RLAT – Rapid Language Adaptation Toolkit

Tim Schlippe

04 June 2013
RLAT – Rapid Language Adaptation Toolkit

- Bridge the gap between technology experts → language experts
  - Automatic Speech Recognition (ASR)
  - Machine Translation (MT)
  - Text-to-Speech (TTS)

- Develop web-based intelligent systems
  - Interactive Learning with user in the loop
  - Rapid Adaptation of universal models to unseen languages

- RLAT webpage: http://csl.ira.uka.de/rlat-dev
RLAT – Rapid Language Adaptation Toolkit

- **Collects:**
  - Appropriate text data
  - Appropriate audio data

- **Defines:**
  - Phoneme set
  - Rich prompt set
  - Lexical pronunciations

- **Produces:**
  - Pronunciation model
  - ASR acoustic model
  - ASR language model
  - TTS voice

- **Maintains:**
  - Projects and users login
  - Data and models
RLAT – Project

Build Your System
- Text and prompt selection (help)
- Text management
- SMT-based text normalization (help)
- Audio collection (help)
- Phoneme selection (help)
- Grapheme-to-phoneme rules (help)
- Lexicon pronunciation creation (help)
- Web-derived pronunciations
- Build acoustic model (help)
- Build language model (help)
- Language model management
- Test ASR system
- Create speech synthesis voice

User: tschlippe Language: English Project: GlobalPhone2_test [Logout]

You must do the following to build support for your language:
- Text collection and selection
- Text management
- Language model management
- Audio collection
- Phoneme set specification
- Lexicon pronunciation creation
- Speech recognition acoustic model creation
- Speech recognition language model creation
- Speech synthesis voice creation

Import data from other projects:
- Text data (copies text folder)
- LM data (copies all lm data)
- Phoneme files (setting, phones, g2p rules)
- Lexicon
- Speech data (creates symbolic link to source project)
- SMT System (import all files)

Source project:
- Chinese - 090129_News
- ENGLISH - 080820_NYTtimesOlympia
- English - GlobalPhone2_test
- English - SPICE
- French - 081223_Test1
Overview – Automatic Speech Recognition

Front End (Preprocessing) → Decoder (Search) → Text

- Acoustic Model
- Lexicon / Dictionary
- Language Model

hi /h///ai /
you /j/u/
we /w///i /
hi you
you are
I am
Overview – Automatic Speech Recognition
Goal:
• Get as much relevant text data as possible
• Use the text data for
  – Generating recording prompts
  – Generating vocabulary lists
  – Build Language Models for ASR
• Possibilities to obtain text data:
  – Web crawler (web pages and search engines)
  – Upload (local) / download texts (remote)
• Language encoding
  – To deal with very common alphabets internally all utf-8
RLAT – Text Collection and Selection – Crawling

- Interface for automatic text collection
  - Upload function, Web crawler, Recrawl: boost data from similar sites
- Filtering:
  - Language Identification
  - Remove HTML tags, XML, JavaScripts, …
  - Convert from different character encodings (ISO-8859, UTF8, …) and file formats into raw text in UTF-8
  - Text Normalization
RLAT – Prompt Selection

• Prompts for recording:
  – Collection without transcription

• Prompts should be:
  – Easy to say (no hard words, numerals etc)
  – Rich in variability

```
(data_00001 "will we ever forget it." )
(data_00002 "there was a change now." )
(data_00003 "it fairly clubbed me into recognizing it." )
(data_00004 "i had faith in them." )
(data_00005 "he moved away as quietly as he had come." )
(data_00006 "suddenly his fingers closed tightly over the handkerchief." )
```
RLAT – Text Collection and Selection – Crawling

- Link Level Depth, Upload Test File to evaluate text being crawled, Upload File with URLs and Search Terms, Clock …

Obtain corpus
You can either crawl the internet for a corpus ...

- Enter URL: ![Image](636x3 to 714x39)
- Link Depth: 5
- Interval (snapshot period): 24 hours
- Check if you want RLAT to build LMs automatically

Crawl (view crawl log) (view wget log)

- Upload text file with URLs (Note: Each URL in one line):
  ![Image](39x4 to 131x43)
  ![Image](43x62 to 672x400)

- Upload text file with search terms (Note: Each search term in one line):

- Upload text file from specific domain for informative and visual feedback about the quality of text being crawled (snapshot functionality)

- Upload vocabulary file to determine the case of words from this file rather than statistically (default method). File should contain only one word per line.

