Progress Report 1 - Human interaction with voice-assistant agents

Current Progress

For my MSc application I have been creating an application that will assist in the capturing of in-home conversations and confrontations that could be causing mental health issues for young children by recording audio surrounding an ‘anchor point’. This way I will be able to capture the conversations preceding and proceeding the point of interest, giving the analyser some key information to dissect the causes and root for the problem. The idea is for the application to be placed in a home to allow for a more natural collection of data. To create the application, I have been utilising the capabilities of pyaudio for audio capture as this seems to be the most widely used and reliable source for obtaining audio from a speaker using Python. I have then been manipulating the audio heard through the main speaker port dependant on which thresholds of sound the audio reaches, there are three levels for the sound to be categorised into: silence, talking and another for shouting.

Currently on starting the application if there is any shouting is detected the application will notify the user and record the shouting and any ensuing conversation for a reasonable amount of time after the application has stopped hearing any sound, if any sound is heard after the shouting the timer is reset and the recording will continue. When all audio has stopped being heard the application will then normalize all the sound so that it doesn’t reach past a certain level for comfortable listening for the user. There will also be a trim of any silence that is heard at the start or the end of the file heard, this way the silence recorded just before the application closes will be removed. When all this processing is complete the file will be saved with a date and time stamp as to not overwrite the previous file recorded.

Currently, if the initial audio heard is talking it will enter a loop which will record a certain amount of talking, if the limit of talking is heard and no shouting detected the audio will reset the audio counter and remove any of the talking heard from that loop of talking. This loop will continue until shouting is detected. When shouting is detected it will follow the same path as if shouting was analysed straight away with the added talking attached to the start of the recording.

For the written aspect of my project I have been researching Social Signal Processing which covers; Modelling (*identification of principles that govern use of social signals*), Analysis (*Automatic detection and interpretation of social signals in terms of principals*) & Synthesis (*Automatic generation of artificial social signals*) of human-to-human and human-to-machine interactions. Although my project is not specifically aimed at the synthesis of social signals, there is still many aspects of the recreation that require the detection of the signals for them to be recreated. Social intelligence is one of the most important qualities humans can possess that leads to success in life so it is a natural progression for the computers that we spend most of our lives around to develop these qualities and be able include facets of social intelligence. For this to take place it is important more research is done into social signal processing.

Timetable

As you can see below I have updated the project timetable to include the new tasks that took a substantial amount of time to complete, I have also updated planned start / finishing and actual start / finishing to reflect when events took place in the development of the project. From the timetable you can see that although some tasks overran and affected the start of other tasks I am still on track to complete the project with time to spare. As I am using an agile approach to the project I am not concerned that not all the tasks I have completed were not on the initial timetable. I also am confident that although there will more than likely be more tasks added in the coming weeks, that the project will be completed in time and be a better application because of the changes.



Future Developments

For the application, in the coming weeks I will be aiming to develop a way to take away talking from the start in seconds as seconds of new talking are heard instead of in blocks of talking which is the way the current system operates. I will also be aiming to implement a way of detecting if a conflict has not been resolved (*e.g. if there is shouting and no talking heard after to resolve the shouting*). I will also be looking at if there would be away to give constructive feedback to the users through technology. I may look into if the application could be imported into a physically external device so that it can be tested on real subjects although there may not be enough time for this to take place within the time frame of the MSc project

For the written aspect of the project I will continue to research Social Signal Processing and I will start to write the specific sections of design decisions, goals, and methodologies for evaluation of the proposed system. I will also be researching effective methods of intervention that could be achieved with the information that is gathered from the application as this could be implemented in future work