Endpoint ID : 02e2587a-a1f7-4fff-9e0f-d519fc45ab9a

The azure speech recognition is pretty good in recognizing a variety of different fields however, there is also a range of words and names that it has quite a hard time recognizing correctly. It surprisingly does really good with medical diseases even the ones that are of Greek origin. It does struggle though with names of plants when given in their Latin form. For example, when I tested with Astragalus Trichopodus it heard 'Estrada lustriopodos' with a pretty low confidence as well 0.27822304, the same for acacia pycnantha where it heard 'AKA Kia Picnanfa' with a confidence of 0.36494043. The speech recognition had a hard time understanding fictional characters and made-up words like spells from Harry Potter. Some instances are the spell ‘ebublio’ where it heard 'A Bob Liol' with 0.24632898 confidence and ‘meteolojinx recanto’ where it heard ‘Metereologists, recanto’ with almost 50% confidence (0.42962146). Finally, I tried some elfish names from the Lord of the Rings and some from Game of Thrones and it also misheard such as ‘Islidur’ it heard "It's leidure" with a really low confidence 0.03499997 and ‘Denethor’ and pretty confidently (0.539783) it heard ‘Then I thought’.

I created two different custom speech models, the first one with only providing as data a plain text and one with pronunciation as well. The first model made barely any difference with the vocabulary I tried. More than 80% of the words were still not recognized. Thus, I decided to created the second model which did far better. Nonetheless, something that surprised me was that when speech studio was processing the data it deleted two instances not sure why but my original model was 18 lines and the second 16. The confidence loggings were still a bit low but it was recognising and logging most of the words correctly. For instance, 0.3855385 'Astragalus trichopodus', 0.30037895 'Ebublio'. The elfish names like Eowyn and Islidur were not successful perhaps a stronger model of audio as training would work.

Overall, I think my accent also played a role in the difficulty the model has of recognising some of the words since I am not a native speaker of English. Thus, audio recordings of different accents would be a valuable addition to training the model.