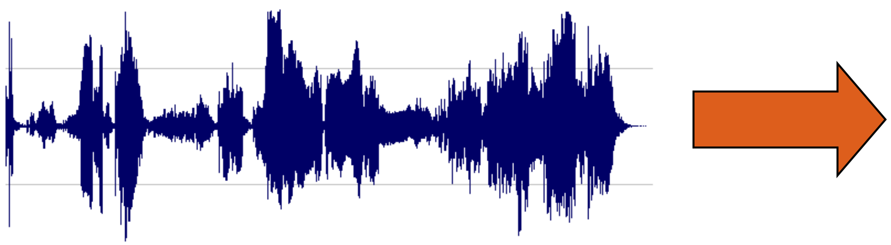
Spoken Arabic Dialect Identification

Fadi Biadsy, Julia Hirschberg, Nizar Habash
Columbia University in the City of New York

EACL 2009 Workshop on
Computational Approaches to Semitic Languages

Motivation

- Given a speech segment of a predetermined language



Dialect = {D1, D2,…,DN}

- **<u>Goal</u>: Arabic dialect Identification**

- Accent and dialect ID have begun to receive attention

- Dialect ID more difficult problem than language ID

Goal

- Test the hypothesis that

  [Gulf, Iraqi, Levantine, Egyptian, Modern Standard Arabic (MSA)]

  can be distinguished based on their phonotactics

- *Phonotactics*: Rules that govern phone sequences
  - e.g., “/p/ /b/” not allowed in English

- Affect the phone sequence distribution of a dialect

Intuition

- Differences between phonetic inventory, lexical choice, and morphology impact the phone sequence distribution

  - For example *"she will meet him"*:

Differences in phonetic inventory and vowel usage

MSA:	/s/ /a/	/t/ /u/ /q/	/A/	/b/	/i/	/l/ /u/	/h/	/u/
Egy:	/H/ /a/	/t/ /?/	/a/	/b/		/l/		/u/
Lev:	/r/ /a/ /H/	/t/ /g/	/A/	/b/		/l/		/u/

  - Phone sequence distribution captures also part of the syllabic structure ➔ models the rhythmic structure

Outline

- Background and Related work
- Arabic Dialects
- Corpora
- Our Approach
  - Phonotatic approach for dialect ID
- Experiments and Results
- Conclusion and Future Work

Dialect ID is Important

I. Infer speaker's regional origin
   - Improve Automatic Speech Recognition (ASR)
     - Model adaptation: Pronunciation, acoustic, morphological, language models
     - For ASR, Vergyri et al. (2005) treated Arabic dialect as different languages
   - Spoken dialogue systems – adapt TTS systems
   - For answering biographical questions

II. Learn about the differences between dialects

III. Call centers – crucial in emergency situations

Related Work

Spoken cues used for language and dialect ID

- Phonotactics
  - Zissman, et al. (1996A) distinguish Cuban and Peruvian dialects
- Spectral differences
  - Torres-Carrasquillo et al. (2004) use Gaussian Mixture Models over MFCCs with shifted-delta-cepstral features to identify Cuban and Peruvian dialects
  - Alorfi (2008) uses an ergodic HMM to model phonetic differences between two Arabic dialects (Gulf and Egyptian Arabic) over MFCC
- Prosody (e.g., intonation and rhythm)
  - Barakat et al. (1999): subjects use **intonational cues** to identify Eastern vs. Western Arabic dialects
  - Hamdi et al. (2004) show rhythmic differences between Eastern vs. Western Arabic Dialects

