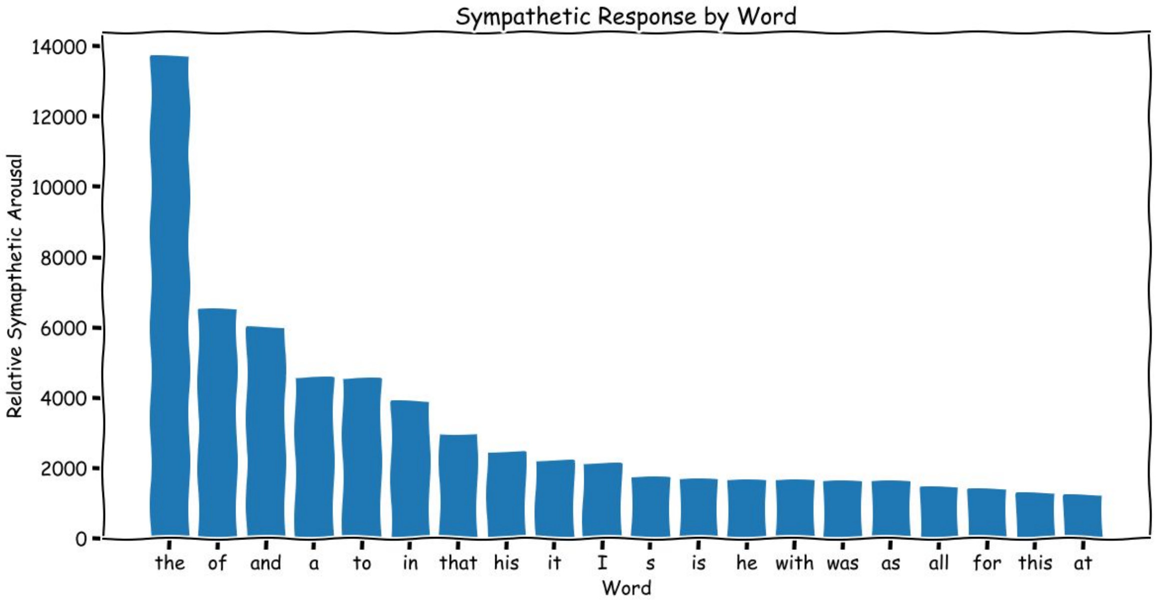
Presentation 2/25/2022

Meeting 4

Big Picture Goal: “Scary Words”

Like with “you,” certain words cause sympathetic nervous system responses. Create a measure of “sympathetic arousal” and visualize what words are the “scariest” in terms of biological arousal. Words -> Arousal prediction.



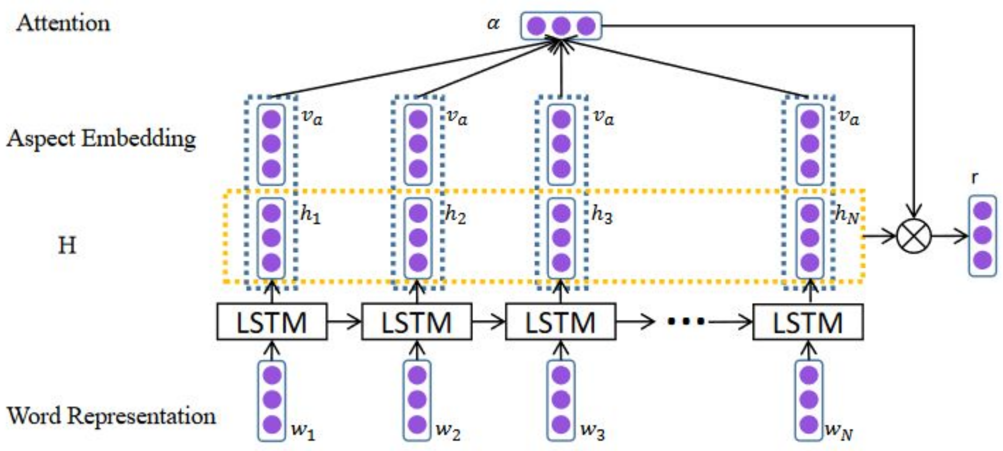
Could help in creating literature meant to evoke a certain response.

