EE 624 Speech Technology
Report on
Kaldi-Automatic Speech Recognition

- Bharat Mohan
194102304

Objective:
Familiarisation of Kaldi-Automatic Speech Recognition toolkit for Language Modelling and Acoustic Modelling.

Obseravation:

a)Language Model

- Perplexity values of the following N grams were computed and the corresponding log likelihoods have been tabulated

N	Perplexity	Log likelihood
1	730.078	-18866.7
2	424.19	-17312.5
3	414.171	-17244.6
4	414.044	-17243.7
5	414.12	-17244.2
6	414.12	-17244.2
15	414.12	-17244.2

- Its observed that the perplexity values decrease till N = 4 and then saturates at 414.12 with further increase in N.
- Significant improvement in perplexity from Language Unigram to Language Bigram
- Backoff factors are created for unseen contexts, for n<N

b)Acoustic Model

- Word Error Rates computed on the decoded development data by the Acoustic Models which were trained and tested using Bigram Language model are

tabluated. Monophone, Triphone and DNN based Models and their corresponding WER are tabulated.

Acoustic Model	WER(%)
Monophone	67
Triphone	48
DNN based(hybrid)	45

- Average Log likelihood values per frame for the test data (in G6) are also tabulated for each Acoustic Model.

Acoustic Model	Log Likelihood per frame
Monophone	-8.55418
Triphone	-8.44221
DNN based	0.166746

Conclusions:

- The performance of Bi-gram or at the most Tri-gram Language models seem adequate for ASR. The complexity increases and the performance doesn't increase significantly for higher- N gram Language models.
- The DNN based Hybrid acoustic model outperforms Mono-phone and Tri-phone acoustic models.