Outline

• Amount of available resources
  – Relatively large → many possible applications (audio indexing, broadcast monitoring)
  – Smaller → speech triage
  – Minimal → binary decision of on-topic vs. off-topic

• Advanced Extraction Technologies
Audio Indexing and Broadcast Monitoring
Audio Indexing System

Audio Source → Audio Feeder → Metadata Composer → XML Metadata

Component Manager:
- Speech Detection
- Speaker Segmentation
- Speaker Tracking
- Speaker Identification
- Speech Recognition
- Name Extraction
- Sentence Detection
- Machine Translation

Dense component integration creates state-of-the-art Rich Transcriptions
Real-Time, Low Latency, Online Algorithms

• Goal: Preserve speech-to-text (STT) accuracy while meeting real-time and low-latency operational requirements

Streaming Audio

Speaker Tracking

Online Speaker Change Detection

Online Speaker Clustering

Speech-To-Text

STT Decode with Online Speaker Adaptation

STT Transcription
Online Speaker Change Detection

- Two-step Approach
  - Speech Detection: Phone decode with 8 classes
    - speech phones: vowel, fricative, obstruent
    - non-speech phones: music, silence, breath, lip-smack, laughter
  - Change Detection
    - Hypothesize speaker change on every phone class boundary
    - Generalized Likelihood Ratio (GLR) test with duration penalty

- Performance
  - Test data: Hub4 97 evaluation, 3 hours

<table>
<thead>
<tr>
<th></th>
<th>Un-adapted WER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without SCD</td>
<td>22.7%</td>
</tr>
<tr>
<td>With SCD</td>
<td>21.4%</td>
</tr>
<tr>
<td>With reference segmentation</td>
<td>20.7%</td>
</tr>
</tbody>
</table>
Online Speaker Clustering

• Online speaker clustering
  – Causal process; decision cannot be changed later
  – Refines speaker boundaries detected by speaker change detection

• Performance
  – Test set: HUB4 98, 3-hours, 827 segments

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misclassification error rate (segment level)</td>
<td>28.5%</td>
</tr>
<tr>
<td>Cluster impurity (frame level)</td>
<td>8.6%</td>
</tr>
</tbody>
</table>
Online Speaker Adaptation

- Online speaker tracking to get speaker-homogeneous, cluster-labeled utterances
- Incremental storing and updating of adaptation parameters for each cluster
- Cluster-based adaptive decoding with adaptation parameters from previously processed utterances

<table>
<thead>
<tr>
<th>Language</th>
<th>Adaptation</th>
<th>WER</th>
<th>%gain</th>
<th>xRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>none</td>
<td>19.4</td>
<td>9.3</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>online</td>
<td>17.6</td>
<td></td>
<td>0.64</td>
</tr>
<tr>
<td>Arabic</td>
<td>none</td>
<td>16.4</td>
<td>7.9</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>online</td>
<td>15.1</td>
<td></td>
<td>0.53</td>
</tr>
<tr>
<td>Chinese</td>
<td>none</td>
<td>16.8</td>
<td>11.9</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>online</td>
<td>14.8</td>
<td></td>
<td>0.73</td>
</tr>
</tbody>
</table>
Sentence Boundary Detection

• Problem: Segment STT output into reasonable-length sentences for efficient Machine Translation
• Approach
  – Discriminative methods
    • Neural Network for acoustic features
    • Perceptron for linguistic features
  – Sentence length constraints
    • Statistical duration modeling
    • Maximum duration constraints
• Performance Comparison
  – Perceptron vs Hidden Event Modeling (HEM)

<table>
<thead>
<tr>
<th>Approach</th>
<th>SBD error rate (Arabic, WER=19.7%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEM</td>
<td>61.2%</td>
</tr>
<tr>
<td>Perceptron</td>
<td>46.0%</td>
</tr>
</tbody>
</table>

– Sentence Duration

<table>
<thead>
<tr>
<th>Approach</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without duration modeling</td>
<td>200s</td>
</tr>
<tr>
<td>With duration modeling</td>
<td>16s</td>
</tr>
</tbody>
</table>
• **Problem**
  – Locate names in speech

• **Approach**
  – Ergodic HMM models the name-class transition
  – Statistical bi-gram model within each class models emission probability of a word given the class

• **Performance**
  – Above 90% F-Measure on English text sources
  – Above 85% F-Measure on Broadcast News speech sources
Machine Translation

- Machine translation component from Language Weaver
  - State of the art statistical phrase translation approach
- Interface between STT and MT preserves named entities in the MT output.

Names in MT linked to STT source
BBN Broadcast Monitoring System

Automatic translation of transcript

Automatic transcription of speech

Real-time streaming video (~3 min delay)

Chinese Vice Premier Wen Jiabao met separately with Beijing China-Africa Cooperation Forum summit leaders of some African countries.

The EU countries starting at the airport more strict new safety standards please see below details.

President Hu Jintao presided over a welcoming ceremony here this afternoon in Beijing Republic of South Africa President Thabo Mbeki's visit to China.

