# 

# Hammad Abdullah

* [14bscshabdullah@seecs.edu.pk](mailto:14bscshabdullah@seecs.edu.pk)
* Pakistan, Islamabad
* National University of Science and Technology(NUST) 2014-2018
* IRC nick: hbd
* Subscribed to Mailing lists
* +92 311 5194032
* [Linkedin](http://linkedin.com/in/hammad-abdullah)
* [github](http://github.com/habdullah)

MLP incorporation in CMUsphinx-GSoC 2017 proposal

**April 3rd, 2017**

# 

# Description

In this project we aim to incorporate MLP acoustic model in sphinx. The expected result is best-in-category accuracy in TED talk recognition.Sphinx traditionally uses a GMM model for training and the accuracy is very reasonable.But it can be improved with the addition of a MLP. So our new model will be a MLP-HMM interface . The addition of MLP-HMM to the Sphinx models will increase accuracy of sphinx systems and will have a significant effect on speech recognition worldwide

# Why I chose this project

I worked with the kaldi ASR toolkit in one of my projects. It had its pros and cons. It was powerful (DNNs) but not portable. CMUsphinx is portable but does not have DNN models.I feel I can contribute to global speech recognition efforts substantially if I can incorporate a DNN model in sphinx. Many people use sphinx for their speech recognition projects and it would be nice to give them an updated toolbox.

# What needs to be accomplished

HMM work like this

p( W | X ) = p ( W ) \* ∑ [ p( s , s’) \* p( xs | s ) ]

The probability highlighted in red will be now be approximated by a residual neural network which will use MLPs instead of convolutional neural networks.

p( W | X ) = p ( W ) \* ∑ [ p( s , s’) \* p( xs | s ) ]

### The architecture

A ResNet with MLPs instead of convolutional Networks.Each MLP will have 2-3\* hidden layers. It will have identity shortcut mapping and identity after-add mapping.

## 

## 

### Training deep architecture

ResNets are very easy to train because the forward and backward props are very smooth due to the Identity shortcut & after add mappings .They also exhibit very low training and test errors. We will train feature to phoneme (feat2ph) alignments using TEDLIUM corpus. The MLP will be discriminatively pretrained and then trained using these alignments.

### Recognition

The input audio will be time sliced.25ms time slices are normally used.To calculate probability of a frame, the preceding and succeeding 4-5 frames will be concatenated and the features computed.40 dimensional log spectrum features will be extracted using filter banks.MLP will produce state probabilities taking mel log spectrum features as input which will be feeded to the HMM.HMM will generate Phone probability. The language models will not be changed.

### Increasing generalization ability

To improve generalization we will use *dropout* and *batch normalizatio*n.

In dropout, each time we input a training example we will randomly omit hidden units with a probability of 0.5. At test time we half the outgoing weights of all the the hidden units.although this will increase our training time and require more hidden units, this will decrease errors substantially.

In batch normalization, we will normalize the training input and then feed it into this function where are parameters which will be learned through back prop.

### SGD optimization

To speed up gradient descent, RMS prop and adagrad algorithms will be used.

# 

# Study Material Consulted

* Samuel Thomas, Patrick Nguyen, Geoffrey Zweig and Hynek Hermansky, MLP based Phoneme Detectors for Automatic Speech Recognition, ICASSP, Prague, Czech Republic, May 2011([pdf](https://www.clsp.jhu.edu/~samuel/pdfs/scarf_mlp.pdf))
* Geoffrey Hinton (2012) Deep Neural Networks for Acoustic Modeling in Speech Recognition ([PDF](https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/38131.pdf))
* CMU ASR [slides](http://asr.cs.cmu.edu/spring2014/) spring 2014

# Goals

### Get TEDLIUM feat2ph alignments

### Get trained MLP classifier

### MLP-HMM expected accuracy

# Milestones

1. Benchmark memory util and performance of sphinx4 on forced alignment for resource planning purposes
2. Solve [tutorial](http://cmusphinx.sourceforge.net/wiki/tutorialam) on sphinx webpage
3. Study ResNet implementation details
4. Write and test function to implement concatenated audio frames -> features.
5. Find number of hidden layers suitable for our model.
6. Design ResNet class
7. Get feat2ph alignments for TEDLIUM using existing methods.
8. Write and test feed forward function.
9. Write and test back propagation function.
10. Write and test method for Discriminative pretraining of hidden layers.
11. Train MLP classifier.
12. Write and test function for MLP-HMM interface.
13. Write and test, result compiling java file
14. Get WER on test audio.
15. Finetune MLP.
16. Port MLP java implementation to C using a cross compiler.
17. Write initial project documentation.
18. Proof-read project documentation with mentor review.
19. Prepare a research report on the project and submit it to a journal after mentor review.
20. (Ongoing after summer) Reply to Questions and comments on Github repo

# Planning Schedule

1. Prior to May 4: milestones 1,2,3

***COMMUNITY BONDING PERIOD STARTS***

1. May 5 - May 15: milestones 4,5
2. May 16 - May 31: milestones 7,6

***CODING PERIOD STARTS***

1. June 1 - June 9: 3 hr/weekdays , 5hr/weekend (due to school).
2. June 10 - June 16: IN ACTIVE
3. June 17 - June 22: milestones 8,9
4. June 23 - June 28: BUFFER
5. June 29 - July 4 : milestones 10,11
6. July 5 - July 10 : BUFFER
7. July 11 - July 16: milestone 12,13
8. July 17 - July 22: milestone 14
9. July 23 - July 28: milestone 15
10. July 29 - August 3 : BUFFER
11. August 4 - August 9 :milestone 16
12. August 10 - August 15: milestone 17
13. August 16 - August 21: milestone 18
14. August 22 - August 29: BUFFER

# Mentors

Dr. Bhiksha Raj

Dr. Nickolay Shmyrev

# Deliverables

Java files with :

* Feature generation methods
* MLP deterministic pre-training methods
* ResNet training methods
* ResNet test methods
* MLP-HMM recognizer methods

# Experience

A big part of my decision to work on this project is because I have relevant experience. I’ve completed Andrew Ng’s machine learning course on coursera and also an AI course during the normal semester.

I finished a Speech recognition course during summer of 16. I worked with the Technische Universität Kaiserslautern TUKL research lab in my department with my professor whose PhD is in ASR and NLP . I was declared best researcher at the end of summer . At the same time I was working on Urdu ASR as a research intern. I continue to work on urdu ASR as a research assistant. My toolkit of choice was kaldi. I familiarized myself with the code base and the configuration of kaldi during this period and since Kaldi and sphinx are similar I will have no problem working with sphinx’s codebase.

# Additional Details

### Is this your first contact with the CMUSphinx project?

No

### We expect you to work on the project 30 hours per week. Are you ready for that?

Ofcourse

### Do you have any commitments during the summer?

None

### Exams or other events you expect to have to deal with during the GSOC period

My finals are from june 1 to june 16 so I won’t be able to work on the project during that time. I have designed my planning schedule accordingly.

### Will you have an Internet access during the summer?

Yes

### We expect you to blog about project success each week. Are you ready for that?

Yes