(Ex. Viscerally-effective horror stories. The Call of Cthulhu Keeper’s Rulebook has a list of human-generated “Lovecraftian Words,” but this is not based on any sort of scientific or literary consensus.)

(Note: The data here is just word counts from Moby Dick, graphed with matplotlib’s xkcd filter)

Possible Models for Words -> Arousal

Attention-based LSTM Neural Networks - Neural network chooses information to pass along when analyzing next sequence using “attention” mechanism.



Picture from “Attention-based LSTM for Aspect-level Sentiment Classification” paper

https://aclanthology.org/D16-1058.pdf

Big Picture Goal: Correlate words heard with sensor readings – produce (likely very inaccurate) guess of what words were heard

"It is possible that this approach could be used to decode information about what words a person is hearing, reading, or possibly even thinking." - Alexander Huth

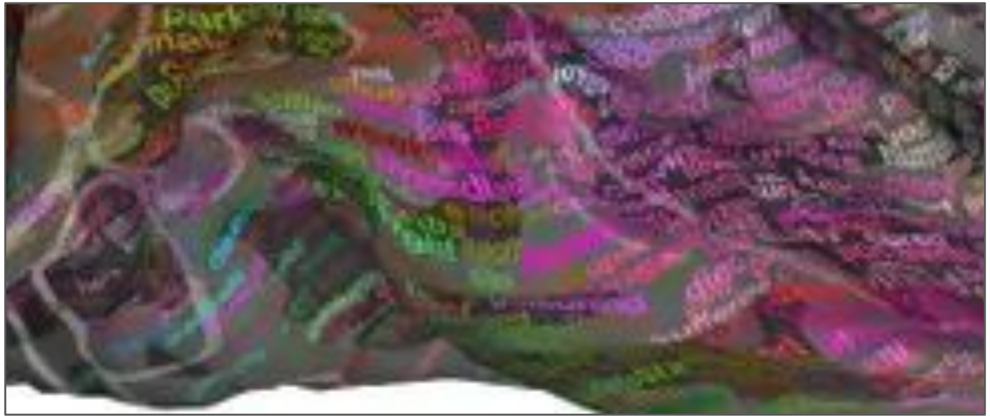
Previous studies have shown that certain words are tied to certain areas of the cerebral cortex. Determine if it would be possible to correlate the readings we have with certain words.

- Note: The study cited used an MRE. It could be our sensors' data is not granular enough.

[Link to Guardian Article]

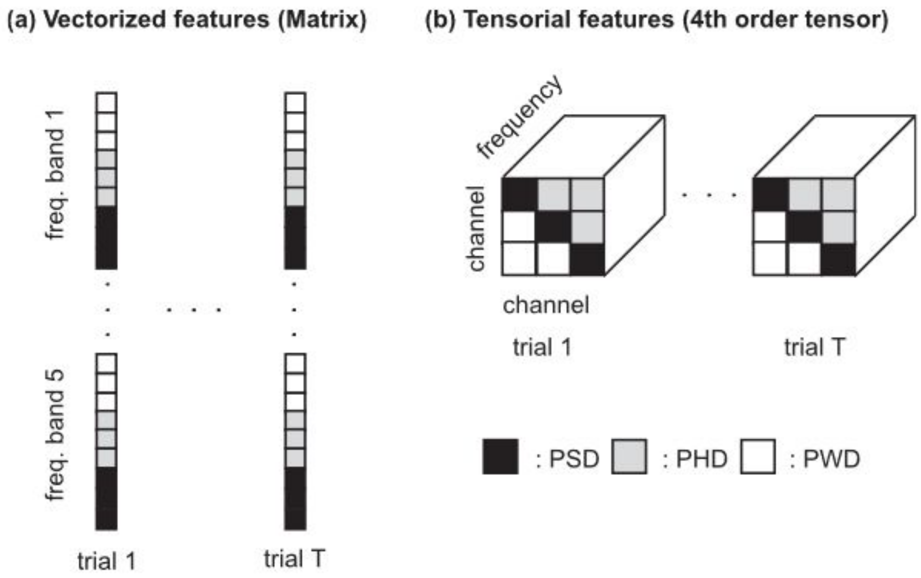
[Link to Study]

