Automatic Speech Recognition for Dialectal Arabic

PhD Examination

Mohamed Ahmed Elmahdy
August 26, 2011

Dialogue Systems Group
Institute of Information Technology
University of Ulm, Germany
Arabic Language

- Largest still living Semitic language
- 250+ million native speakers

- Modern Standard Arabic (MSA)
  - Standardized
  - A lot of ASR research
  - Not used in everyday life

- Used in everyday life
  - Not standardized (mainly spoken)
  - Many different dialects
  - Very few ASR research

Significant differences between MSA and Dialectal Arabic
  - Considered as completely different languages
MSA Versus Dialectal Arabic

- Egyptian Colloquial Arabic (ECA) has been chosen as a typical Arabic dialect
  - Phonological
    - /t/, /s/ in ECA instead of /T/ in MSA
      - e.g. /talaː tah/ (three) in ECA versus /Talaː Tah/ in MSA
  - Lexical
    - /tˈArAbEːzA/ (table) in ECA versus /tˈawila/ in MSA
  - Syntactic
    - SVO in ECA versus VSO in MSA
Automatic Speech Recognition

- High level diagram for a state-of-the-art ASR system

\[ \hat{W} = \arg \max_{W \in L} P(O \mid W)P(W) \]

For dialectal Arabic, sparse and low quality speech corpora are available
Thesis Objectives

- Acoustic modeling for dialectal Arabic where little speech data exists
- To benefit from existing MSA speech data to improve dialectal Arabic speech recognition
- Acoustic modeling for dialectal Arabic where phonetic transcription is not possible
Outline

- Introduction
- Previous work
- Approaches
- Experiments and results
- Conclusions and future directions
Previous Work

- **Data Pooling Acoustic Modeling**
  - A speech data pool of MSA and Egyptian Colloquial Arabic (ECA) is used to train the acoustic model
  - By adding more MSA data, less contribution of ECA data

Outline

- Introduction
- Previous work
- Approaches
- Experiments and results
- Conclusions and future directions
Proposed Approaches for Dialectal Arabic ASR

- Phonemic acoustic modeling
  - Dialectal speech data where phonetic transcription is available

- Graphemic acoustic modeling

- Unsupervised acoustic modeling

- Arabic Chat Alphabet-based acoustic modeling
Phonemic Cross-Lingual Acoustic Modeling

- Benefit from existing large MSA speech corpora

  - Assumptions:
    → MSA is always a 2nd language for any Arabic speaker
    → Large amount of MSA speech data (large number of speakers) implicitly cover all the acoustic features of the different Arabic dialects

  - Approach:
    → Train an acoustic model using a large amount of MSA speech data
    → Adaptation of the MSA acoustic models with a little amount of dialectal speech data
Phonemic Cross-Lingual Acoustic Modeling (cont.)

- State-of-the-art AM adaptation techniques include:
  - Maximum Likelihood Linear Regression (MLLR)
    \[ \Phi_{MLLR} = A\Phi + b \]
  - Maximum A-Posteriori (MAP)
    \[ \Phi_{MAP} = \arg\max_{\Phi} P(O \mid \Phi)P(\Phi) \]
    - Requirement: adaptation data and the AM have to share the same language and phoneme set

- Egyptian Colloquial Arabic (ECA) is chosen as a typical dialect

- INITIALLY: MSA and ECA do not share the same phoneme inventory
Phonemic Cross-Lingual Acoustic Modeling (cont.)

- **SOLUTION**: Phoneme sets normalization
  - AM adaptation is possible

- Phoneme sets normalization
  - Several phone mapping rules are applied
  - Map ECA phonemes to their origins in MSA (even if they are acoustically different)

\[
\begin{align*}
\text{ECA} & \quad /b/ \quad /g/ \quad /j/ \quad /e/ \quad /i/ \quad /o/ \quad /u/ \quad /t/ \quad \ldots \ldots \\
\text{MSA} & \quad /b/ \quad /dZ/ \quad /i/ \quad /u/ \quad /t/ \quad \ldots \ldots \\
\end{align*}
\]

\[
\begin{align*}
\text{جزر} & \quad (\text{carrot}) \\
/g/ /A/ /z/ /A/ /r/ & \quad \rightarrow \quad /dZ/ /a/ /z/ /a/ /r/
\end{align*}
\]
Phonemic Cross-Lingual Acoustic Modeling (cont.)

- Block diagram for the proposed approach
- The adapted ECA AM is evaluated against the ECA baseline AM
Proposed Approaches for Dialectal Arabic ASR

- **Phonemic acoustic modeling**
  - Dialectal speech data where phonetic transcription is available

- **Graphemic acoustic modeling**
  - Phonetic transcription is not possible/difficult
  - Short vowels are missing
  - Phonetic transcription is approximated to be word letters

- **Unsupervised acoustic modeling**
  - Transcriptions are not available at all
  - Dialectal speech was automatically transcribed using a MSA model

- **Arabic Chat Alphabet-based acoustic modeling**
  - Latin letters are used instead of Arabic ones
  - Include short vowels that are missing in traditional Arabic orthography
Outline

- Introduction
- Previous work
- Approaches
- Experiments and results
- Conclusions and future directions
Speech Corpora

- **Egyptian Colloquial Arabic (ECA) corpus**
  - High quality read speech
  - Accurate phonetic transcription
  - Different speech domains to ensure good acoustic features coverage: time/date, spelling, restaurant, numbers, reservation, etc

