1 What I'm doing & Why it's Extremely Cool

I'm creating Neural Nets which better classify data from unseen conditions, without any explicit adaptation of model parameters or data transformations.

My approach is extremely cool because it doesn't require tons of data, it is not specific to some dataset, and it can be used to build any Neural Net (not just for speech recognition).

2 Overview of Speech Recognition

This section contains an overview of the training and testing procedures for standard automatic speech recognition (ASR) pipelines. The overview will provide the reader with a technical grounding in ASR, so that the rest of the dissertation will have some point of reference.

- **Gaussians + HMMs**
- **Neural Nets + HMMs**
- **end-to-end Neural Nets**

3 Background Literature

Here I will cover the literature relevant to working with small (or completely new) datasets. There are two main approaches, (1) adapt a model from one training dataset to a new, smaller dataset; (2) create a model that is robust enough to handle data from multiple domains.

- **Model Adaptation: (e.g. Speaker; Language)**
- **Model Robustness: (e.g. Noise; Channel)**

4 Experiments

This section contains the main contributions of the dissertation research.

This dissertation investigates training methods for acoustic modeling in the Neural Net + HMM ASR pipeline.

I aim to produce acoustic models which perform better (i.e. lower Word Error Rates) on datasets which are not similar the original training dataset.

I investiage the effectiveness of different tasks (eg. linguistic tasks vs machine learning tasks) in a Multi-task Learning framework.

4.1 Data

The differences between the training and testing data will be (1) the recording noise conditions, (2) who the speaker is, or (3) what language the speaker is using. The following table shows which data sets are used for each audio condition.

		Corpus	
		Train	**Test**
Audio Condition	**Noise**	TIDIGITS	Aurora 5
	Speaker	LibriSpeech-A	LibriSpeech-B
	Language	LibriSpeech	Kyrgyz Audiobook

Table 1: Speech Corpora

4.2 Model Training Procedure

I investigate methods of neural net training with Multi-Task Learning (MTL).

Typically, the tasks used in MTL are assumed to exist and be useful in some other application (e.g. POS tagging as an additional task to learn dependency parsing). However, a less common approach to MTL training is to create completely new tasks.

This dissertation investigates the creation of new tasks for MTL, either using (1) linguist-expert knowledge, (2) ASR Engineer-expert knowledge, or (3) general Machine Learning knowledge.

The latter two knowledge sources are useful for buidling acoustic models, but not much else. On the other hand, the final knowledge source (general machine learning concepts) can be applied to *any* classification problem.

The three knowledge sources will be abbreviated as such:

- (LING) **Linguistic Knowledge**
- (ASR) **Traditional Speech Recognition Pipeline**
- (ML) **General Machine Learning**

Each of these categories contains a wealth of ideas, but I will consolidate each into three experiments. With three experiments for each knowledge source, my dissertation will contain nine (9) experimental conditions (for each audio condition).

Specifically, I will use the following concepts to create new tasks to be used in MTL training:

	Knowledge Source		
	LING	**ASR**	**ML**
Experiments	voicing	monophones	k-means
	place	1/2 triphones	random forests
	manner	triphone variations	bootstrapped resamples

Table 2: Experimental Setup

Each of these tasks will be added to a Baseline model. More specifically, the Baseline model will be a Neural Net with a single output layer (Task A), and the tasks above will be added as a second task (Task B). You can think of the tasks as simply a new set of labels for the existing data set. For example, when the Ling task of voicing is used, any audio segment labeled `[b]` will be assigned the new label `voiced`.

When these experiments will be applied to each of the three audio conditions, we get the following 30 experiments:

Data Condition	**Train Data**	**Test Data**	**MTL Training Tasks**	**Num. Exps**
Noise	TIDIGITS	Aurora 5	Basline Baseline + LING Baseline + ASR Baseline + ML	1 3 3 3
Speaker	LibriSpeech-A	LibriSpeech-B	Baseline Baseline + LING Baseline + ASR Baseline + ML	1 3 3 3
Language	LibriSpeech + Kyrgyz-A	Kyrgyz-B	Baseline Baseline + LING Baseline + ASR Baseline + ML	1 3 3 3
				30

Table 3: Experimental Setup