[Link to "Brain Atlas"]



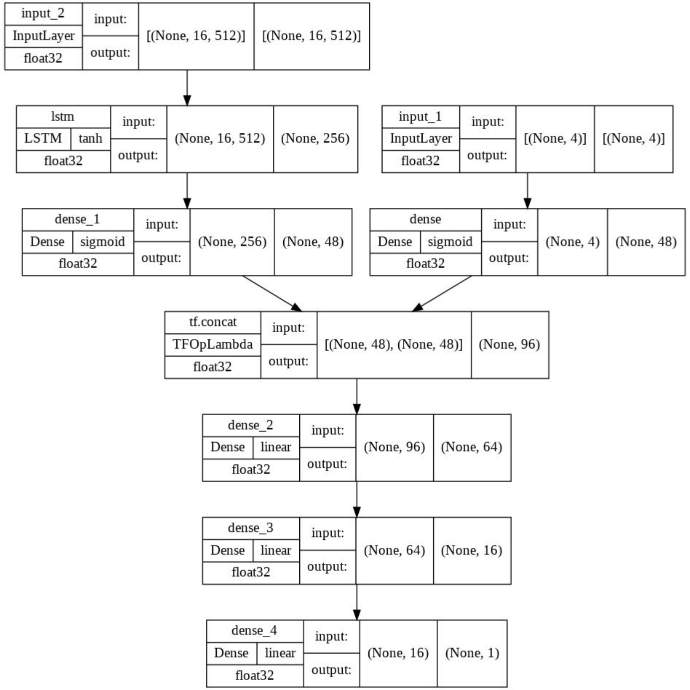
Possible Models for EEG -> Words

Tensor-based dimensionality reduction and regression - reducing the dimensionality of EEG data then performing a regression prediction (paper).



