

# Assignment Lab3

## Tuning ASR and TTS

---

### Part A: Hard cases for Speech Recognition

Already in Lab2 I noticed that some people names I added in the grammar were not properly recognized and I tried to take care of that issue by adjusting the grammar to the transcripts I got from the ASR. 'Gerlof', 'Staffan' and 'Sharid' are indeed transcribed as 'Gerloff', 'Stephan' and 'Shahid', with a confidence of 7.7%, 5.7% and 4.6% respectively. But even as 'Guerlov'/'Girl off', 'Stefan' and 'Shahrid'/'Sherid'.

I tested the ASR with some fictional names from Star Wars but the most famous are actually rather well recognized: 'Luke Skywalker', 'Han Solo' and 'Princess Leia' are understood with a confidence of 83%, 86% and 80% respectively. Therefore I tried with some minor characters from the same saga. 'Yaddle', 'Kabe', and 'Nien Nunb' are not picked up but transcribed as 'Yadol'/'Yadol'/'Yadel' with low confidence (4-6%), 'Cabe'/'Caib' (7-12%), 'Nyan Numb'/'Niyan Numb' (6-8%).

The real people names are not recognized because they are not names of English origin and the model is for English. If I chose a model for Swedish and told 'Staffan' it would certainly be easily recognized. The current model does know some non-English words/names (Gerloff, Shahid) but of course not all of them and if so only some version/spelling of them. For the fictional names, it's the same issue of spelling unusual words/names. The low scores even for the recognized misspelling ('Cabe' instead of 'Kabe' for instance) are probably due to the misspelled names not appearing too much in the training data used to build the model.

So to conclude, the first problem is the absence of the desired word in the model's training data and a possible second problem is the low frequency of the recognized misspelled word in the training data.

This speech recognition issue could be solved by adding the desired words to the training data, repeated a sufficient number of times and used in context, for the model to be able to recognize them.

I used Azure Custom Speech, providing plain text data with 66 lines (6 names, 11 lines/sentences by name) to train a customized model. The targeted names are now not only supported but even comfortably recognized:

- Gerlof : 85%
- Staffan : 84%
- Sharid: 60%
- Yaddle : 63%
- Kabe: 68%
- Nien Nunb : 56%

Here is my Endpoint ID: 9fb1d792-feb6-47ef-aac9-ec2e46998109