Grant Yap
gyap@u.rochester.edu
11/11/2022
CSC 412
HW # 5
Youtube link: https://youtu.be/R7_CbbQ6oCo

a) Speaking Rate and Feedback
   i) My thoughts on the accuracy of the speech recognition system is that it works pretty well, if the user is speaking continuously. The issue comes with the lag times that the computer takes to transcribe the speech to the text box. It basically breaks on words that have the same exact sound. It becomes difficult for the speech recognition to distinguish between the different “toos”. When I was trying to test the system for the credit card numbers, it constantly kept changing the “two” that I meant to “to” which was very annoying, but with basic speech recognition and conversations it does a pretty good job. The speech recognition also cannot input periods, or any type of punctuation.
   ii) I would argue that it is not a super reliable metric to measure speaking rate because of the fact that speaking is very fluid. For example, when I was trying to speak and I ran out of my train of thought, my wpm became very low, but when I came back to what I needed to say, it sped up to a normal pace very quickly.
   iii) I believe that 133 being the slow threshold was very harsh for me and for others. I tended to speak at around 100-120 wpm which was more comfortable to me and I felt I could get my message across much easier. I would agree that 188 wpm is a decent threshold for a fast speech since I had to blurb my words out to get a decent handle of it being displayed for the videos.
   iv) The feedback mechanism I used was a change in the “div” where I made a new div with id “words” using this, I was able to display information after every long enough pause which would be added to the final transcript. I then used a time variable by taking the current date, and the last timestamp that happened from the beginning or last long pause. Using this, I would be able to update every time there was a long pause like the assignment prompt suggested. I considered using the “interim” variable but found that this became too inconsistent. I ended up dividing the total words spoken, with the time * 1000 all multiplied by 60 to simulate words per minute. I used the div as this was the easiest for me to manipulate and use to draw attention.
b) Automated Agent
   i) I thought my responses were generally appropriate as I had trouble implementing the agent in an efficient manner. I ended up using a bunch of if statements which is poor coding but I was time crunched. Some cases were left out such as an overarching check to make sure that the user made the correct responses, but I felt that from my speech, you were able to reasonably get all the responses.
   ii) I essentially took the most recent phrase that the user spoke and made sure to check for a reasonable keyword that would be given. If the keyword I was looking

for wasn't obvious, I would give options with the prompt that I gave to the user. I also started the automated agent by adding a button similar to the copy and email button that was already part of the api.

iii) I personally do not think the google speech API is good enough to have actual conversations with people because people would need to verbally speak the punctuation needed to make proper sentences for the computer to understand. With better programming, conversations could become very efficient, but if there was a special conversation that has never been heard or accounted for before, then it would be difficult to cope with all the responses needed. My thoughts on automated agents are that they are a worthy endeavor to go for, and I hope to work on those types of problems in the future as that is what I feel is an advance in technology at this time.