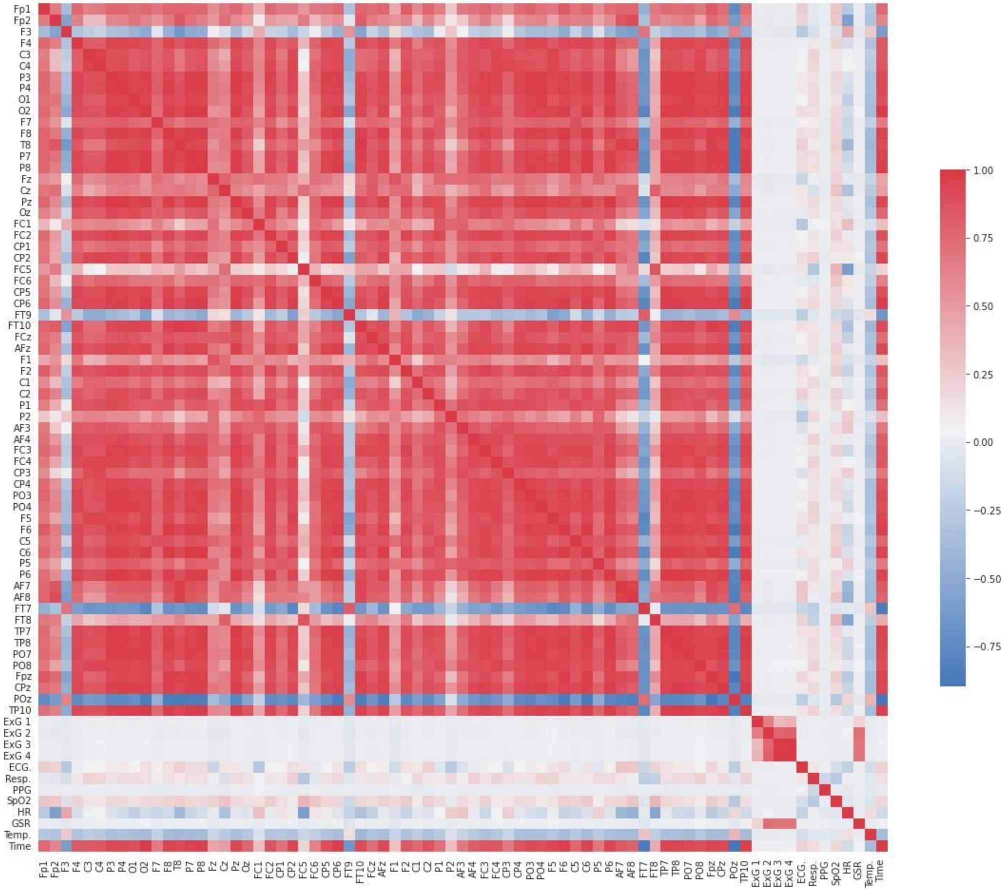
Initial Basic EEG -> Words Model

Directly using EEG dataset with only FFT as preprocessing. Might improve more with more preprocessing.

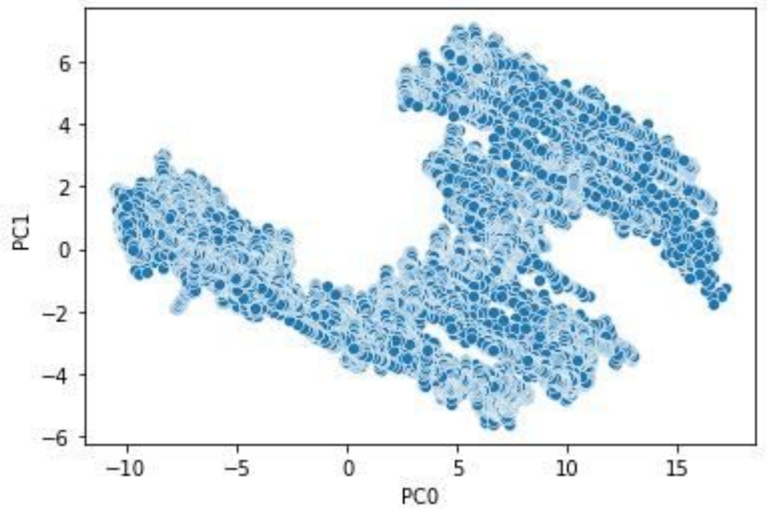



```
Epoch 1/5
111095/111095 [==============================] - 1703s 15ms/step - loss: 0.6935 - accuracy: 0.5001 - val_loss: 0.6873 - val_accuracy: 0.8515
Epoch 2/5
111095/111095 [==============================] - 1718s 15ms/step - loss: 0.6932 - accuracy: 0.5003 - val_loss: 0.6972 - val_accuracy: 0.8515
Epoch 3/5
111095/111095 [==============================] - 1728s 16ms/step - loss: 0.6932 - accuracy: 0.5004 - val_loss: 0.6886 - val_accuracy: 0.8515
Epoch 4/5
111095/111095 [==============================] - 1733s 16ms/step - loss: 0.6932 - accuracy: 0.5005 - val_loss: 0.6891 - val_accuracy: 0.8515
Epoch 5/5
111095/111095 [==============================] - 1740s 16ms/step - loss: 0.6932 - accuracy: 0.5000 - val_loss: 0.6921 - val_accuracy: 0.8515
```

Basic EDA



Fp1
Fp2
F3
F4
C3
C4
P3
P4
O1
O2
F7
F8
T8
P7
P8
Fz
Cz
Pz
Oz
FC1
FC2
CP1
CP2
FC5
FC6
CP5
CP6
FT9
FT10
FCz
AFz
F1
F2
C1
C2
P1
P2
AF3
AF4
FC3
FC4
CP3
CP4
PO3
PO4
F5
F6
C5
C6
P5
P6
AF7
AF8
FT7
FT8
TP7
TP8
PO7
PO8
Fpz
CPz
POz
TP10
ExG 1
ExG 2
ExG 3
ExG 4
ECG.
Resp.
PPG
SpO2
HR
GSR
Temp.
Time
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75