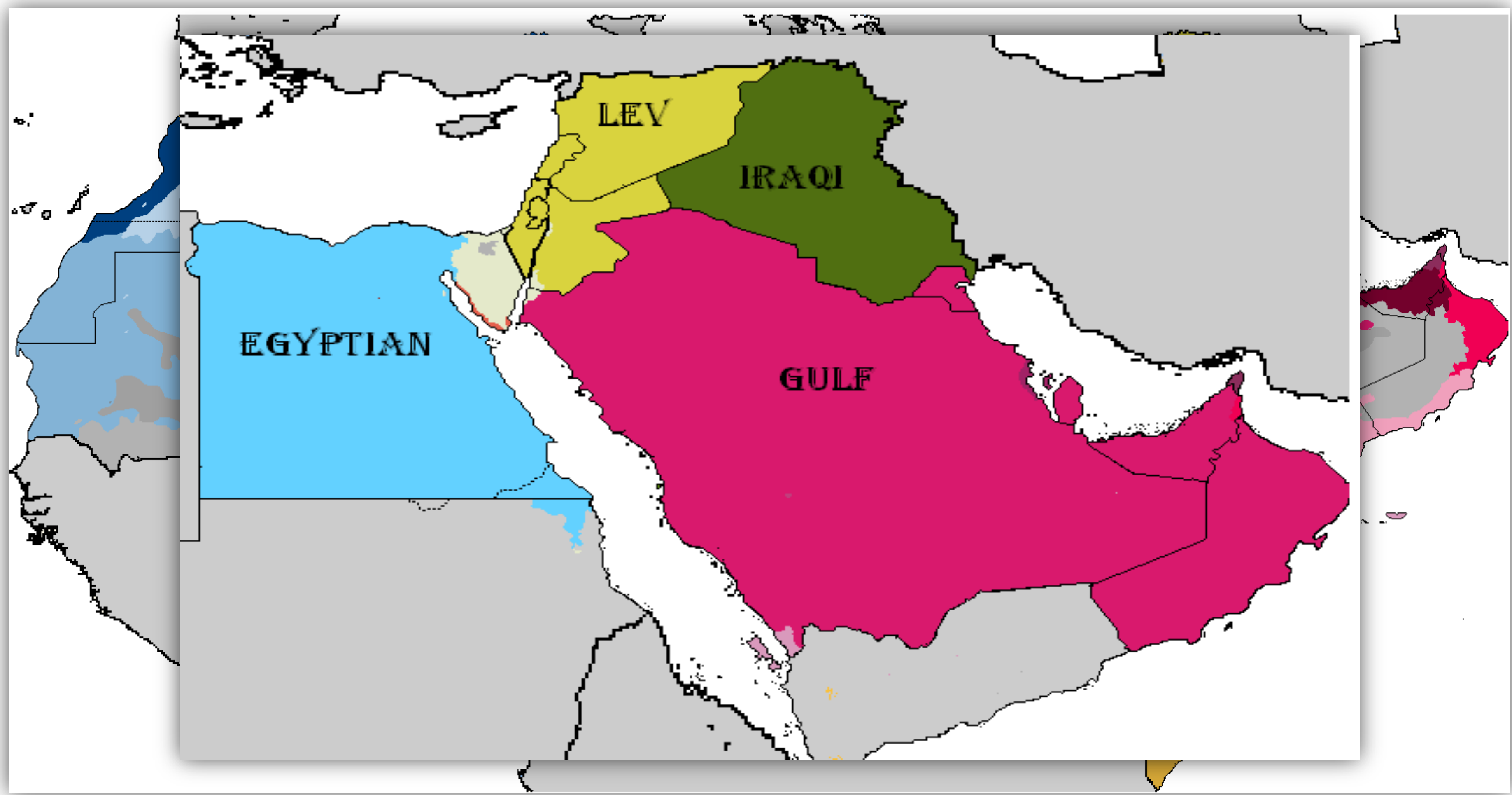
Arabic Dialects

- Arabic is a collection of multiple variants
  - Modern Standard Arabic (MSA) has a special status:
    - formal written standard language of media, culture and education across the Arab world
  - Colloquial Arabic: spoken dialects are the means for communication in daily life

- Variants differ greatly from each other
  - Lexical choice, morphology, syntax, phonology and prosody

- ***Code-switching*** between MSA and colloquial Arabic ➔ problems for ASR

Arabic Dialects



(by Arab Atlas)

Corpora – four dialects

- Recordings of spontaneous telephone conversation produced by native speakers of the four dialects available from LDC

Dialect	# Speakers	Total Duration	Test Speakers	Corpus
Gulf	965	41.02h	150	Gulf Arabic conversational telephone Speech database (Appen Pty Ltd, 2006a)
Iraqi	475	25.73h	150	Iraqi Arabic conversational telephone Speech database (Appen Pty Ltd, 2006b)
Egyptian	398	75.7h	150	CallHome Egyptian and its Supplement (Canavan et al., 1997) CallFriend Egyptian (Canavan and Zipperlen, 1996)
Levantine	1258	78.89h	150	Arabic CTS Levantine Fisher Training Data Set 1-3 (Maamouri, 2006)

Corpora – MSA

- No data with similar recording conditions for MSA
- So we use TDT4 Arabic broadcast news
  - 47.6 hours of speech (downsampled to 8khz)

- 150 speakers, identified automatically, from a corpus used in the DARPA GALE program (12.06 hours of speech)
  - Non-speech data was removed manually

