

Neurotech/ML Stuff (best read in-order):

☐ Speech Recognition with RNNs: <https://arxiv.org/pdf/1303.5778.pdf>

- Using only **audio**, not EMG
- Bidirectional end-to-end **LSTM RNNs + CTC**
  - Long Short-term Memory (LSTM) - used for context/memory in longer sequences
  - Bidirectional Recurrent Neural Network (RNN) - layers feed forward but also backward, used for sequences
  - Connectionist Temporal Classification (CTC) - type of loss function for training RNNs
- Achieved 17.7% error rate on TIMIT

☐ EMG-UKA Data Corpus:

[https://www.researchgate.net/publication/271824674\\_The\\_EMG-UKA\\_Corpus\\_for\\_Electromyographic\\_Speech\\_Processing](https://www.researchgate.net/publication/271824674_The_EMG-UKA_Corpus_for_Electromyographic_Speech_Processing)


- Two datasets, full one has 7:32 hours of data (452 minutes), trial version has 1:52 hours of data (112 minutes)
- Has EMG readings of **audible, whispered, and silent** reading
- Includes audio, transcription, and (badly) aligned phonemes
- <https://www.kaggle.com/datasets/xabierdezuazo/emguka-trial-corpus>

☐ End-to-end with Subvocal Speech:

[http://www.pamelatoman.net/wp/wp-content/uploads/2018/06/subvocal\\_speech\\_recognition\\_paper.pdf](http://www.pamelatoman.net/wp/wp-content/uploads/2018/06/subvocal_speech_recognition_paper.pdf)

- Uses **EMG-UKA** data corpus
- Uses **LSTM RNNs + CTC** in a similar way to Paper #1
- Attempted data augmentation, but no significant improvement
- Achieved 70.2% character error rate on EMG-UKA

☐ Attention is All You Need: <https://arxiv.org/pdf/1706.03762.pdf>

-  Transformer Neural Networks - EXPLAINED! (Attention is all you need)
- **Transformers** are incredible for **sequence-to-sequence**, examples include:
  - GPT-3, Stable Diffusion, Whisper, Gato
- Typically much more accurate and faster to train than RNN/LSTM architecture

- ☐ Whisper: <https://cdn.openai.com/papers/whisper.pdf>
  - **Transformer model** that transcribes audio to text
- ☐ TEMGNet: <https://arxiv.org/pdf/2109.12379.pdf>
  - **Transformer model** using sEMG (surface EMG) for hand gesture recognition
  - ~80% accuracy on 18 classes
- ☐ ViT: <https://ieeexplore.ieee.org/document/9834070>
  - Seems to be better than TEMGNet
  - ~90% accuracy on 18 classes (same dataset as TEMGNet)

Extra:

- <https://link.springer.com/article/10.1007/s41133-016-0001-z>
- Transformers for Image Recognition at Scale: <https://arxiv.org/pdf/2010.11929.pdf>
- EEG based Continuous Speech Recognition using Transformers:  
<https://arxiv.org/pdf/2001.00501.pdf>

## Transformer Implementation Tutorials:

Transformers in Keras:

- ▶ Introduction to Attention, Transformers and NLP in Keras

Transformers in PyTorch (2 yrs ago):

- ▶ Pytorch Transformers from Scratch (Attention is all you need)

Transformers in PyTorch (6 months ago):

- ▶ Transformer: Concepts, Building Block, Attention, Sample Implementation in PyTorch

Transformers in PyTorch (Recent Paper by Microsoft):

<https://paperswithcode.com/paper/torchscale-transformers-at-scale>

GPT (Generative Pretrained Transformer) in PyTorch (very recent)

- ▶ Let's build GPT: from scratch, in code, spelled out.