The NPC Standing Committee Vice-Chairman Cheng Siwei State Councilor Tang Jiaxuan vice chairman of the CPPCC National Committee Luo Hachai attended today's welcoming ceremony.

President Hu Jintao presided over a welcoming ceremony here this afternoon in Beijing Republic of South Africa President Thabo Mbeki's visit to China.

The NPC Standing Committee Vice-Chairman Cheng Siwei State Councilor Tang Jiaxuan vice chairman of the CPPCC National Committee Luo Hachai attended today's welcoming ceremony.

President Hu Jintao presided over a welcoming ceremony here this afternoon in Beijing Republic of South Africa President Thabo Mbeki's visit to China.

The NPC Standing Committee Vice-Chairman Cheng Siwei State Councilor Tang Jiaxuan vice chairman of the CPPCC National Committee Luo Hachai attended today's welcoming ceremony.
Query in English Retrieves Multi-Lingual Results

Focus cues on search terms

Click in result to jump to streaming video for review

Saved queries monitor incoming video automatically
Speech Triage
Open-Domain Speech Challenges

- Uncontrolled recording conditions
- Informal colloquial speech
- Variety of input audio encodings and formats
- Error rates in STT and MT machine outputs are high
Approach – Translingual Speech Triage

- **Query by example**
  - Easy for users to specify (“more like this”)
  - System creates parallel queries from English and foreign-language passages

- **Supervised relevance feedback**
  - Users give both positive and negative supervised feedback
  - System creates an updated discriminative retrieval model

- **Iterative refinement**
  - User and system repeat cycle until retrieval results converge
Speech Triage Prototype

Search results returned in both languages

Simple text query used to begin iterative retrieval

Relevance judgments can be positive or negative
Relevance Ranked Retrieval with Feedback
Information Extraction from Speech Using Minimal Resources
Information Extraction using Minimal Resources

• What could you do with a few hours of speech from a new language with no transcripts?

• Limits utility to the “topic” view of information
  – Audio of interest “On-topic” and not of interest “Off-topic”
  – Examples
    • Speaker
    • Language
    • Words
    • Topics
System Architecture

Acoustic Unit Models

Speech

Recognition

Acoustic Units

Classifier

Classifier Model

On-Topic

Off-Topic
Recognition with Limited Resources

• **Universal phone recognition**
  – HMM recognizer with a large phone set
  – Built from a collection of corpora in different languages
  – Models may not match the language/domain of interest

• **Self-organizing recognizers**
  – Create/discover acoustic units from the available data
  – No cross domain issues
Acoustic Models

• **Frame-based Models**
  – No time sequencing information (e.g., Gaussian Mixture Model of cepstrum and derivatives)
  – Useful for speaker ID and language ID

• **Segmental Models**
  – Model sequences of frames
    • Parametric Trajectory Model of cepstral vectors
  – Useful for words, topics, as well as speaker and language ID
Training/Recognition

**Training**

1. Speech from Domain
2. Segmentation
3. Segment Clustering and Labeling
4. Training of Segmental Models

**Recognition**

1. Speech
2. Segmentation
3. Recognition Search
4. Acoustic Unit Labels
Classification

• Run recognizer on On-topic and Off-topic data
• Create features
  – e.g., n-gram counts of acoustic-unit labels for each class
• Train SVM (or other classifier) with feature vectors
• Recognize on acoustic utterance in question
  – Perform n-gram counts and classify
Experiment 1

• OnTopic: any of the following topics from Switchboard corpus – credit card use, vacation spots, buying a car
• Recognizer training
  – 3.5 hours of Switchboard data
  – 200 discovered acoustic units employed
• Classifier training
  – 110 5min conversations on-topic; 339 off-topic
  – Bigram features (with some feature selection based on frequency of occurrence and tf-idf)
  – SVM-light (Joachim) package employed; used with classification tree
• Performance
  – 24% EER using self-organizing recognizer
  – For comparison
    • 35% classification EER obtained using conventional English phone recognizer which had a 35% phone accuracy on Switchboard
Experiment 2: Arabic Broadcast News

- **OnTopic**: Cairo Radio broadcasts about Husni Mubarak
  - Broadcasts were nominally about 2 minutes duration
- **Recognizer training**
  - 1.5 hours of domain data
  - 50 acoustic units
- **Classifier training**
  - 1.4 hours on-topic; 3.4 hours off-topic
  - 5-gram features
- **Performance**
  - 19% EER
- **Best Features**
  - Two 5-grams of acoustic units that were variants of the name Mubarak being spoken
Waveforms-Mubarak
Spectrograms-Mubarak

Mubarak 1  Mubarak 2
Advanced Extraction Technologies
Advanced Extraction Technologies

• The Information Extraction Landscape
  – Names
  – Entities
  – Relations
  – Events

