Context-Dependent for LVCSR: TANDEM, Hybrid of Both?

T. V. Aditya

Abstract

- Gaussian Mixture Model (GMM) and Multi Layer Perceptron (MLP) based acoustic models are compared on a French large vocabulary continuous speech recognition (LVCSR) task.
- However, the best performance is achieved when deep MLP acoustic models are trained on concatenated cepstral and context-dependent bottle-neck features.
- Furtherexperimentsreveal the importance of the neighbouring frames in case of MLP based modeling, and that its gain over GMM acousticmodels is strongly reduced by more complex features.

Introduction and Corpus Description

- Bottle-neck TANDEM
- GMM Based

Table 1: *Training and testing corpora*

	total data [h]	# running words
Train	257	9,800k
Dev10	3.7	41k
Eval10/Dev11	2.9	36k
Eval11	3.1	38k

Experimental setups

- Features
  - Cepstral features
  - TANDEM MLP features
- Acoustic Models
  - GMM-HMM
  - Hybrid MLP-HMM

Results

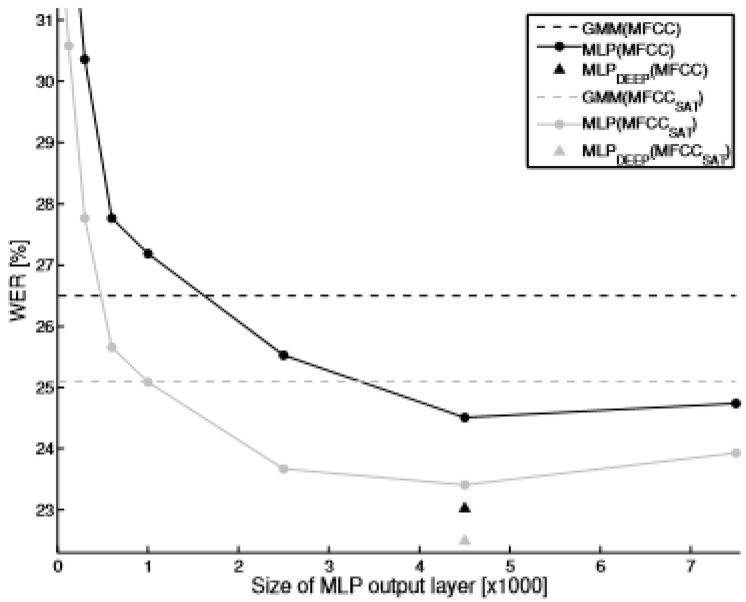

Figure 1: *Effect of the output layer size on Word Error Rate (WER) obtained on Eval10 test set using hybrid MLP acoustic models and cepstral features*

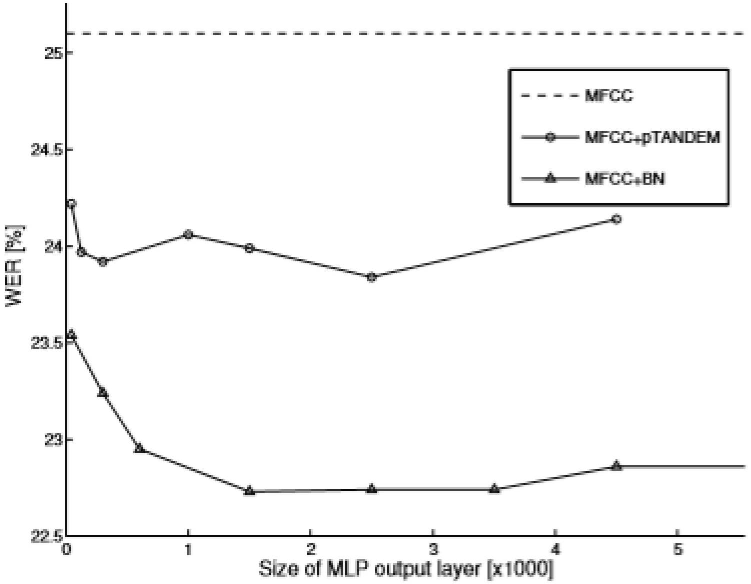

Figure 2: *Effect of the output layer size on Word Error Rate (WER) obtained on Eval10 test set after SAT using concatenated cepstral and probabilistic TANDEM (pTANDEM) or bottle-neck (BN) features*

Results

Table 2: *Comparison of GMM (baseline) and deep MLP based acoustic modeling (AM) with different features. Results are given as word error rate (WER).*

	Test set	Dev10			Eval10			Dev11			Eval11		
	AM	GMM	MLP		GMM	MLP		GMM	MLP		GMM	MLP	
	# input frames	1	1	9	1	1	9	1	1	9	1	1	9
Features	MFCC+Δ+$\Delta\Delta$	27.4	27.3	21.9	29.8	29.3	23.0	28.5	27.8	21.8	26.7	27.2	20.5
	$MFCC_{LDA}$	24.6	23.6	22.1	26.5	25.3	23.2	25.3	24.0	21.9	23.6	22.8	20.8
	$MFCC_{SAT}$	23.8	23.5	22.0	25.1	24.5	22.5	23.8	23.1	21.4	21.6	21.1	19.4
	$(MFCC_{LDA}+BN)_{SAT}$	21.6	21.8	21.4	22.7	22.7	21.9	21.6	21.4	20.6	**19.0**	19.1	**18.4**

CONCLUSIONS

- From the results in Table 2 we can conclude that MLP based acoustic modeling outperforms the GMM based one.
- The difference between the two acoustic modeling method is over 10% relative when linear transformed (derivatives, LDA, CMLLR) MFCC features are applied.
- Since our research was limited to short-term TANDEM features, we intend to carry experiments with long-term features (e.g. MRASTA) and even deeper MLPs, as well.