After clicking "Crawl", the crawl will run on its own in the background. When it has finished, the symbol next to it will turn green.

13 days 11 hours 48 minutes 35 seconds
• **Snapshot functionality**
  - Informative and visual feedback about the quality of text being crawled
  - Results which indicate the quality (PPL, OOV) of the collected texts are computed and displayed periodically (to be defined by the user) during the crawling process

```
Obtain corpus
You can either crawl the internet for a corpus ...
```

```
<table>
<thead>
<tr>
<th>Enter URL: <a href="http://www.leansien.fr">www.leansien.fr</a></th>
<th>Link Depth:</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval (snapshot period):</td>
<td>No Snapshot</td>
<td>12 Hours</td>
<td>24 Hours</td>
<td>3 Days</td>
<td>7 Days</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Upload text file with URLs (Note: Each URL in one line):</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="button" alt="Durchsuchen" /> <img src="button" alt="Upload URLs" /> text/url</td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Link Depth:</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval (snapshot period):</td>
<td>No Snapshot</td>
<td>12 Hours</td>
<td>24 Hours</td>
<td>3 Days</td>
<td>7 Days</td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Upload text file with search terms (Note: Each search term in one line):</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="button" alt="Durchsuchen" /> <img src="button" alt="Upload Search Terms" /> text/url</td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Upload text file from specific domain for informative and visual feedback about the quality of text being crawled (snapshot functionality)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="button" alt="Durchsuchen" /> <img src="button" alt="Upload Test Data" /> (view uploaded test text)</td>
</tr>
</tbody>
</table>
```

After clicking “Crawl”, the crawl will run on its own in the background. When it has finished, the symbol next to it will turn green.

```
13 days 11 hours 48 minutes 35 seconds
```

```
Language Model Performance of 5-fold cross validation,
 generated on: December 02 2009 10:16:04

Perplexity (Averaged) 54.04752
Perplexity (min) 35.5676
Perplexity (max) 86.5922
OOV Rate (% Averaged) 1.2065
1-gram Coverage (% Averaged) 99.94216
2-gram Coverage (% Averaged) 86.2481
3-gram Coverage (% Averaged) 59.42910
Vocabulary Size (Avg of Training) 96374
Total Words 8248002
```
Example: [www.cnn.com](http://www.cnn.com) evaluated on English GlobalPhone dev set
Approach

1. User supplies an URL to RLAT for crawling
2. Crawler retrieves N documents (web pages)
3. Compute the statistics (TF-IDF) from the N documents
4. Terms with highest TF-IDF score form query terms
5. Query search engine (Google) to get the URLs for the query terms
6. Crawl the URLs for the data
Additional Data
We have found domain terms that might help get more text data relevant to the domain:

- **Check** those terms that are not relevant and hit the **Crawl_Again** button to get additional data.

- Upload text file with search terms for additional data to build interpolated Language Model. **Note**: Each search term in one line.

- Number of hits from search engine: 1

**Calculate frequency statistics**
- **Calc Stats** calculate word frequencies
- **word frequencies**
1. Pure Rule-Based text normalization

2. Text normalization based on Statistical Machine Translation and user support
   - Pre-Normalization
     - Language-independent rule-based normalization / language-specific rule-based normalization (if available)
   - Language-specific normalization by RLAT users
     - Web-based user interface for text normalization
     - Keep the effort for the users low:
       - No use of sentences with more than 30 tokens to avoid horizontal scrolling
       - Sentences to normalize are displayed twice:
         - The upper line shows the non-normalized sentence, the lower line is editable.

RLAT – Text Collection and Selection – Text Management

- Allow to store and handle multiple text files
Language model management

(View LMMangement log file) (View LMMangement input & result file) (View interpolate LMs log)

Select the ngram order: 

Language models

If you provide language model weights, the sum of the selected language model weights should add up to 1. If the sum of the weight of the selected language models is smaller than 1, the weight of the second selected language model will be increased in a way that the weights of the selected language models add up to 1.