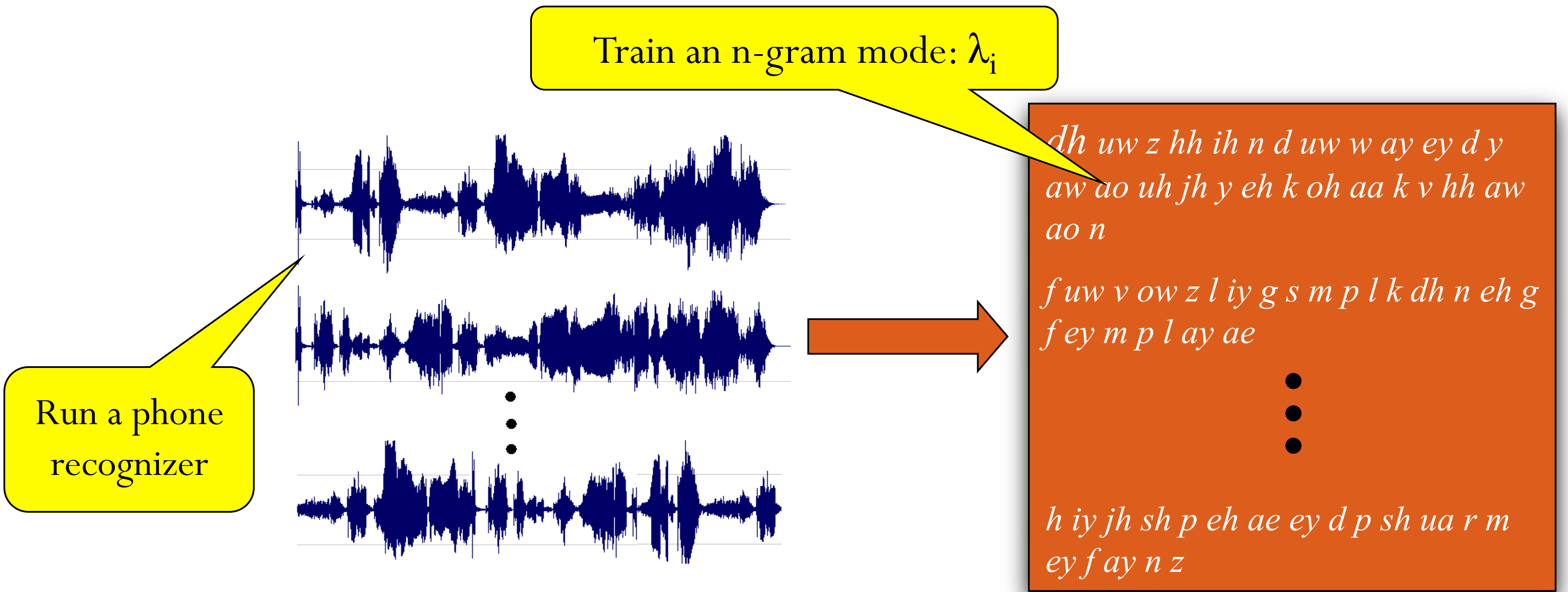
Our approach

- Adopt the Parallel Phone Recognition followed by Language Modeling (Parallel PRLM) used for Language ID
  - (Zissman et al., 1996B)

- We use Parallel PRLM to show that Arabic dialects can be distinguished based on phonotactics

