September 5, 2019

Technical Learning

Learned the following in Python 3

- General syntax and operators
- Functions (defining, function headers)
- Structures (arrays, dictionaries)
- Printing
- Indexing
- Input/Prompts
- Conditionals (if, elif, else)
- Iteration (for, while)

Next to learn

- File I/O
- Built-in functions

September 6, 2019

EMG Research

Questions:

- What is EMG?
- What are EMG sensor limitations?
- Has is EMG placement determined?

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1455479/

General EMG Information

- EMG = Electromyography
- Measures electrical signals related to muscle action potentials (outputs proportional voltage)
- Methods to measure muscle fiber action potentials from muscle fibers of a single motor unit action potential
    - Invasive electrode: Consists of a needle/lead that is placed directly in a muscle fiber
    - Non-invasive electrode: superficial electrode that records signal composite of all muscle fiber action potentials occurring under electrode
- Complications: noise and signal distortion
    -

https://www.sciencedirect.com/science/article/pii/S1877050917324717

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3821366/

https://www.ncbi.nlm.nih.gov/pubmed/15298438

Physiology Research
Questions:

- For the scope of this prosthetic, what does “grip” mean for the body? What does it mean for the prosthetic? (How is grip being translated?)
- What muscle regions/groups are of interest?
- For the expected muscles/muscle regions, what are expected measured values for said muscle activity?

https://study.com/academy/lesson/forearm-muscles-anatomy-support-movement.html

https://www.ncbi.nlm.nih.gov/pubmed/10048518

https://www.ncbi.nlm.nih.gov/pubmed/10048518

Team Meeting

- As EMG sensor research progresses, I was assigned work regarding the voice of Shimon.
- Task: mark 9:00 minute audio files, identifying usable audio, silence, and lyrics.

September 8, 2019

Shimon Audio Work

- 9:00 minutes of the first audio file were labeled per phrase with understood lyrics from the song. The 9 minutes were uploaded to Richard’s dropbox and Richard was notified via Slack. Awaiting response to continue or fix the work done so far.

September 9, 2019

Shimon Audio Work

- Separate audio files, splitting all audio segments and removing silences.

September 11, 2019

Shimon Audio (under Rishi)

- Instructions for Shimon audio were clarified. Songs to be assigned later.

September 13, 2019

General

- Participated in survey for Shimon.

Meeting with Jason for Drumming Arm

- Reviewed Python basics, importing databases, basics of machine learning. Waiting on Jason to share basic machine learning model.

September 18, 2019

Shimon Audio

- Completed assigned task regarding segmenting audio and writing txt files with appropriate lyrics. This was done for the songs: All I Ask, Love Fell Down, My Friend, Paint it Black, and You Gotta Die Sometime. The files were uploaded to the provided dropbox link.

September 20, 2019

Shimon Audio

- Awaiting response for audio synthesis. (Delayed due to late submission of audio files)
- Meeting Time confirmed: every Friday at 2pm (the assigned course time)
- New Task: read up on assigned research paper for synthesis and discussion next week

Learning Machine Learning

- Use Google's Machine Learning Crashcourse
- Goal: Understand classifier models? (relevant to Drumming Arm ML model)

September 21, 2019

Background Information: https://skymind.ai/wiki/generative-adversarial-network-gan

ML Voice Synthesis Papers: https://arxiv.org/pdf/1908.01919.pdf
Guiding Inquiries:

- What methodology was used?
- What datasets were used?
- What method is used to evaluate the model?
- Are there examples/demos of voice samples generated by the model?
- Other interesting findings from the paper (ex. *code repository*)

Adversarilly Trained End-to-End Korean Singing Voice Synthesis System

Introduction

- Problem/Motivation: for neural network-based SVS systems, predicting vocoder features is limited
- Proposition: end-to-end framework that directly generates linear-spetogram
- Constraints: incorporating end-to-end framework → complicates the SVS system model because the complex target (the linear-spectrogram) requires more training data
- Novel Approaches/Contributions:
    - End-to-end Korean SVS system
    - Phonetic enhancement masking method → efficient use of training data
        - Modeled low-level acoustic features related to pronunciation from test info → more accurate pronunciation
    - Reusing input data at super-resolution stage and training with adversarial manner → improve sound quality and realism of singing.