✓ www.tagesschau.de-2010-03-18_1501 www.tagesschau.de-2010-03-18_1501.info
   www.tagesschau.de-2010-03-18_1501.lm  optional weight: 0.5

✓ upload-cnnArt-2010-03-18_1511 upload-cnnArt-2010-03-18_1511.info
   upload-cnnArt-2010-03-18_1511.lm  optional weight: 0.1

✓ upload-heiseArt-2010-03-18_1411 upload-heiseArt-2010-03-18_1411.info
   upload-heiseArt-2010-03-18_1411.lm  optional weight: 0.4

☐ upload-tagArt-2010-03-18_1412 upload-tagArt-2010-03-18_1412.info
   optional weight: 

☐ upload-zdfArt-2010-03-18_1527 upload-zdfArt-2010-03-18_1527.info
   optional weight: 

(View interpolated language model)

Select the language model cutoffs: 2-gram: 0 3-gram: 0 4-gram: 0 5-gram: 0

Delete selected language models.

Delete selected language models

Build language models from selected texts. Note that this function will overwrite old language models.

Build LMs from selected texts

Create an interpolated language model with user specified weights.

Note that this function will overwrite the interpolated language model. If language models, which are corresponding to the selected texts, do not exist, they will be build.

Interpolate language models with user specified weights

Upload a development set. Note that this function will overwrite the interpolated language model. If language models, which are corresponding to the selected texts, do not exist, they will be build.

Durchsuchen... Upload_dev_set_and_interpolate_language
RLAT – Language Model Building

Language Model Performance:

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perplexity (Averaged)</td>
<td>27.6301</td>
</tr>
<tr>
<td>Perplexity (min)</td>
<td>26.8446</td>
</tr>
<tr>
<td>Perplexity (max)</td>
<td>30.0191</td>
</tr>
<tr>
<td>OOV Rate (% Averaged)</td>
<td>1.3308</td>
</tr>
<tr>
<td>1-gram Coverage (% Averaged)</td>
<td>98.85626</td>
</tr>
<tr>
<td>2-gram Coverage (% Averaged)</td>
<td>87.1453</td>
</tr>
<tr>
<td>3-gram Coverage (% Averaged)</td>
<td>55.63288</td>
</tr>
<tr>
<td>Vocabulary Size (Avg of Training)</td>
<td>1052</td>
</tr>
<tr>
<td>Total Words</td>
<td>26560</td>
</tr>
</tbody>
</table>

Interpolated Language Model Performance:

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perplexity (Averaged)</td>
<td>30.63308</td>
</tr>
<tr>
<td>Perplexity (min)</td>
<td>28.2386</td>
</tr>
<tr>
<td>Perplexity (max)</td>
<td>33.5173</td>
</tr>
<tr>
<td>OOV Rate (% Averaged)</td>
<td>0.5566</td>
</tr>
<tr>
<td>1-gram Coverage (% Averaged)</td>
<td>99.52136</td>
</tr>
<tr>
<td>2-gram Coverage (% Averaged)</td>
<td>88.37218</td>
</tr>
<tr>
<td>3-gram Coverage (% Averaged)</td>
<td>55.9979</td>
</tr>
<tr>
<td>Vocabulary Size (Avg of Training)</td>
<td>53015</td>
</tr>
<tr>
<td>Total Words (Base+Recrawled)</td>
<td>914067</td>
</tr>
<tr>
<td>Weight (% Base Model)</td>
<td>98.36546</td>
</tr>
</tbody>
</table>

To build a language model, click this button:

Build Language Model

Select the n-gram order: 3

Select number of topics in your data: 0

Select the language model cutoffs: 2-gram: 0 3-gram: 0 4-gram: 0 5-gram: 0

It appears you have collected additional data. You may want to:
- Build language model using this data and interpolate with main LM (recommended).
- Add this data to the original data and build a language model.
- Do not use additional data.

Upload a language model for (re) building the current (possible interpolated) LM interpolating it with the uploaded LM. The LM should be in plain text.

Upload Language Model (.lm or .g2) If there is already another LM build it will be replaced.
Notice that the upload limit is 30MB. Another possibility is to upload a clean text and build the language model.
Overview – Automatic Speech Recognition
RLAT – Phoneme Selection

- Selection from standard IPA chart
- User’s names for phonemes
  - Can match their lexicon (if one exists)
  - Can match their familiarity
- Audio feedback
  - Click to hear recording of each phone
**RLAT – Phoneme Selection**

**Phoneme set specification**

This is a tool which will display all IPA phonemes. You can choose and give names to phonemes you wish your models to use. After you have finished, click the "Submit" button.