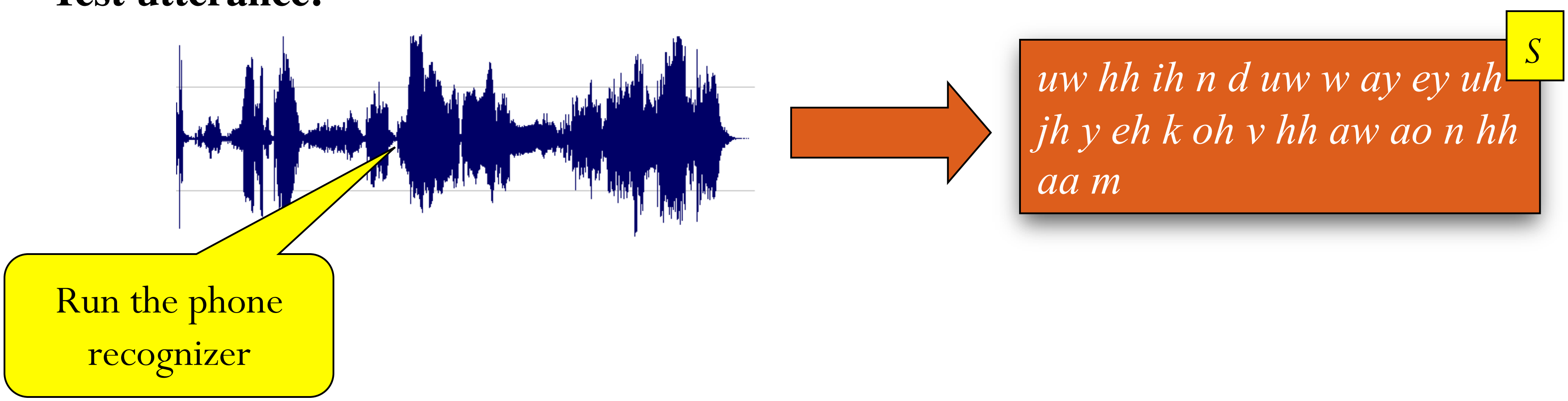
PRLM– Training

For each dialect i:



PRLM – Identification

Test utterance:



$$D^* = \arg\max_i P(D_i \mid S) = \arg\max_i P(S \mid D_i) P(D_i)$$

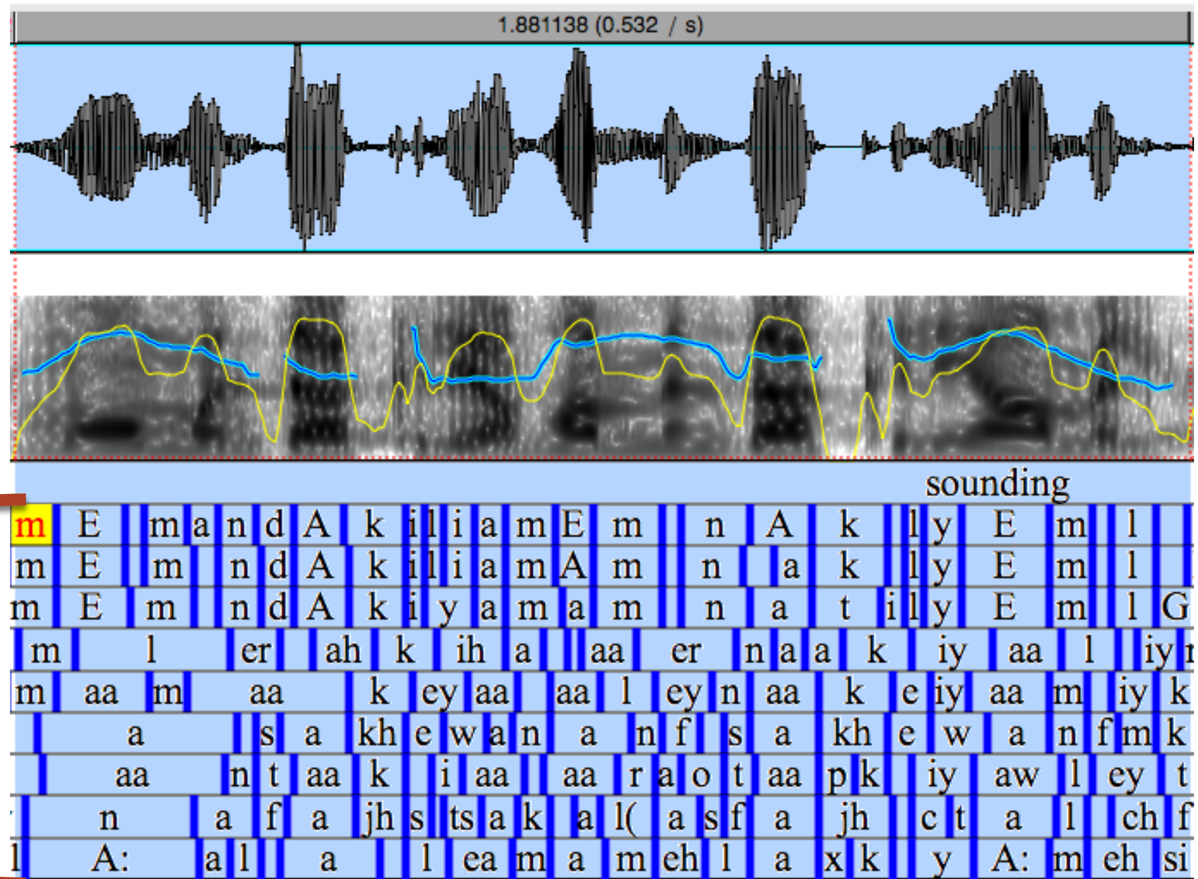
$$= \arg\max_i (S \mid \lambda_i) P(\lambda_i)$$

