Lab3. Tuning ASR and TTS

Part A: Hard cases for Speech Recognition

1. Names for fictional places, people or objects that are not recognized (in your project)?

   My project is based on the folder 'in case of crisis or war' that can be found on msb.se (the 'home preparedness' section). I am not using any fictional places or names in the project, however, the dialogue is in English but some terminology have strong Swedish connotations which may impact ASR:

   - 'MSB' (Myndigheten för Samhällsskydd och Beredskap): Swedish speaker might use Swedish pronunciation / ɛmesbe/
   - MSB is called 'Swedish Civil Contingencies Agency', but SCCA is not an establish acronym
   - The Swedish public broadcaster is named: Sveriges Radio. There may be several different variants of this that the ASR needs to handle:
     - Sveriges Radio pronounced in Swedish with Swedish speaker
     - Sveriges Radio pronounced in Swedish by non-Swedish speaker
     - Various English translations of Swedish Radio (e.g. 'Sweden's radio', 'Radio Sweden', 'Swedish Radio' etc)
     - Information about local P4 channel frequency is found at: krisinformation.se.
2. I tested the confidence score for saying 'MSB' as a Swedish person would and the results that I got were:
   - 'Then misspell' – confidence 0.074
   - 'Atmosphere' – confidence 0.046
   - 'MSPO' – confidence 0.049

   When speaking with English pronunciation 'MSB' was recognized correctly with confidence scores >0.688. Confidence scores for the Swedish pronunciations are low.

3. The recognition falters – I suspect – because the ASR is trained to recognise English speech and the training data does not include acoustical data for Swedish spelling sounds. As a next step I would look how to fine-tune the ASR to recognize Swedish spelling of 'MSB' by including text + recording sample data to the model. Furthermore, I can use the ASR confidence score as a threshold in the dialogue manager to trigger "I'm not sure I understood what you said." error handling. Low confidence score – along with analysing the text recognized – can also be used during test and evaluation of the system to highlight problem areas.