First Quarter 2018:

I am currently working on two research projects: (1) Recognition of code-switched speech and (2) Building speech synthesis systems with intent.

Code-switched speech recognition:
I have been working on this project from over a year now. The main idea is to (a) derive either linguistically motivated or data driven 'styles' of switching and (b)incorporate the derived information into the model training procedure. For the first part, we have worked on identifying the language at the frame level. In addition, we have segmented the switched corpus into multiple styles. For the second part, we are exploring different ways to incorporate the derived information. We began with language model. However, this has not yielded the expected performance as of now. I have a stack of experiments to do in this space that spans summer.

Building speech synthesis systems with intent:
I have begun working on this project this semester. The main idea is to modify the training procedure of a vanilla voice by incorporating information about the content to be focused. This will help us build speech synthesis with explicit control. In other words, we will be able to differentiate between “I am **here**” and “**I** am here”. I have finished the baseline implementation for this by a rudimentary separation of the utterances based on the identity of the word to be focused. Over the summer I will be working on approaches that can implicitly learn this.

Retrospection:

I had the following goals:

(1) Finish the course requirements which meant taking three courses.
(2) Submit our system to the ASR Speech Recognition Challenge by Microsoft Research India.
(3) Submit our system to the Blizzard Speech Synthesis Challenge with WaveNet based vocoding.
(4) I personally had an objective of making all of our code docker compatible and notebook friendly.
(5) I also wanted to be very responsive to the users that have issues in our core software stack: Festvox, flite, speech tools and Festival.

I am genuinely pleased and satisfied with the amount of work I have got done this semester. I believe that I have been able to accomplish about 90% of what we actually expected to at the beginning of the year. In addition, I have also involved myself in multiple aspects that I have not anticipated at the beginning of the semester. I will explain this in more detail below:

(1) *What went well:*

I have always envisioned myself continuing at LTI for my PhD. Therefore, it was important for me to kickstart a longer term action plan in one of the semesters during MLT. In other words, I wanted to do at least one of these before beginning PhD: (A) Build a good enough neural infrastructure for our speech pipelines that enables us to experiment freely with Natural Language Understanding during PhD. (B) Collaborate with the other brilliant students here at CMU. Since my advisor gives me an enormous amount of space and time to grow as a researcher, I found that the best way to accomplish (A) was to participate in the challenges. For example, I have revamped our acoustic modeling module as a part of our submission to Voice Conversion Challenge 2018. For Blizzard Challenge 2018, I have designed a Wavenet based implementation with a quality that forced us to revamp our vocoder in our submission. I want to be able to experiment freely with conversational agents during my PhD which would need expertise in multiple areas of speech and text. As a part of this, I have forced myself to build a large chunk of our submission to the Microsoft ASR Challenge 2018. Similarly, when an opportunity presented itself to participate in the paralinguistics challenge (COMPARE 2018), I took it and built from scratch our current submissions in all the three tracks (recognizing self assessed intent, recognizing intent from speech from disabled people and recognizing different styles of children's cries). (B) was powered by two realizations: (i) I realize that it is foolish to not take advantage of the tremendous resource available in CMU: brilliant students. (ii) I want to be able to take up an academic position after my PhD. Therefore I wanted to ensure that I collaborate with people and do it organically. For each of the 5 challenges I did this semester, I have deliberately collaborated with people. Specifically I have worked with people from LTI, ECE, MCDS and MIIS programs within CMU. Most of them have been organic and fruitful. I plan to keep doing through my PhD.

(2) What did not go well:

I have clearly overestimated the amount of time and energy I needed to spend on all the things I got involved in. I originally anticipated to spend around 20 percent of my time on challenges. However, I ended up spending more than 37%. The coursework which I estimated would take up 30% ended up taking more than 43%. This has directly affected the amount of time I get to spend on my core research. Most of the collaborations I got involved in were early stage which meant I had to put in effort for knowledge transfer. However, in retrospect, I do think this is a good thing. It is essential to fall short, reevaluate and correct the course. I did a variant of this. For example, midway through the semester, I have realized that things were getting overwhelming. To counter this, I have done two things: (a) I forced myself to have a clear write up for every meeting. After spring break, I had write up in conference submission format for more than 80% of the meetings I took. (b) I began working in isolation by cocooning myself. This meant I needed to be more communicative and online via platforms like slack and multitask which is opposite to the way I prefer operating. But I have applied all the techniques learnt in the time, energy & stress management seminars offered by CPDC last semester to be able to do

this. For example, I have built chatbots to explicitly handle my time and energy allocations for each of the projects I was doing.

(3) *What do I aspire to improve upon:*

One thing that I am dissatisfied about is the reduction in percentage of seminars and talks I attended this semester. I usually attend more than 60% of all the presentations that happen in CMU including GCC, CPDC and Eberly. In fact most of them have shaped how I structured my workload during this semester. But this semester, I have attended less than 30% of all the talks including the ones of LTI. I think this is a huge opportunity cost from my end, especially today given the pace at which AI research is progressing. This is one of the things I plan to balance going forward.

LookAhead:

I am very excited for the next semester. I will be taking only one course which is primarily project based one. I anticipate that this will double the amount of time I have. Usually people do one of two things when this happens: (1) Do twice the number of things or (2) Do the same things twice as well. At this point, I believe I will be taking a portfolio balancing view of this. I will allocate 20% of my extra time to do new things but a major chunk (80%) of my time and energy will directly go into the core research. Given the developments in AI, I believe that it is imperative for students to be aggressive and play offense even at the cost of overburdening themselves. Usually early stages of PhD are the time to take risks and explore things. Instead of doing multiple things in the name of exploration, I plan to take a trinity approach: I believe that going forward, success in deep learning will be based on the trinity of (a)data, (b)algorithms and (c)infrastructure. CMU has always been great at (a)algorithms. Personally, I have built the algorithmic prowess required to further our research via participation in challenges. I will dedicate a chunk of my summer to empower the data(b) portion of the trinity. Specifically, I plan to do a semi large scale data aggregation for our code-switching research. There will be one component of this which is one-time(recording in our studio) and another which will be perennial (crawling YouTube and Twitter continually). This aggregation will be continued in the Fall semester and I can see us exploiting active learning for data augmentation. We are making steps to build the infrastructure stack(c) required. This will also include two components one of which will be maintained by us locally (similar to the clusters we have) and the other will be utilizing the developments being done by the tech giants. We have shortlisted the GPU stack we are interested in maintaining locally. I already have access to the TPU infra of Google and I plan to acquire access to the AWS GPU infra as well.

To conclude, I believe this semester was a crucial building block for kick starting my PhD. Based on the way Google rates its employees, I rate myself a 3.2 out of 4. This implies that I have done more things than was originally anticipated (hence score of > 3) but fell short on some aspects that I should have done better. Specifically, not being able to submit our work on code-switching to Interspeech 2018 (< hence score of 3.5).