Parallel PRLM

- Instead of using one phone recognizer, use multiple (M) different phone recognizers
  - ➔ *M* n-gram models for each dialect
    - **English, Arabic, Mandarin, etc.**

- Advantages:
  - Capture subtle phonetic differences
  - PRs are prone to errors, so relying upon multiple phone streams may lead to more robust model overall

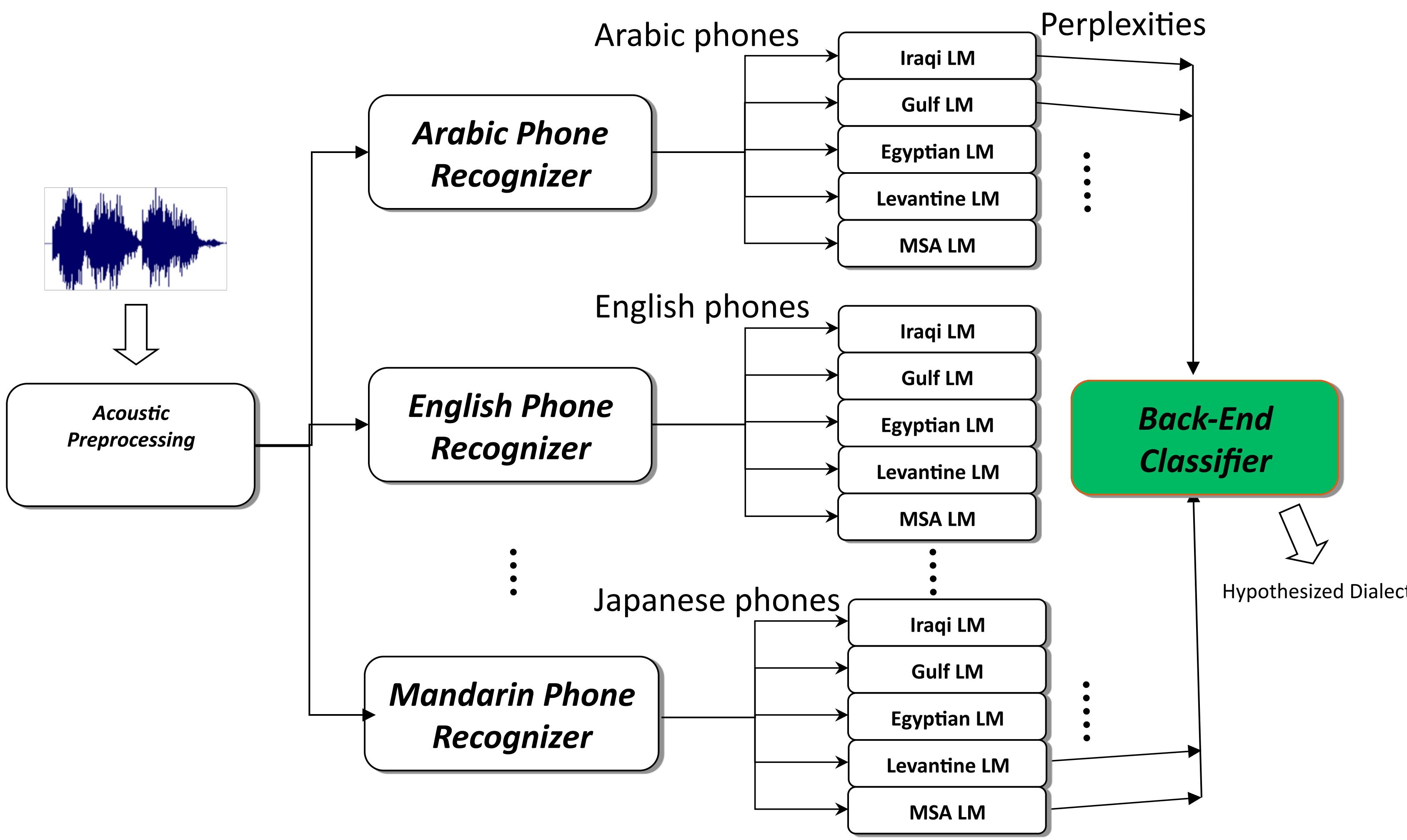
Example



9 phone streams produced by 9 different phone recognizers

Parallel PRLM – Identification

+ Back-End Classifier



Phone Recognizers

- Six open-loop phone recognizers for
  - **English, German, Japanese, Hindi, Mandarin, Spanish**
  - A toolkit developed by Brno University of Technology (Matejka et al., 2005)
    - Trained on OGI multi-language corpus

