres-Garcia, C. A. Reyes-Garcia, and L. Villasenor-Pineda, “Toward a silent speech interface based on unspoken speech,” in BIOSIGNALS 2012 – Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing, 02 2012.
- [2] Shunan Zhao and Frank Rudzicz (2015) Classifying phonological categories in imagined and articulated speech. *In Proceedings of ICASSP 2015*, Brisbane Australia.
- [3] A. Balaji, A. Haldar, K. Patil, T. S. Ruthvik, V. CA, M. Jartarkar, and V. Baths, “Eeg-based classification of bilingual unspoken speech using ann,” in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017, pp. 1022–1025.
- [4] Soh, Harold & Demiris, Yiannis. (2014). Spatio-Temporal Learning With the Online Finite and Infinite Echo-State Gaussian Processes. IEEE transactions on neural networks and learning systems. 26. 10.1109/TNNLS.2014.2316291.
- [5] G. Krishna, C. Tran, Y. Han, M. Carnahan, and A. H. Tewfik, “Speech synthesis using eeg,” in ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 1235–1238.
- [6] https://www.allaboutcircuits.com/technical-articles/fixed-point-representation-the-q-format-and-addition-examples/
- [7] https://emotiv.gitbook.io/epoc-x-user-manual/introduction/technical-specifications
- [8] https://emotiv.gitbook.io/epoc-flex-user-manual/epoc-flex/technical-spec