**PhoneMapFile settings file**

**Consonants (Pulmonic)**

Please choose the consonant sounds you’d like to have in your new acoustic models by giving it a name in the textbox next to it.

<table>
<thead>
<tr>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Postalveolar</th>
<th>Retroflex</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Pharyngeal</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>f</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)**

<table>
<thead>
<tr>
<th>Consonants (Pulmonic)</th>
<th>Front</th>
<th>Central</th>
<th>Back</th>
<th>Nasal</th>
<th>TRILL</th>
<th>TAP-FLOT</th>
<th>Fricative</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>T TH D DJ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>j G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f v</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dh dz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zh z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sh s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zh z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>th n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dh dz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shadowed areas denote articulations; judged impossible.
NOTE: If you do not specify a name for a diacritic phoneme, it will be ignored!

**Selected letter:** a  **Selected diacritics**  **Possible diacritics**

**Suprasegmentals:** a: Long a· Half-long ã Extra-short

**Syllabic:** a Syllabic a Non-syllabic

**Consonant-release:** aʰ Aspirated aⁿ Nasal release Û No audible release aˡ Lateral release

**Phonation:** ã Voiceless ã Voiced ã Breathy voiced ã Creaky voiced

ã Dental ã Linguolabial ɑ Apical ɑ Laminal

**Articulation:** ã Advanced ã Retracted ã Centralized ã Mid-centralized

ã Raised ã Lowered

ã More rounded ã Less rounded aⁿ Labialized aˡ Palatalized

**Co-articulation:** a³ Velarized a⁵ Pharyngealized ɑ Velarized ɑ Advanced tongue root

ɑ Retracted tongue root ɑ Nasalized ɑ Rhotacized

Phoneme to dictionary assignment:

a:  AL

**Diacritics**

Diacritics may be placed above a symbol with a descender, e.g. ʃ.
Vowels

Please choose the consonant sounds you’d like to have in your new acoustic models by checking the check box.

Front
- [Y]
- [i]
- [UE]

Central
- [X]
- [u]

Back
- [W]
- [U]
- [UW]

Close-mid
- [E]
- [e]
- [OE]
- [ø]

Close
- [AX]
- [θ]

Open-mid
- [EH]
- [ɛ]
- [OL]
- [æ]

Open
- [AE]
- [æ]

Vowel Triangle Reference Graph
Here you can see the relation between different IPA vowels.
RLAT – Lexicon Pronunciation Creation

- 3 options to obtain pronunciations:
  - Lex Learner
  - IPA Crawler (Web-derived pronunciations)
  - Keynounce – A game for pronunciation generation
Initial Grapheme to Phoneme Rules

Please input an initial Grapheme to Phoneme (G2P) rule of your language.

Based on this rule, our system will “guess” the correct pronunciation of words in your language. You are able to view the predicted pronunciation, change it, delete it, or type a correct pronunciation for this word. The correct pronunciation will be saved into your dictionary and our system will make use of this information to make a better “guess” in predicting pronunciation of new words.

Now please type in Grapheme to Phoneme rule (G2P) for us. Just type one of the most common pronunciation for each grapheme. Thanks.

```
Upload g2p
```

```
Upload char.info
```

Download/view current settings: g2p char.info

<table>
<thead>
<tr>
<th>Grapheme</th>
<th>Upper Case</th>
<th>Lower Case</th>
<th>Punctuation Mark</th>
<th>Number</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>b</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>c</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>d</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>e</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>f</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>g</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
<tr>
<td>h</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
<td>☑</td>
</tr>
</tbody>
</table>
John Kominek and Alan W. Black,
Learning Pronunciation Dictionaries: Language Complexity and Word Selection Strategies,
HLT Conference of the NAACL, 2006
**sein**

**Hilfsverb**

Anmerkung:
Alle Verbindungen mit "sein" schreibt man nach neuer Rechtschreibung getrennt (da sein, weg sein, zusammen sein ...).