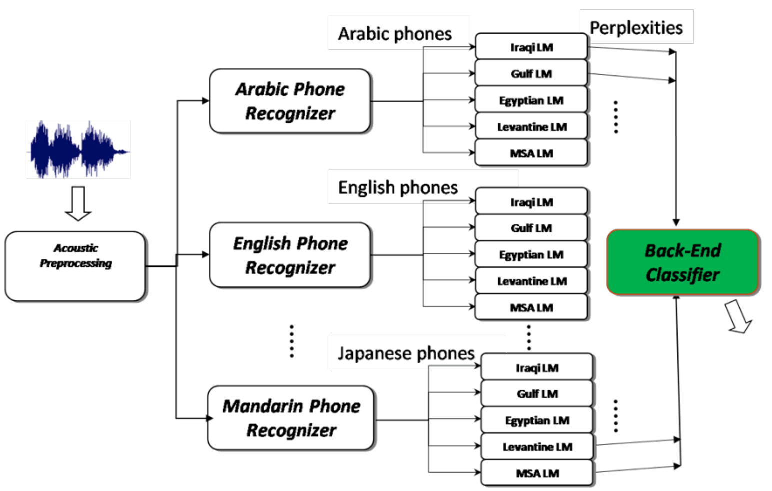
Arabic Phone Recognizers

- We built three MSA phone recognizers using HTK
  - Pronunciation Dictionary following (Biadsy et al., 2009, NAACL)

1. With the standard 6 vowels
2. Models **emphatic vowels** (6 standard + 6 emphatic vowels)
   - Emphatic vowels: vowels that precede and/or succeed emphatic consosnants {E,T,D,Z}
   - e.g., b A s (kiss) vs. /b/ ***/A/ /S/*** (bus)
3. With a bigram LM and emphatic vowels (6 standard + 6 emphatic vowels)

Experiments

- Dialect Identification:
  1. Use the LMs to produce perplexity scores for each of the **150** test speakers **for each dialect –** total 600 feature vectors
  2. Report 10 fold cross validation of the back-end classifier

