Sharing Acoustic Models Between Languages

Partha Lal

CSTR

27th November 2006/ Thesis Proposal Talk
Few languages have ASR training data

- Speech recognition systems require large amounts of labelled data but such data exists for only a few languages.
- How can we perform recognition on languages with little or no speech data?
- Can we improve performance on languages with medium amounts of data?
- Maybe we could use existing data in other languages?
Few languages have ASR training data

- Speech recognition systems require large amounts of labelled data but such data exists for only a few languages.
- How can we perform recognition on languages with little or no speech data?
- Can we improve performance on languages with medium amounts of data?
- Maybe we could use existing data in other languages?
Few languages have ASR training data

- Speech recognition systems require large amounts of labelled data but such data exists for only a few languages.
- How can we perform recognition on languages with little or no speech data?
- Can we improve performance on languages with medium amounts of data?
- Maybe we could use existing data in other languages?
Few languages have ASR training data

- Speech recognition systems require large amounts of labelled data but such data exists for only a few languages.
- How can we perform recognition on languages with little or no speech data?
- Can we improve performance on languages with medium amounts of data?
- Maybe we could use existing data in other languages?
How can we use data from other languages?

Knowledge from one language could be encapsulated in a trained classifier.

It could then be applied to another language.
How can we use data from other languages?

- Knowledge from one language could be encapsulated in a trained classifier.
- It could then be applied to another language.
What classes should the classifier use?

- From which set of classes should $p$ be?
- Will this vary between languages?
What classes should the classifier use?

- From which set of classes should $p$ be?
- Will this vary between languages?
Outline

1. Background
   - Applications
   - Sub-word Units
   - Direct vs. Indirect Modelling
2. Direct Models
   - Phoneme
   - Grapheme
   - Articulatory Feature
3. Indirect Models
   - Tandem Features
   - Experiments
4. MLP Modifications
   - New Output Layer
   - Multi-task Learning
Outline

1 Background
   - Applications
   - Sub-word Units
   - Direct vs. Indirect Modelling

2 Direct Models
   - Phoneme
   - Grapheme
   - Articulatory Feature

3 Indirect Models
   - Tandem Features
   - Experiments

4 MLP Modifications
   - New Output Layer
   - Multi-task Learning
Train on A, recognise X

- A is a language with lots of data
- Language X has little data
- The trained model is being used cross-lingually
- Source languages similar to the target language are expected to work best
Train on A, recognise X

- A is a language with lots of data
- Language X has little data
- The trained model is being used cross-lingually
- Source languages similar to the target language are expected to work best
Train on $A$, recognise $X$

- $A$ is a language with lots of data
- Language $X$ has little data
- The trained model is being used cross-lingually
- Source languages similar to the target language are expected to work best
Train on $A$, recognise $X$

- $A$ is a language with lots of data
- Language $X$ has little data
- The trained model is being used cross-lingually
- Source languages similar to the target language are expected to work best
Train on \( \{A, B, C\} \), recognise \( X \)

- **Language \( X \) has little data**
- The source languages have 10–100 hours of data each
- The trained model is again being used cross-lingually
- Choice of \( \{A, B, C\} \) …
  - Similar languages vs. different, diverse languages
- Vary amount of data from each language
Train on \( \{A, B, C\} \), recognise \( X \)

- Language \( X \) has little data
- The source languages have 10–100 hours data each
- The trained model is again being used cross-lingually
- Choice of \( \{A, B, C\} \)...
  - Similar languages vs. different, diverse languages
- Vary amount of data from each language
Train on \( \{A, B, C\} \), recognise \( X \)

- Language \( X \) has little data
- The source languages have 10–100 hours data each
- The trained model is again being used cross-lingually
- Choice of \( \{A, B, C\} \)...
  - Similar languages vs. different, diverse languages
- Vary amount of data from each language
Train on \( \{A, B, C\} \), recognise \( X \)

- Language \( X \) has little data
- The source languages have 10–100 hours data each
- The trained model is again being used cross-lingually
- Choice of \( \{A, B, C\} \)
  - Similar languages vs. different, diverse languages
- Vary amount of data from each language
Train on \( \{A, B, C\} \), recognise \( X \)