- **Levantine Colloquial Arabic (LCA) corpus**
  - Spontaneous speech
  - No phonetic transcription

- **Modern Standard Arabic (MSA) corpus**
  - News broadcast speech
  - Accurate phonetic transcription

(English: First class ticket)
Graphemic: تذكرة درجة أولى
Phonemic: /tazkara daraga ?uula/
ACA: /tazkara daraga 2uula/
Phonemic Cross-Lingual Adaptation Results

- ECA corpus:
  - 65% for training/adaptation
  - 35% for testing

Word Error Rate (WER)

\[
WER = \frac{Sub + Ins + Del}{N}
\]

- ECA Phonemic AM

<table>
<thead>
<tr>
<th>Approach</th>
<th>WER (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECA baseline</td>
<td>15.00</td>
</tr>
<tr>
<td>MSA only</td>
<td>41.80</td>
</tr>
<tr>
<td>MSA+ECA data pooling</td>
<td>10.00</td>
</tr>
<tr>
<td>MSA+ECA adaptation</td>
<td>15.00</td>
</tr>
</tbody>
</table>

41.8% Relative reduction in WER
Effect of MSA Speech Data Amount

- Varying the amount of MSA speech data
- Effect on phonemic cross-lingual adaptation

![Graph showing the decrease in WER with increasing MSA speech amount]

**Consistent decrease in WER**
Graphemic Cross-Lingual Adaptation Results

- Tested on two Arabic dialects:
  - Egyptian Colloquial Arabic (ECA)
  - Levantine Colloquial Arabic (LCA)

### Results

<table>
<thead>
<tr>
<th></th>
<th>ECA graphemic AM</th>
<th>LCA graphemic AM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dialect baseline</td>
<td>18.00%</td>
<td>60.00%</td>
</tr>
<tr>
<td>MSA only</td>
<td>20.00%</td>
<td>52.00%</td>
</tr>
<tr>
<td>MSA+dialect adaptation</td>
<td>22.5% for ECA</td>
<td>56.9% for LCA</td>
</tr>
</tbody>
</table>

Relative reduction in WER
- **22.5%** for ECA
- **15.9%** for LCA
Outline

- Introduction
- Previous work
- Approaches
- Experiments and results
- Conclusions and future directions
Conclusions and Future Directions

- **Conclusions**
  - Supervised/unsupervised cross-lingual acoustic modeling for dialectal Arabic
  - Improvements are observed in both phonemic and graphemic modeling
  - Consistent reduction in WER by adding more MSA data
  - The approach can be extended to other dialects e.g. Levantine
  - Novel technique using the Arabic Chat Alphabet for phonetic transcription

- **Future directions**
  - Extension to all the Arabic dialects
  - Cross-lingual language modeling for dialectal Arabic using existing MSA text resources
  - Automatic dialect recognition
Selected Publications


Conclusions and Future Directions

- **Conclusions**
  - Supervised/unsupervised cross-lingual acoustic modeling for dialectal Arabic
  - Improvements are observed in both phonemic and graphemic modeling
  - Consistent reduction in WER by adding more MSA data
  - The approach can be extended to other dialects e.g. Levantine
  - Novel technique using the Arabic Chat Alphabet for phonetic transcription

- **Future directions**
  - Extension to all the Arabic dialects
  - Cross-lingual language modeling for dialectal Arabic using existing MSA text resources
  - Automatic dialect recognition

Thank you for your attention
Arabic Dialects Classification

Arabic

- Standard Arabic
  - Modern Standard Arabic (MSA)
  - Classical Arabic
    - Northern
      - Levantine
    - Iraqi
- Dialectal Arabic
  - Western
  - Eastern
    - Maghreb
    - Libyan
    - Tunisian
    - Algerian
    - Moroccan

- Peninsula
  - Yemeni
  - Omani
  - Saudi
  - Gulf
Phonetic Transcription in English

Orthographic transcription

one three five

Audio file

Lexicon

<table>
<thead>
<tr>
<th>Word</th>
<th>Phonemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONE</td>
<td>W AH N</td>
</tr>
<tr>
<td>TWO</td>
<td>T UW</td>
</tr>
<tr>
<td>THREE</td>
<td>TH R IY</td>
</tr>
<tr>
<td>FOUR</td>
<td>F AO R</td>
</tr>
<tr>
<td>FIVE</td>
<td>F AY V</td>
</tr>
<tr>
<td>SIX</td>
<td>S IH K S</td>
</tr>
<tr>
<td>SEVEN</td>
<td>S EH V AH</td>
</tr>
<tr>
<td>NEIGHT</td>
<td>EY T</td>
</tr>
<tr>
<td>NINE</td>
<td>N AY N</td>
</tr>
<tr>
<td>ZERO</td>
<td>Z IY R OW</td>
</tr>
<tr>
<td>OH</td>
<td>OW</td>
</tr>
</tbody>
</table>

Feature extraction

Phonetic transcription

W AH N TH R IY F AY V

Feature vectors
Phonetic Transcription in Arabic

- Diacritics (short vowels) are not written (ambiguity)
- High OOV rate due to morphological complexity
Unsupervised Acoustic Modeling

Normalized MSA Corpus

MSA Corpus

ECA Corpus

Adaptation set

Unlabeled ECA speech

ECA Language model

Speech recognizer/ Evaluation

WER %

Normalized ECA Corpus

Testing set

Backup slides