Related Work

- End-to-end text-to-speech system trained as an autoregressive manner → improved performance than conventional
    - More controllable elements of speech (prosody, style, et.)
- For adversarial training, conditional GAN (w/ projection discriminatior) and R1 regularization → train autoregressive network and super-resolution network (ie achieve end-to-end network)

Proposed Network

- Mel-synthesis network produces mel-spectrogram (from time-aligned text and pitch inputs)
- Super-resolution network converts the generated mel-spectrogram to linear spectrogram
- Discriminator uses upsampled result w/ mel-spectrogram to train network adversarially

- Input Representation
    - Inputs: Korean text, MIDI pitch input, audio or pronunciation
    - Korean words can be split into onset, nucleus, and coda
        - Nucleus generally occupies majority of the pronunciation

        - Onset and coda were assigned to first and last frames of the word
            - Not an accurate timing for each phoneme
            - BUT a convolution-based network could handle the issue
- Mel-Synthesis Network
    - Modified SVS System uses:
        - Pitch encoders (similar to text encoder) [7]
        - Conditioning method of encoded pitch on the mel decoder [17]
        - Phonetic enhancement mask decoder → allows for data with various pronunciation-pitch cominantion and more realistic pronunciations
- Super-Resolution Network
    - Reused aligned text and pitch info in SR network → exploiting useful info in generation process
    - Utilize adversarial training methods to make SR network produce realistic sounds

Experiments

- Dataset
    - Created their own Korean singing dataset
    - Prepped Singing voice MIDI files
    - Recorded female vocalist
    - Aligned MIDI files with recorded audio
    - Assigned the syllables in lyrics to each MIDI note
    - 49 songs were used for training, 1 song for validation, 10 for testing
- Evaluation
    - Method 1: phonetic enhancement → pronunciation accuracy
    - Method 2: Local Conditioning pitch and text to SR →
    - Method 3: adversarial training method → sound quality
    - Quantitative Eval
        - Model generated a singing voice with correct pitch and timing
    - Qualitative Eval
        - Listening test with native Korean speakers to evaluate pronunciation accuracy, quality, and naturalness
    - Result
        - The best results achieved when combining all three methods

---

<u>September 28, 2019</u>

VIP

- Asked other members of research how to submit this logbook (instructions are unclear)

Shimon Voice Audio

- Identified correct phonemes/pronunciation of data set

        - Used https://en.wikipedia.org/wiki/ARPABET as reference for sounds
- Noticed possible changes in lyrics written: (current → whats said in audio)
    - Big_iron-04-02
        - Arizonia → Arizona
    - Breathe_owl_breathe-05-03
        - Fire in my throat → fire up my throat
    - An email will be sent to Rishi tomorrow regarding these possible changes

---

September 29, 2019

Shimon Voice Audio

- Emailed Rishi, awaiting response for change/submission

October 03. 2019

Shimon Voice Synthesis

- Submitted corrected txt phoneme info

Drumming Arm

- Continued with ML Crash Course (Learning how to use Tensor Flow)

October 04, 2019

New Assignments

- Familiarize myself with ML model for the Drumming Arm and Shimon Voice Synthesis

---

October 09, 2019

Shimon Voice Audio

- Understood code until reducing error and transforming data

Drumming Arm

- Almost completely anotated code
- Useful resources: sklearn documentation and frequent function searching
    - https://scikit-learn.org/stable/tutorial/basic/tutorial.html

---

October 11, 2019

October 16, 2019

Drumming Arm

- Finished code anotation
- Began watching additional ML videos (https://www.youtube.com/watch?v=HcqpanDadyQ&list=PLIivdWyY5sqJxnwJhe3etaK7utrBiPBQ2&index=1)

October 17, 2019

Shimon Voice Synthesis

- Recurrent Neural Networks (https://towardsdatascience.com/learn-how-recurrent-neural-networks-work-84e975feaaf7)
- GRU Networks (https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be)
- LSTM Networks (http://colah.github.io/posts/2015-08-Understanding-LSTMs/)
- Seq-to-Seq Modeling (https://towardsdatascience.com/understanding-encoder-decoder-sequence-to-sequence-model-679e04af4346)

October 31, 2019

* shift to Shimon experiment under Richard

Experimental Design

- Objective: justify the application of robotics to music creation/generation/performance
- Method: Comparing Shimon to same audio via speaker
    - Each participant will listen to 3 songs, once performed by Shimon and once via a speaker
        - Randomized order of the 3 songs
        - Randomized order of delivery (Shimon or speaker)

- Use survey/quiz to assess “is it creative?” measuring
    - Emotional response
    - Enjoyment
    - Engagement/entertainment

How To Measure Creativity?