- Language \( X \) has little data
- The source languages have 10–100 hours data each
- The trained model is again being used cross-lingually
- Choice of \( \{A, B, C\} \)...
  - Similar languages vs. different, diverse languages
- Vary amount of data from each language
Train on \( \{A, B, C\} \), recognise \( X \)

- Language \( X \) has little data
- The source languages have 10–100 hours data each
- The trained model is again being used cross-lingually
- Choice of \( \{A, B, C\} \)...
  - Similar languages vs. different, diverse languages
- Vary amount of data from each language
Train on \( \{A, B, C, X\} \), recognise \( X \)

- As the previous scenario, but the target language is also one of the source languages
  - Language \( X \) could again have little data
  - Or we could be improving on an already well trained model
  - This is a multilingual case
  - Vary amount of data from each language
Train on \( \{A, B, C, X\} \), recognise \( X \)

- As the previous scenario, but the target language is also one of the source languages
- Language \( X \) could again have little data
- Or we could be improving on an already well trained model
- This is a multilingual case
- Vary amount of data from each language
Train on \( \{A, B, C, X\} \), recognise \( X \)

- As the previous scenario, but the target language is also one of the source languages
- Language \( X \) could again have little data
- Or we could be improving on an already well trained model
- This is a multilingual case
- Vary amount of data from each language
Train on \( \{A, B, C, X\} \), recognise \( X \)

- As the previous scenario, but the target language is also one of the source languages
- Language \( X \) could again have little data
- Or we could be improving on an already well trained model
- This is a multilingual case
- Vary amount of data from each language
Train on \( \{A, B, C, X\} \), recognise \( X \)

- As the previous scenario, but the target language is also one of the source languages
- Language \( X \) could again have little data
- Or we could be improving on an already well trained model
- This is a multilingual case
- Vary amount of data from each language
Sub-word Units
Desirable Properties for Speech Units

Sound the same across languages

- Having this property would mean models trained in one language can be easily transferred to another.
Desirable Properties for Speech Units
Evenly distributed across languages

- Units in the target language may be under-represented in the source languages
Desirable Properties for Speech Units

Easily derived

- Speech unit labels are derived from a pronunciation dictionary
- Producing pronunciation dictionaries for new languages is an expensive task
  - Also, dictionary pronunciations are rarely used in conversational speech
Desirable Properties for Speech Units

Easily derived

- Speech unit labels are derived from a pronunciation dictionary
- Producing pronunciation dictionaries for new languages is an expensive task
  - Also, dictionary pronunciations are rarely used in conversational speech
Desirable Properties for Speech Units

Easily derived

- Speech unit labels are derived from a pronunciation dictionary
- Producing pronunciation dictionaries for new languages is an expensive task
  - Also, dictionary pronunciations are rarely used in conversational speech
Phonemes

CONSONANTS (PULMONIC)

<table>
<thead>
<tr>
<th></th>
<th>Biliteral</th>
<th>Labiodental</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Postalveolar</th>
<th>Retrals</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Pharyngeal</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>P b</td>
<td>t d</td>
<td>t q</td>
<td>c j</td>
<td>k g</td>
<td>q g</td>
<td>?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>n l</td>
<td>n</td>
<td>n l</td>
<td>n</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td>B r</td>
<td>r</td>
<td>r l</td>
<td>r</td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>f v</td>
<td>ð s ð</td>
<td>s z</td>
<td>s z</td>
<td>ç j</td>
<td>ð x</td>
<td>x ñ</td>
<td>h ñ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral Fricative</td>
<td>H k</td>
<td>H k</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>v j</td>
<td>j</td>
<td>j</td>
<td>j</td>
<td>u</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral Approximant</td>
<td>l l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

VOWELS

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.
Phonemes

- Phonemes that are nominally the same are produced differently in different languages
- Phonemes are differently distributed across languages
- Pronunciation dictionaries need to be produced for the target language
Phonemes

- Phonemes that are nominally the same are produced differently in different languages.
- Phonemes are differently distributed across languages.
- Pronunciation dictionaries need to be produced for the target language.
Phonemes

