Malachi Shoffner

1. Basic performance analysis of the speech quality

The first section focused on basic analysis of the speech quality. The biggest challenge in this section was simply becoming familiarized with the format of the label.json and log.json files themselves. Once the structure of the data became clear, it was as simple as accessing dictionary keys. Thus, no deficiencies were noted in this section.

An interesting realization I made is that only 9 of my 98 tracked dialog turns were not understood, which were denoted by an "N" in the CSV output file. This indicates that, while the state tracker can definitely be improved, it is trained well enough to at least partially understand the sentences spoken to it.

2. Automatic annotation of dialog state change

Section 2 involved annotation of dialog state tracking changes. I faced several challenges during this stage, a far cry from the comparative simplicity of section 1. The very first challenge I ran into was determining whether a value should be removed or not. It was straightforward enough to figure out whether the machine was adding or modifying a request, since those two actions had "act" values such as "inform" and "reqalt" in the log files. However, I was at a loss when it came to filling in the "d" key in the dictionary.

I came upon a solution to this challenge when inspecting the provided sample data. I realized that whenever the previous turn contained a request, and the current turn being tracked contained a different request, the previous request was deleted. This is reflected in my program by the "prevCam" and "curCam" variables, that track the previous and current turns, respectively.

I made the observation that in both the sample data, and my own provided data, there were very few attributes for which a value was removed. This would seem to be a function of the words spoken to the machine, as opposed to a reflection of the speech quality.

3. Automatic generation of dialog feature information for miscommunication detection

Section 3 includes generation of feature information. The purpose of my own dialog feature information was to determine if there was any correlation between the length of an utterance, and the state tracker's understanding of that utterance. By creating a frequency distribution containing sentence length and system understanding, I was provided with a visual representation of these two pieces of information.

I hypothesized that the longer the sentence was, the greater the chance of miscommunication, being that each word carries with it a probability of being misunderstood. However, the results ren-

dered my hypothesis false. Out of the 8 utterances that were not understood, none of them had 10 or more words, and only one had more than seven words. Given the small amount of sentences with 10 or more characters (a mere 11), it is impossible to find conclusive evidence that a longer utterance leads to better understanding, but based on my sample size, that seems to be the trend.