- Engagement:
    - Evaluate instances of participation during stimulus
        - Stimulus Types (indep): live music, recorded music
        - Participation (dep): active, passive, none
            - Active: sing, hum
            - Passive: toe tapping, hand tapping, clapping, following provided song sheet
            - None: just listens
        - Factors that can influence participation:
            - Are they songs people know, easy to sing, or song sheet provided
    - Reference: https://scholarworks.wmich.edu/cgi/viewcontent.cgi?article=2641&context=honors_theses
- Enjoyment:
    - Have participants hooked up to an ECG and analyzed heart rate and high/low frequencies. HF/LF and LF/HF ratios provide information about participants sympathetic response to the music delivery/conditions.
    - Reference: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4841601/

November 13, 2019

Research of the effect of *Engagement* on *Perception*

- https://www.ncbi.nlm.nih.gov/pubmed/15495906/
    - Active memory (short-term) results in time perception distortion, increasing w/ greater temporal production
    - Derived Ideas for Experiment
        - Require subjects to recall aspects of the performance/video and follow-up with asking them about what they perceived
            - Compare accuracies of their recall
            - Ask about time perception of the performance/video and compare their accuracy
- https://www.ncbi.nlm.nih.gov/pubmed/7808276
    - Findings: The subjective perception of time shortens as the subject devotes less attention to a temporal task
    - Derived Ideas for Experiment
        - Have all subjects report time perception for one of the songs via both mediums

- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5496904/
    - Perception of time can be affected by the engagement of a person in an activity
    - The brain commits both time and attentional resources to an engaging stimulus
    - Results from this study showed that:
        - Across viewers, there was more uniform report of time's passage with more similar neural processing
        - Time perception accuracy was not necessarily statistically significantly different between engagement states (however, different content was provided for these measurements, they did not compare the same content delivered via multiple mediums)
    - Derived Ideas for Experiment
        - Give the subjects the option to stop when they like
            - Possible complications that could invalidate data:
                - Subjects want to spend the least amount of time in the test and will stop either stimulus out of convenience.
                - Subjects my want to stop watching/listening to a song the second time they are set up to see/hear it.
- https://www.psychologicalscience.org/observer/the-fluidity-of-time
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3152725/

November 13, 2019

Ideas regarding the experimental design and methodology

Experiment Outline:

- Receive student
- Retrieve their info & agreement to participate in the study (req)
- Spectator will keep track on <Observation Form>
- Listen to <Song 1> in both mediums
- Answer <Questionnaire 1>
- Listen to <Song 2> in both mediums
- Answer <Questionnaire 1>
- Listen to <Song 3> in both mediums
- Answer <Questionnaire 1>

* order of songs and their mediums are randomized
* questionnaire types will randomly but equally be distributed among the song pairs

Observation Form:

- Per song and per medium, record occurrences of active and passive engagement
    - Active Engagement: sing, hum
    - Passive Engagement: toe tapping, hand tapping, clapping, snapping, nodding to rhythm

- Example:

Engagement Type		Song 1		
		Live	Recording	
Passive	Toe Tapping			
	Hand Tapping			
	Nodding			
	Other			
Active	Sing			
	Hum			
	Other			

Subject's Questionnaire Example Questions:

* before <song>*
Keep track of <element> during the (performance/video).

* after <song#> via <medium1> *
Estimate the number of times <element> occurred: _____
Estimate the duration of <medium1>: _____min

* after <song#> via <medium2> *
Estimate the number of times <element> occurred: _____
Estimate the duration of the <medium2>: _____min

* after <song#> via both mediums *
Which medium was more enjoyable: (medium1 or medium2)

— or —

On a scale from 0 to 5, rate which medium was more (enjoyable, entertaining, appealing, etc.):

0 (medium1) ———— 1 ———— 2 ———— 3 ———— 4 ———— 5 (medium2)