

Cross-lingual DNN pretraining for low-resource ASR systems



Edinburgh – Cambridge – Sheffield

Pawel Swietojanski



THE UNIVERSITY
of EDINBURGH

24 April 2012

Major cross-lingual approaches:

- Phone mapping
- SGMMs
- Cross-lingual MLP features (numerous approaches: posteriors, bottleneck, ..., obtained using various NNs structures)

Assumption : (Good) ASR donor systems are required in order to prepare and transfer cross-lingual knowledge.



Question : Can we make **any** use of **any** speech data without **any** knowledge about it and without **any** supporting stuff to improve (low-resource) ASR system ?



Motivation: Pretraining

Idea : Transfer only generatively initialised MLP parameters (here: on cross-lingual speech data).



Experiments. Baseline results.

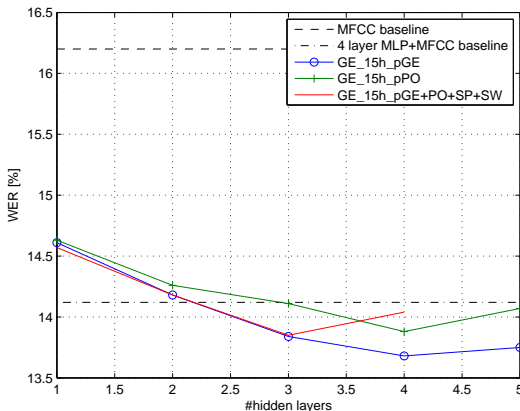
The baseline 4 layer MLP (3 hidden layers) used here as posteriors generator was trained on PLP_E_D_A_Z features forming 9 frames context window.

Dataset	WER [%]		
	MFCC	MLP	MLP+MFCC
GE_15h	16.2	15.79	14.12
GE_5h	18.35	19.47	17.41
GE_1h	23.10	25.68	21.31



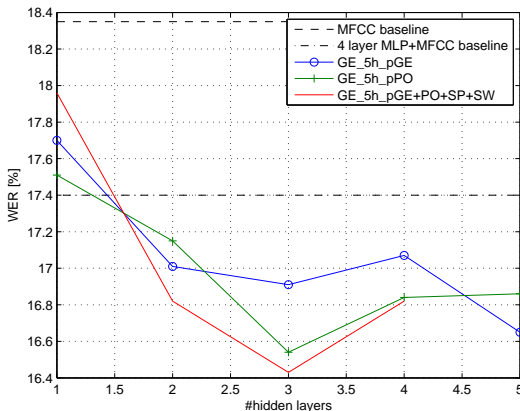
Experiments: Whole training set

GE_Xh_pPO means that the target language is **GE**rman with limited training set to **X hours** and MLP was pretrained on **PO**rtuguese.



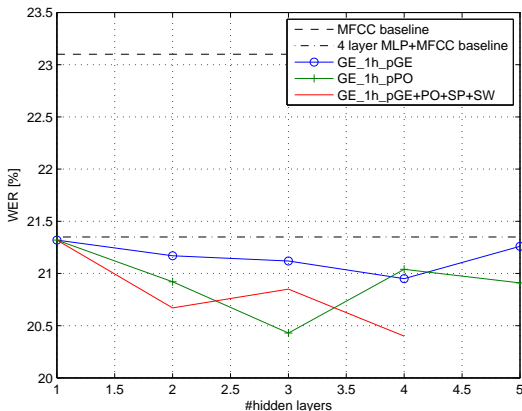
Experiments: Training set limited to 5 hours

GE_Xh_pPO means that the target language is **GE**rman with limited training set to **X hours** and MLP was pretrained on **PO**rtuguese.

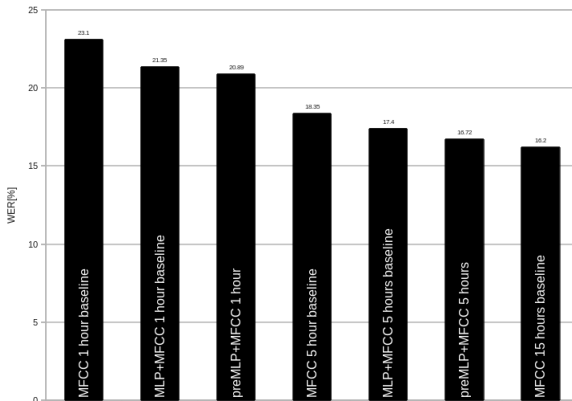


Experiments: Training set limited to 1 hour

GE_Xh_pPO means that the target language is **GE**rman with limited training set to **X hours** and MLP was pretrained on **PO**rtuguese.


 UNIVERSITY
EDINBURGH

Summary



Potential future directions

- Hybrid approach - use DNN to directly model acoustics.
- Fully generative features?