- Phonemes that are nominally the same are produced differently in different languages
- Phonemes are differently distributed across languages
- Pronunciation dictionaries need to be produced for the target language
Graphemes

- Graphemes are definitely produced differently in different languages
- Graphemes are distributed differently across languages
- Logographic languages pose a challenge
- Avoids the (error-prone) task of pronunciation modelling
Graphemes

- Graphemes are definitely produced differently in different languages
- Graphemes are distributed differently across languages
  - Logographic languages pose a challenge
  - Avoids the (error-prone) task of pronunciation modelling
Graphemes

- Graphemes are definitely produced differently in different languages
- Graphemes are distributed differently across languages
- Logographic languages pose a challenge
- Avoids the (error-prone) task of pronunciation modelling
Graphemes

- Graphemes are definitely produced differently in different languages
- Graphemes are distributed differently across languages
- Logographic languages pose a challenge
- Avoids the (error-prone) task of pronunciation modelling
# Articulatory Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place</td>
<td>labial, labio-dental, dental...</td>
</tr>
<tr>
<td>Manner</td>
<td>vowel, approximant, flap...</td>
</tr>
<tr>
<td>Nasality</td>
<td>+, -, silence</td>
</tr>
<tr>
<td>Glottal state</td>
<td>voiced, voiceless, aspirated...</td>
</tr>
<tr>
<td>Rounding</td>
<td>+, -, silence</td>
</tr>
<tr>
<td>Vowel</td>
<td>aa, ae, ah, ao, aw1, aw2, ax...</td>
</tr>
<tr>
<td>Height</td>
<td>very high, high, mid-high...</td>
</tr>
<tr>
<td>Frontness</td>
<td>back, mid-back, mid...</td>
</tr>
</tbody>
</table>
Articulatory Features

- Lower-level, more language-independent mapping to acoustics
- Some difference in distribution but less so than other units
Articulatory Features

- Lower-level, more language-independent mapping to acoustics
- Some difference in distribution but less so than other units
How can we use the classifier outputs?

- Given class posteriors, we could use them
  - **Directly**  Hybrid ANN/HMM models
    - Sub-word units used by the classifier are used in recognition
    - Can work without data in target language
  - **Indirectly**  Tandem features
    - Classifier output is post-processed so that the classifier’s classes and the recognition units can be different
    - Choice of classes is less important
How can we use the classifier outputs?

- Given class posteriors, we could use them
  
  **Directly** Hybrid ANN/HMM models
  
  - Sub-word units used by the classifier are used in recognition
  - Can work without data in target language

- Indirectly Tandem features
  
  - Classifier output is post-processed so that the classifier’s classes and the recognition units can be different
  - Choice of classes is less important
How can we use the classifier outputs?

- Given class posteriors, we could use them
  - **Directly** Hybrid ANN/HMM models
    - Sub-word units used by the classifier are used in recognition
    - Can work without data in target language
  - **Indirectly** Tandem features
    - Classifier output is post-processed so that the classifier's classes and the recognition units can be different
    - Choice of classes is less important
How can we use the classifier outputs?

- Given class posteriors, we could use them
  - **Directly** Hybrid ANN/HMM models
    - Sub-word units used by the classifier are used in recognition
    - Can work without data in target language
  - **Indirectly** Tandem features
    - Classifier output is post-processed so that the classifier’s classes and the recognition units can be different
    - Choice of classes is less important
How can we use the classifier outputs?

- Given class posteriors, we could use them
  - **Directly** Hybrid ANN/HMM models
    - Sub-word units used by the classifier are used in recognition
    - Can work without data in target language
  - **Indirectly** Tandem features
    - Classifier output is post-processed so that the classifier’s classes and the recognition units can be different
    - Choice of classes is less important
How can we use the classifier outputs?

- Given class posteriors, we could use them
  
  **Directly** Hybrid ANN/HMM models
  
  - Sub-word units used by the classifier are used in recognition
  - Can work without data in target language

- Indirectly Tandem features
  
  - Classifier output is post-processed so that the classifier’s classes and the recognition units can be different
  - Choice of classes is less important
How can we use the classifier outputs?