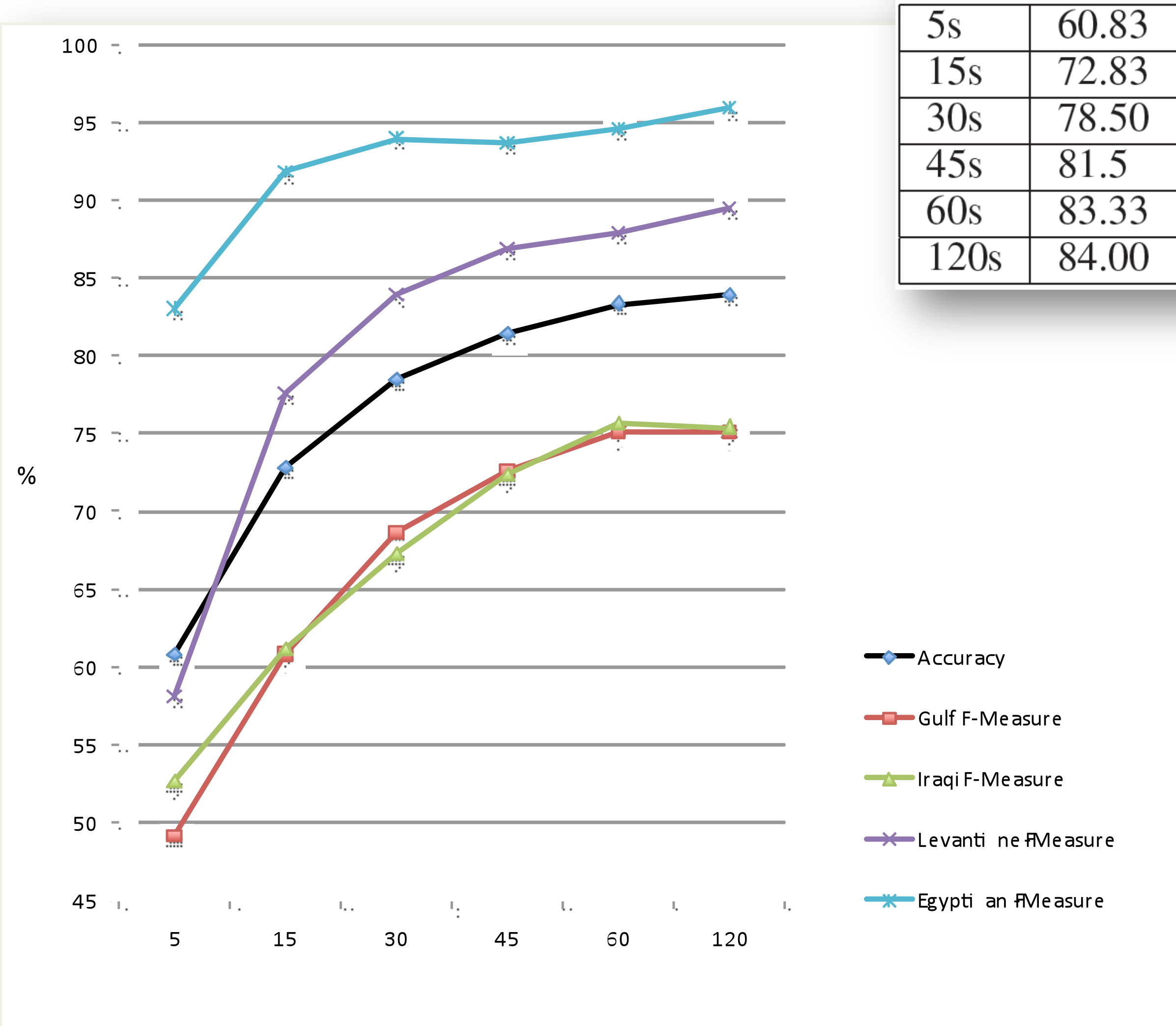


Results

Gulf vs. Egyptian Dialect ID

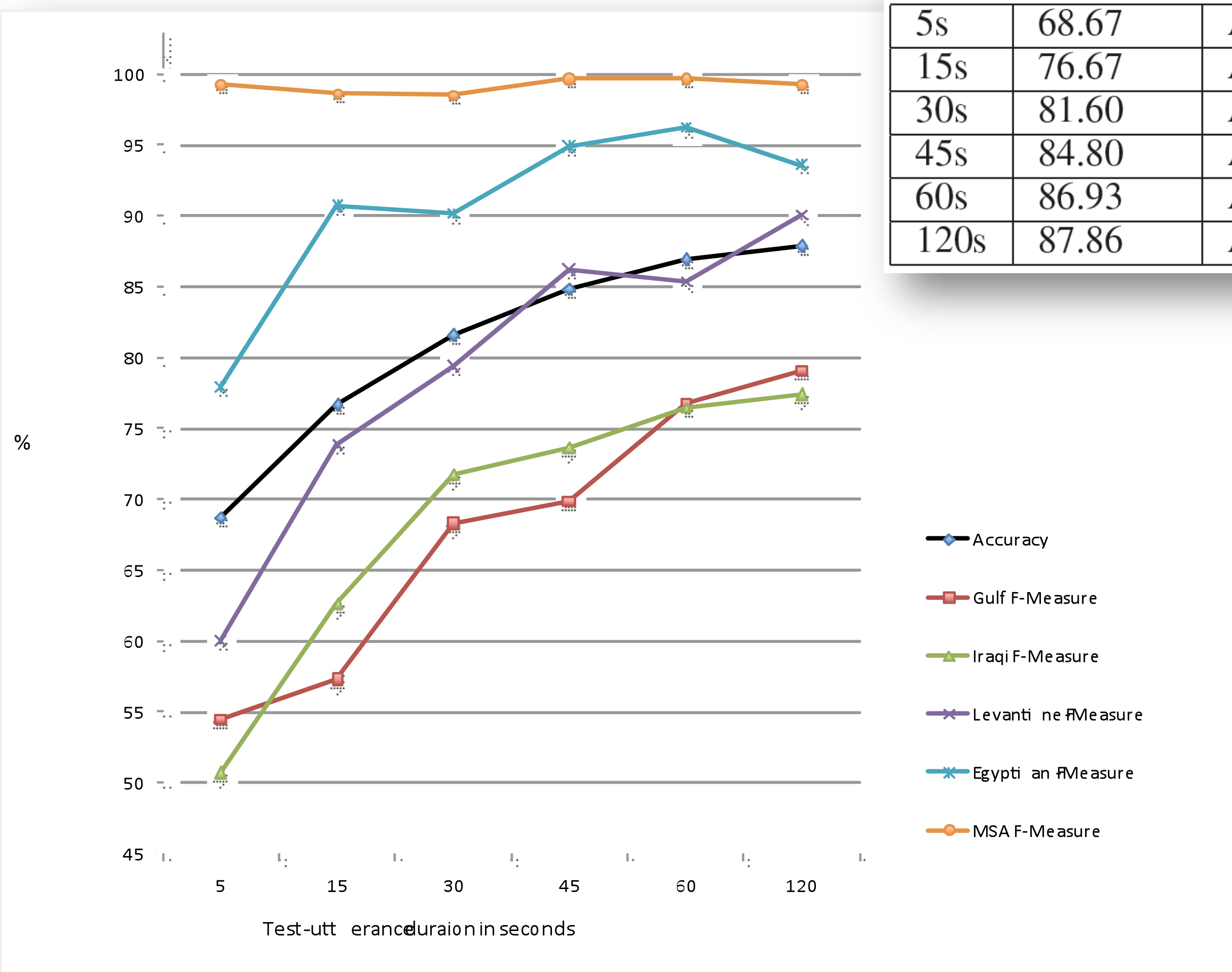
- Previous work (Alorfi 2008): best result is **96.67%.**
  - Data: 40 speakers (20 **Gulf** collected from TV soap and 20 CallHome **Egyptian**)
- Our best result is **97.00%** (Egyptian and Gulf F-Measure = 0.97)
  - when using the following phone recognizers:
    - Arab open loop emphatic, English, Japanese, and Mandarin
- Advantages:
  - Our data from same recording conditions as opposed to mix of different genres
  - Our system tests 300 speakers as opposed to 40 ➔ may be more reliable
  - Our test data includes female speakers too ➔ more general