Silbentrennung:
sein, Präteritum: war, Partizip II: ge-wen

Aussprache:
IPA: [zaɪn]

**sein (Französisch)**

**Substantiv, m**

Silbentrennung:
sein, Plural: seins

Aussprache:
IPA: [zɛ̃]

Hörbeispiele: [zɛ̃] sein, Plural: —
**Web derived pronunciations**

This function allows you to extract pronunciations in IPA format from one or more websites. It is intended for Wiktionary or similar online dictionaries that have separate pages for individual words. Each source URL is queried with each word from a search vocabulary.

**Source URLs**

Please enter a list of URLs into the textbox below, one URL per line. Each URL must contain the placeholder "$$" or "!!!". The former will be substituted by case variations of words from your search vocabulary, whereas the latter will be substituted by the true case of words exactly as written in your vocabulary.

http://de.wiktionary.org/wiki/$$
http://de.wikipedia.org/wiki/$$

**Search vocabulary (word list)**

Please upload a list of words whose IPA pronunciations you would like to find on the websites given by the source URLs. The upload must be a plain UTF-8 text file. It can be a vocabulary list (one word per line) or a reference text (several words per line). A list of unique words will be automatically extracted from the upload. However, no text normalization or filtering is performed, so you have to make sure that the upload contains only characters that are compatible with the specified source URLs.

**Status**

IPA crawler is running.
Searching URLs for word "Keyboard" (7 of 13) ...

**Result**

(view vocab) (view log) (view wget log) (view output) (view generated dictionary)
Keynounce – A Game for Pronunciation Generation through Crowdsourcing
Overview – Automatic Speech Recognition

[Diagram showing components of Automatic Speech Recognition system, including AM, Lex, and LM]
RLAT – Audio Collection

Audio collection - RLAT - Mozilla Firefox

Build Your System
- Text and prompt selection (help)
- Audio collection (help)
- Phoneme selection (help)
- Grapheme-to-phoneme rules (help)
- Lexicon pronunciation creation (help)
- Build acoustic model (help)
- Build language model (help)
- Test ASR system
- Create speech synthesis voice

User: demo
Language: English
Project: DemoSystem
[Logout]

Audio collection

Record audio: [Watch Demo Video]

File: data_00001.wav
Length: 0.0
Position: 0.0

will we ever forget it.

Sessions Panel
- Set Working Dir
- New Speaker
- Close Speaker
- Play
- Record
- Upload

Process Log
1: SUCCESS: Server path set to demo/English/DemoSystem
RLAT – Acoustic Model Building

- Acoustic Model Building requires:
  - Recorded Speech Data
  - Phone set definition
  - Pronunciation Lexicon

- Two step process:
  1. Configuration
  2. Model Training
Step 2. Acoustic Model Training

- Select method of acoustic model training.
  1. Either the specified speakers (see below) or 10% random speakers will be selected for tuning (devSet) and evaluating (testSet) the performance of the system. The remaining speakers will be selected for training the acoustic models.
  2. 90% of each speaker’s utterances are used for training and the rest of the utterances are used for testing the acoustic models.
  3. All available data will be used for training the acoustic models. Testing can be done via the 'Test ASR system' interface.

Train_Acoustic_Models LOG_FILE

- Initialization: IN PROGRESS...
- Computing Labels: NOT STARTED YET.
- Computing Cepstral Means: NOT STARTED YET.
- Computing LDA matrix: NOT STARTED YET.
- K-Means Clustering over the codebooks: NOT STARTED YET.
- EM Training: NOT STARTED YET.
- Segmentation of Speech: NOT STARTED YET.
- Making Polyphone Trees: NOT STARTED YET.
- EM Training over the code-books: NOT STARTED YET.
- Clustering Contexts: NOT STARTED YET.
- Performing the Splits: NOT STARTED YET.
- Context Dependent Acoustic Models Training: NOT STARTED YET.

View/Hide Results >>

Step 3. Acoustic Model Testing

Currently selected Language Model for testing: Default LM - See page 'Build language model'

- Decode on specified test set.
- Decode on specified development set.
Overview – Automatic Speech Recognition

[Diagram showing the process of automatic speech recognition with labeled components AM, Lex, and LM, along with phonetic transcriptions for 'hi', 'you', 'I am' and 'we' with associated symbols for each stage.]
Simple echo back testing function
Thanks for your interest!
References