- Given class posteriors, we could use them
  - **Directly** Hybrid ANN/HMM models
    - Sub-word units used by the classifier are used in recognition
    - Can work without data in target language
  - **Indirectly** Tandem features
    - Classifier output is post-processed so that the classifier’s classes and the recognition units can be different
    - Choice of classes is less important
This work concentrates on acoustic modelling.

It doesn’t look at other areas, e.g. language modelling.
This work concentrates on acoustic modelling

It *doesn’t* look at other areas, e.g. language modelling
Outline

1. Background
   - Applications
   - Sub-word Units
   - Direct vs. Indirect Modelling

2. Direct Models
   - Phoneme
   - Grapheme
   - Articulatory Feature

3. Indirect Models
   - Tandem Features
   - Experiments

4. MLP Modifications
   - New Output Layer
   - Multi-task Learning

Sharing Acoustic Models Between Languages
Baseline Monolingual Triphone HMM

- Serves as a baseline and
  - provides forced alignments
  - provides lattices
Baseline Monolingual Triphone HMM

- Serves as a baseline and
  - provides forced alignments
  - provides lattices

X → mixture of Gaussians → O
Baseline Monolingual Triphone HMM

- Serves as a baseline and
  - provides forced alignments
  - provides lattices

Baseline Monolingual Triphone HMM with a mixture of Gaussians.
Given a shared set of phonemes, conventional triphone HMM systems will be built.

For phonemes in the target language that don’t appear in the source language(s) tying or mapping schemes will be used.
Baseline Triphone HMM

- Given a shared set of phonemes, conventional triphone HMM systems will be built.
- For phonemes in the target language that don’t appear in the source language(s) tying or mapping schemes will be used.
Neural Networks are trained to classify acoustic observations

ANN Posteriors are then used to provide observation likelihoods

\[ P(o_t|x_t) \propto \frac{P(C=c|o_t)}{P(C)} \]
**Hybrid ANN/HMM**

- Neural Networks are trained to classify acoustic observations.
- ANN Posteriors are then used to provide observation likelihoods:
  \[ P(o_t|x_t) \propto \frac{P(C = c|o_t)}{P(C)} \]
Advantages

- Uses a **wider context**
- MLPs are trained **discriminatively** (unlike GMMs)
- MLPs can be trained on different data
- No Gaussians used — the model has a smaller, fixed size
Advantages

- Uses a **wider context**
- MLPs are trained **discriminatively** (unlike GMMs)
- MLPs can be trained on different data
- No Gaussians used — the model has a smaller, fixed size
Advantages

- Uses a wider context
- MLPs are trained discriminatively (unlike GMMs)
- MLPs can be trained on different data
- No Gaussians used — the model has a smaller, fixed size
Advantages

- Uses a wider context
- MLPs are trained **discriminatively** (unlike GMMs)
- MLPs can be trained on different data
- No Gaussians used — the model has a smaller, fixed size
Standard Hybrid

- MLP uses the same phonemes as the recognition model

Output: 1

Input: x
Hybrid with Mismatched Phoneme Sets

MLP uses source language phonemes
- Mapping to target language phonemes applied
- Mapping may be context-sensitive
Hybrid with Mismatched Phoneme Sets

- MLP uses source language phonemes
- Mapping to target language phonemes applied
- Mapping may be context-sensitive

 VE about a from MLP

mapping between phoneme sets

x

a

=1
Hybrid with Mismatched Phoneme Sets

- MLP uses source language phonemes
- Mapping to target language phonemes applied
- Mapping may be context-sensitive

Partha Lal
Sharing Acoustic Models Between Languages
Hybrid with Multiple Mismatched Phoneme Sets

- MLPs use source language phonemes
- Mappings to target language phonemes applied
Hybrid with Multiple Mismatched Phoneme Sets

- MLPs use source language phonemes
- Mappings to target language phonemes applied
Hybrid with Mismatched Grapheme Sets

- MLP uses source language graphemes
- Mapping to target language graphemes applied

Diagram:

1. \( x \)
2. Mapping between grapheme sets
3. \( a \)
4. VE about \( a \) from MLP
5. \( =1 \)
Hybrid with Mismatched Grapheme Sets

MLP uses source language graphemes
- Mapping to target language graphemes applied

