07/17/2017

**JSALT Rosetta sub-team**

Shruti, Francesco, Lucas, Mark, Florian, and Odette

**Aim/research question**

Can we find the phone set of an under-resourced language through adaptation of an ASR system trained on a resourced language?

*Assumption 1*: we know “the” phoneme inventory of the unknown language (e.g., from Wikipedia, knowing that these are far from the “truth” (because what is the actual phoneme inventory of a language? There will always be disputes about that).

*Assumption 2*: we know (roughly) what language family the under-resourced language belongs to

*Assumption 3*: there is a resourced language that is somewhat related to the under-resourced language (ideally from the same language family); although in principle the setup should work from any resourced language not necessarily one close to the under-resourced language.

**Methodology**

Well-resourced language: Dutch (read speech from CGN)

Under-resourced language: English (FlickR\_8K)

1. Train a DNN-based ASR (in EESEN) on a well-resourced language → Baseline system
2. Compare the phone inventories of the well-resourced and under-resourced language, identify phones from the under-resourced language (= ‘L2 phones’) that are missing from the well-resourced language and identify the closest phone from the well-resourced language (= ‘L1 phone’)
3. Copy the weights from the L1 phones in the Baseline system to create the L2 phones in the DNN; adapt the weights for the L2 phones by visually comparing similar contrasting phone pairs (using visualisation of the DNN hidden layers) that do exist in the well-resourced language using Mark’s already world-famous equation:

eps\_ENG = eps\_DU + Vheight

ash\_ENG = eps\_DU - Vheight

1. Run a free phone recognition pass on the under-resourced language. This will result in a phone alignment of the under-resourced language
2. Extract confidence scores for each sentence in the under-resourced language:
   1. get confidence scores from the ASR
   2. get confidence scores from the speech2image task; trained on the under-resourced language (which is of course not possible when it really is an under-resourced language)
   3. get confidence scores from the speech2translation task; trained on the under-resourced language (which is of course not possible when it really is an under-resourced language)
   4. get confidence scores from Alan’s TTS

→ in principle these steps should only be done to figure out which confidence measure is the ‘best one’; when we move to a ‘proper’ under-resourced language such as Mboshi/Basraa we have to do things differently

1. Retrain the Baseline system:
   1. On all utterances/alignments from step 4)
   2. On only those utterances/alignments with the ‘highest’ confidence score according to 5a, 5b, 5c, 5d, and/or a combination of these

AND train a system from scratch only using the utterances from the under-resourced language

* 1. all utterances/alignments from step 4)
  2. On only those utterances/alignments with the ‘highest’ confidence score according to 5a, 5b, 5c, 5d, and/or a combination of these

1. Evaluation of the Baseline system (before copying the weights) and after each iteration of step 4-6 (= Iter model <nb>)
   1. Retrieval rate on the speech2image task (= phone posteriors to image)
   2. Visualisation of the DNN hidden layers:
      1. Which phones were correctly learned? Compare to the system trained from scratch
      2. How are the phones in the Iter model <nb> spread out compared to the Baseline model and the system trained from scratch?
      3. Compare the phone distribution of the Iter model <nb> with a monolingual model trained on the under-resourced language (supervised training, which is of course not possible in the case of an actual under-resourced language)

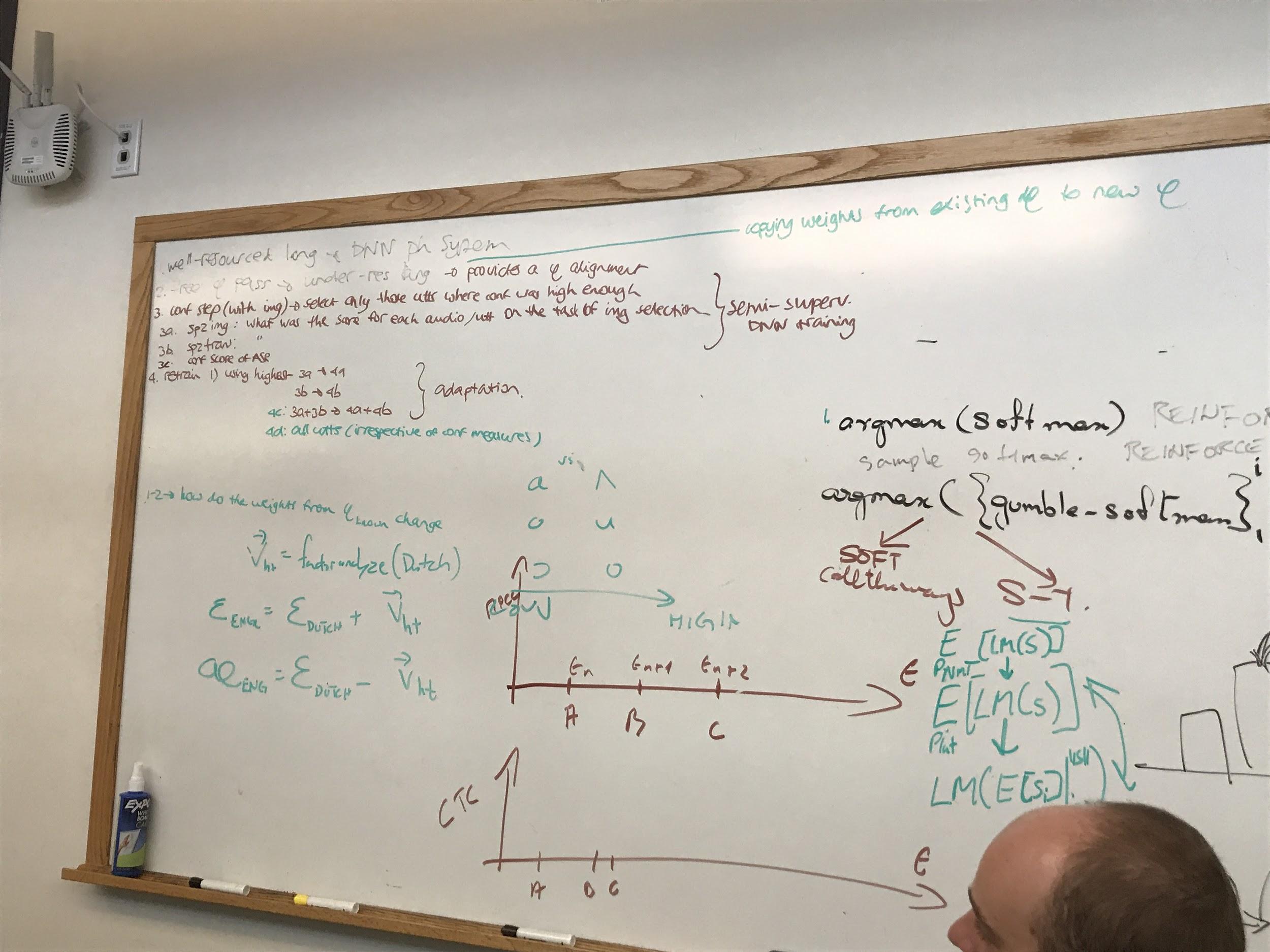
→ This is the reason why we want to carry out this experiment on Dutch and English, to get a feel for what to expect when we move to an actual under-resourced language for which we can’t build a monolingual supervised model

**Repeat process with**

Well-resourced language: English

Under-resourced language: Mboshi/Basaa

**White board discussion on Monday 7/17/17**



**White board discussion on Tuesday 7/18/17**