6
4
2
0
−2
−4
−6
PC1
−10
−5
0
5
10
15
PC0

Future Improvements

- Implement data preprocessing/cleaning/feature extraction based on EDA results to improve dataset
- Experiment with various model layers to see what works best (try using RNN/LSTM)

Analysis of Possible Directions

Analysis of Possible Directions

(Biometrics) -> (Audio) or (Biometrics) -> (Text)

(Text) -> (Audio) or (Audio) -> (Text)

(Text) -> (Biometrics) or (Audio) -> (Biometrics)

(Audio, Text) -> (Biometrics)

(Biometrics, Audio) -> (Text)

(Biometrics, Text) -> (Audio)

We could also do something totally different. This is just a guide to explain my thinking.

Directions I don’t like

- Directions that have been done before; May be hard to make further progress in time span. Also, they miss the point of using biometric data.

  (Audio) -> (Text) - Transcription

  (Text) -> (Audio) - Text-to-speech

Directions I don’t like

- Very difficult. Too many possible sequences of words for a given biometric response. If we choose one of these, we should relax the problem, e.g. trying to detect when an emotional word is spoken.

  (Biometrics) -> (Audio) and (Biometrics) -> (Text)

- Doesn’t seem useful

  (Biometrics, Text) -> (Audio)

Analysis of Possible Directions

(Biometrics) -> (Audio) or (Biometrics) -> (Text)

(Text) -> (Audio) or (Audio) -> (Text)

(Text) -> (Biometrics) or (Audio) -> (Biometrics)

(Audio, Text) -> (Biometrics)

(Biometrics, Audio) -> (Text)

(Biometrics, Text) -> (Audio)

Analysis of Possible Directions

Audio + Text -> Biometrics, or

Audio -> Biometrics

- We could use techniques for extracting emotional state from text/audio and test them against real biometric data.
- In some applications it would just be easier to measure biometric data directly.
- A possible real world application would be predicting biometric responses to a radio advertisement (for a marketing firm).
- Another case is predicting biometric response to reading a piece of text, but our current data doesn’t support that.

Analysis of Possible Directions

Biometric + Audio -> Text

- I am pretty keen on this idea.
- Clear real world application: wearable technology can record both audio and biometrics.
- What if we used both to predict words spoken/heard, sentiment of words, importance of words spoken (for summarization), or even emotional analysis?
- Using additional sources of data may allow us to make progress in these areas

Advantages and Disadvantages

- Clear real world application.
- Biometric data may resolve some ambiguities in transcription, but it also complicates the task.
- Audio and biometric data may similarly complicate the task of sentiment analysis.
- There are existing approaches to transcription and sentiment analysis.
- We don’t have clear targets for sentiment analysis. Would probably require some hand-labelling.
- Very limited dataset for sentiment analysis.
- Maybe we could use existing sentiment scores of words.

Advantages and Disadvantages

- Consider the types of data we have: biometrics, audio, text, and maybe video. Text stands on its own as the only one that cannot be measured directly by a sensor, because it is really an abstraction made by humans.
- Other abstractions from audio: tone of voice and sentiment of speech. These cannot be captured by text alone.
- Since I like it so much, I made a codename:

  Improving the Transcription of Sentiment and Language with bIometric Tools

  (ITS LIT)

Semantic analysis

- Goal of semantic analysis: systematically identify, extract, quantify, and study affective states and subjective information
- Some approaches use a knowledge base of known “emotional words”: “angry”, “mad”, “hate”, etc.
- Perhaps we could use our data to automatically assign such labels to words.
- Not sure if it has been done before
- A similar task, emotion recognition, seeks to identify whether someone is angry, sad, etc.

Using the Transcript

- The audio data captures the exact time a word is spoken, but doesn’t specify which (English) words are being spoken
- The transcript doesn’t capture the the exact time a word is spoken, but it does tell us the words spoken and in which order
- How might we consolidate or synchronize these sources of data?
- One approach – make a time-stamped transcript using existing automatic caption software.