$x \rightarrow a \rightarrow \text{mapping between grapheme sets} \rightarrow \text{VE about } a \text{ from MLP} \rightarrow =1$

Partha Lal
Sharing Acoustic Models Between Languages
Hybrid with Multiple Mismatched Grapheme Sets

MLPs use source language graphemes

Mappings to target language graphemes applied
Hybrid with Multiple Mismatched Grapheme Sets

MLPs use source language graphemes
Mappings to target language graphemes applied
The mapping between AFs and phoneme state $x$ can be:

- manually specified using linguistic knowledge
- learnt from data

Implemented with monolingual data at CLSP Workshop 2006
The mapping between AFs and phoneme state $x$ can be
- manually specified using linguistic knowledge
- learnt from data

Implemented with monolingual data at CLSP Workshop 2006
AF-based Hybrid

- The mapping between AFs and phoneme state $x$ can be
  - manually specified using linguistic knowledge
  - learnt from data

- Implemented with monolingual data at CLSP Workshop 2006

**Background**
- Direct Models
- Indirect Models
- MLP Modifications
**Summary**

**Articulatory Feature**

**Phoneme**

**Grapheme**

**MLP Modifications**
AF-based Hybrid

- The mapping between AFs and phoneme state $x$ can be
  - manually specified using linguistic knowledge
  - learnt from data

- Implemented with monolingual data at CLSP Workshop 2006
AF labels are initially derived through forced alignment
After that, the Hybrid system can be used to produce new labels for the data
That can be iterated...
Embedded Training

- AF labels are initially derived through forced alignment
- After that, the Hybrid system can be used to produce new labels for the data
- That can be iterated...
Embedded Training

- AF labels are initially derived through forced alignment
- After that, the Hybrid system can be used to produce new labels for the data
- That can be iterated...
Outline

1 Background
   - Applications
   - Sub-word Units
   - Direct vs. Indirect Modelling

2 Direct Models
   - Phoneme
   - Grapheme
   - Articulatory Feature

3 Indirect Models
   - Tandem Features
   - Experiments

4 MLP Modifications
   - New Output Layer
   - Multi-task Learning

Partha Lal
Sharing Acoustic Models Between Languages
Indirect Models

- The speech units used don’t appear explicitly in the model structure
- The output of MLPs are post-processed such that they can be modelled alongside acoustic features with a GMM
- Units used in the MLP can differ from those in the HMM
- Units from different languages can be used together
Indirect Models

- The speech units used don’t appear explicitly in the model structure.
- The output of MLPs are post-processed such that they can be modelled alongside acoustic features with a GMM.
- Units used in the MLP can differ from those in the HMM.
- Units from different languages can be used together.
Indirect Models

- The speech units used don’t appear explicitly in the model structure
- The output of MLPs are post-processed such that they can be modelled alongside acoustic features with a GMM
- Units used in the MLP can differ from those in the HMM
- Units from different languages can be used together
Indirect Models

- The speech units used don’t appear explicitly in the model structure.
- The output of MLPs are post-processed such that they can be modelled alongside acoustic features with a GMM.
- Units used in the MLP can differ from those in the HMM.
- Units from different languages can be used together.
A window of acoustic observations is input to the MLP and the output activations are then computed.

$$MLP(O_{PLP})$$
Extracting Tandem Features
Make more Gaussian

Since the MLP outputs will later be modelled using Gaussian distributions, logs are taken to make the distribution more Gaussian.

$$\log(MLP(O_{PLP}))$$
Principal Components Analysis is used to decorrelate the log MLP outputs.

\[ PCA(\log(MLP(O_{PLP}))) \]
Extracting Tandem Features
Combine with Acoustic Features

- The massaged MLP output vector is now concatenated with a vector of conventional acoustic observations.
- Different, complementary features could be used here:

\[ O_{MLP} = PCA(\log(MLP(O_{PLP}))) \]
\[ O = O_{MLP} + O_{MFCC} \]
The massaged MLP output vector is now concatenated with a vector of conventional acoustic observations. Different, complementary features could be used here.

\[ O_{MLP} = \text{PCA}(\log(\text{MLP}(O_{PLP}))) \]
\[ O = O_{MLP} + O_{MFCC} \]
Extracting Tandem Features
Model with a GMM

- The resulting vector is fed to a conventional HMM using mixture-of-Gaussians output densities and recognition is performed.