Results
Four colloquial dialects



Dur.	Acc. (%)	Phone Recognizers
5s	60.83	ArbOE+ArbLME+G+H+M+S
15s	72.83	ArbOE+ArbLME+G+H+M
30s	78.50	ArbO+H+S
45s	81.5	ArbE+ArbLME+H+G+S
60s	83.33	ArbOE+ArbLME+E+G+H+M
120s	84.00	ArbOE+ArbLME+G+M

Experiments
Four colloquial dialects + MSA



Dur.	Acc. (%)	Phone Recognizers
5s	68.67	ArbO+ArbLME+H+M
15s	76.67	ArbLME+G+H+J+M
30s	81.60	ArbO+ArbOE+E+G+H+J+M+S
45s	84.80	ArbOE+ArbLME+E+G+H+J+M+S
60s	86.93	ArbOE+ArbLME+G+J+M+S
120s	87.86	ArbO+ArbLME+E+S

***MSA results might be inflated due to:**

1. MSA is a mix of BN, read speech, telephone speech
2. Different recording conditions
3. Speaker IDs in MSA corpus were determined automatically

Back-End Classifier (4 way, 2m test)

Classifier	Accuracy %
Average and Max (Zissman et al., 1996B)	65.5
SVM (linear kernel)	72.5
SVM (quadratic kernel)	80.0
Multilayer Neural Network	79.67
Logistic Regression	**84.0**

Phone recognizers (4 way, 2m test)

Phone Recognizers	Accuracy %
Our 3 Arabic phone recognizers	80.16
The other 6 phone recognizers	76.16
Combination (without feature selection)	**83.5**

Conclusion

- **Hypothesis Confirmed**: Arabic dialects and MSA significantly differ from each other in terms of their phonotactic distribution
- Parallel PRLM approach is effective also for identifying Arabic dialects with considerable accuracy:
  - **5-way: 87.86% (with 120s of test utterance)**
  - **4-way: 84.0% (with 120s of test utterance)**
- A back-end classifier significantly improves over a simple combiner
- Typically our MSA phone recognizers' sequences with emphatic vowels are the most valuable sequences
- The most distinguishable dialects: (using 30s test utterance duration, for example)
  1. **MSA** (F-Measure is **always** above **98.00%**).
  2. **Egyptian** (F-Measure: **90.2%,** with **30s**)
  3. **Levantine** (F-Measure: **79.4%**, with **30s**)
  4. **Iraqi (**F-Measure: **71.7%**, with **30s**)
  5. **Gulf** (F-Measure **68.3%**, with **30s**)

Most confusable dialects

Future Work

- Explore the prosodic difference across Arabic dialects
  - e.g., intonation, rhythm, and pitch accent distribution
- Attempt to improve the accuracy of the system using these prosodic features
- Reduce the duration of test utterances necessary to identify the dialect
- Identify code switching points

Thank you!

- Acknowledgments:
  - Thanks to: Dan Ellis, Kathy McKeown, Bob Coyne, Kevin Lerman, Michal Mandel, Andrew Rosenberg, and Kapil Thadani for useful discussions.

Confusion Matrix

	Gulf	Iraqi	Levantine	Egyptian
Gulf	<u>115</u>	24	10	1
Iraqi	27	<u>112</u>	10	1
Levantine	8	7	<u>132</u>	3
Egyptian	2	3	3	<u>142</u>