- This step needs at least some data in the target language.
The resulting vector is fed to a conventional HMM using mixture-of-Gaussians output densities and recognition is performed.

This step needs at least some data in the target language.
Phoneme Tandem

The MLP posteriors can be

- Phoneme posteriors from one language
- From multiple, language-specific MLPs combined as posteriors
- From multiple, language-specific MLPs combined as tandem features
The MLP posteriors can be
- Phoneme posteriors from one language
- From multiple, language-specific MLPs combined as posteriors
- From multiple, language-specific MLPs combined as tandem features
The MLP posteriors can be

- Phoneme posteriors from one language
- From multiple, language-specific MLPs combined as posteriors
- From multiple, language-specific MLPs combined as tandem features
Factored Tandem

- $o_{MLP}$ and $o_{MFCC}$ could be weighted differently
- No common covariance parameter
- Separate parameter tying can be done with each observation stream
- Factored Tandem has been shown to perform better than concatenated
Factored Tandem

- $o_{MLP}$ and $o_{MFCC}$ could be weighted differently
- No common covariance parameter
- Separate parameter tying can be done with each observation stream
- Factored Tandem has been shown to perform better than concatenated
Factored Tandem

- $o_{MLP}$ and $o_{MFCC}$ could be weighted differently
- No common covariance parameter
- Separate parameter tying can be done with each observation stream
- Factored Tandem has been shown to perform better than concatenated

Factored Tandem

mixture of Gaussians

MLP

MFCC

Partha Lal
Sharing Acoustic Models Between Languages
Factored Tandem

- $o_{MLP}$ and $o_{MFCC}$ could be weighted differently
- No common covariance parameter
- Separate parameter tying can be done with each observation stream
- Factored Tandem has been shown to perform better than concatenated
Grapheme Tandem

- \(X\) connection
- Mixture of Gaussians
- MLP
- MFCC

Partha Lal
Sharing Acoustic Models Between Languages
AF Tandem

- Shown to work at least as well as phoneme-based Tandem
- Separate parameter tying can be done with each observation stream
AF Tandem

- Shown to work at least as well as phoneme-based Tandem
- Separate parameter tying can be done with each observation stream
Outline

1. Background
   - Applications
   - Sub-word Units
   - Direct vs. Indirect Modelling

2. Direct Models
   - Phoneme
   - Grapheme
   - Articulatory Feature

3. Indirect Models
   - Tandem Features
   - Experiments

4. MLP Modifications
   - New Output Layer
   - Multi-task Learning
Replace and Re-train the Output Layer

- Train on the source language
- Replace the output layer with one suitable for the target language
- Retrain on the target language
Replace and Re-train the Output Layer

- Train on the source language
- Replace the output layer with one suitable for the target language
- Retrain on the target language
Replace and Re-train the Output Layer

- Train on the source language
- Replace the output layer with one suitable for the target language
- Retrain on the target language
Multi-task Learning Scenarios

Do phoneme classification in two languages

- Have output units for all phonemes in both languages i.e. around twice as many units
Multi-task Learning Scenarios
Do phoneme and language classification

- Have a phoneme target (from a global phoneme set) and a language ID target
- These could equally apply to other speech units
Multi-task Learning Scenarios
Do phoneme and language classification

- Have a phoneme target (from a global phoneme set) and a language ID target
- These could equally apply to other speech units
Large amounts of speech data exists but not all of it is in the particular language being worked on

Knowledge from data in other languages can be encapsulated in a trained classifier

We will compare different ways in which classifiers could be used
Summary

- Large amounts of speech data exists but not all of it is in the particular language being worked on.
- Knowledge from data in other languages can be encapsulated in a trained classifier.
- We will compare different ways in which classifiers could be used.
Summary

- Large amounts of speech data exists but not all of it is in the particular language being worked on
- Knowledge from data in other languages can be encapsulated in a trained classifier
- We will compare different ways in which classifiers could be used