• Applications
  – (Semi-)Automatic fill of data bases from language sources
    • Social Network Analysis
    • Organizational structure & leadership
    • Expert systems/reasoning systems
  – Improved translation, summarization, and detection algorithms
American and Iraqi forces have captured a former member of Saddam Hussein's government and two of the dictator's relatives. The relatives, Abdullah Maher Abdul Rashid and his cousin Marwan Taher Abdul Rashid, were seized on March 8 in Tikrit, Mr. Hussein's hometown. Marwan Taher Abdul Rashid worked as a bodyguard for Mr. Hussein. Abdullah Maher Abdul Rashid is a brother-in-law of Mr. Hussein's son Qusay, who was killed in a battle with American soldiers in July 2003. The third man arrested, Omar Hassan Chiad, was an official in Mr. Hussein's government and was caught by American soldiers.
American and Iraqi forces have captured a former member of Saddam Hussein's government and two of the dictator's relatives. The relatives, Abdullah Maher Abdul Rashid and his cousin Marwan Taher Abdul Rashid, were seized on March 8 in Tikrit, Mr. Hussein's hometown. Marwan Taher Abdul Rashid worked as a bodyguard for Mr. Hussein. Abdullah Maher Abdul Rashid is a brother-in-law of Mr. Hussein's son Qusay, who was killed in a battle with American soldiers in July 2003. The third man arrested, Omar Hassan Chiad, was an official in Mr. Hussein's government and was caught by American soldiers.

<table>
<thead>
<tr>
<th>ID</th>
<th>Type</th>
<th>Names</th>
<th>Other Mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-25</td>
<td>PER</td>
<td>“Abdullah M. A. Rashid”</td>
<td>“his”, “brother-in-law of Qusay”</td>
</tr>
<tr>
<td>E-26</td>
<td>PER</td>
<td>“Marwan T. A. Rashid”</td>
<td>“his cousin”, “bodyguard”</td>
</tr>
<tr>
<td>E-54</td>
<td>PER</td>
<td>“Qusay”</td>
<td>“Mr. Hussein’s son”</td>
</tr>
<tr>
<td>E-29</td>
<td>PER</td>
<td>“Omar Hassan Chiad”</td>
<td>“an official in Mr. Hussein’s government”</td>
</tr>
</tbody>
</table>
American and Iraqi forces have captured a former member of Saddam Hussein's government and two of the dictator's relatives. The relatives, Abdullah Maher Abdul Rashid and his cousin Marwan Taher Abdul Rashid, were seized on March 8 in Tikrit, Mr. Hussein's hometown. Marwan Taher Abdul Rashid worked as a bodyguard for Mr. Hussein. Abdullah Maher Abdul Rashid is a brother-in-law of Mr. Hussein's son Qusay, who was killed in a battle with American soldiers in July 2003. The third man arrested, Omar Hassan Chiad, was an official in Mr. Hussein's government and was caught by American soldiers.

<table>
<thead>
<tr>
<th>ID</th>
<th>Relation Type</th>
<th>Argument 1</th>
<th>Argument 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-34</td>
<td>ORG-affiliation.employment</td>
<td>E-18 “forces”</td>
<td>E-28 “American”</td>
</tr>
<tr>
<td>R-27</td>
<td>ORG-affiliation.employment</td>
<td>E-18 “forces”</td>
<td>E-84 “Iraqi”</td>
</tr>
<tr>
<td>R-83</td>
<td>ORG-affiliation.employment</td>
<td>E-29 “Omar”</td>
<td>E-35 “government”</td>
</tr>
<tr>
<td>R-57</td>
<td>Personal.family</td>
<td>E-34 “relatives”</td>
<td>E-75 “Hussein”</td>
</tr>
<tr>
<td>R-29</td>
<td>Physical.located</td>
<td>E-34 “relatives”</td>
<td>E-79 “Tikrit”</td>
</tr>
<tr>
<td>R-42</td>
<td>GPE-affiliation.origin</td>
<td>E-75 “Hussein”</td>
<td>E-79 “Tikrit”</td>
</tr>
</tbody>
</table>
American and Iraqi forces have captured a former member of Saddam Hussein's government and two of the dictator's relatives. The relatives, Abdullah Maher Abdul Rashid and his cousin Marwan Taher Abdul Rashid, were seized on March 8 in Tikrit, ... Qusay, who was killed in a battle with American soldiers in July 2003.

<table>
<thead>
<tr>
<th>ID</th>
<th>Event Type / Roles</th>
<th>Agent</th>
<th>Person</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>V-34</td>
<td>Justice.arrest</td>
<td>E-18 &quot;forces&quot;</td>
<td>E-29 “Omar”, E-34 “relatives”</td>
<td>Past</td>
</tr>
<tr>
<td>V-52</td>
<td>Justice.arrest</td>
<td></td>
<td>E-34 “relatives”</td>
<td>2005-03-08</td>
</tr>
<tr>
<td>V-27</td>
<td>Life.die</td>
<td></td>
<td>E-54 “Qusay”</td>
<td>2003-07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E-63 “soldiers”</td>
<td></td>
</tr>
</tbody>
</table>
Levels of Linguistic Analysis

Name finding
Parsing
Co-reference

Person: Slobodan Milosevic
Position: president
Organization: Yugoslavia

Person: Milos Milosavljevic
Position: President
Organization: Association of Yugoslav Banks

Person: Milos Milosavljevic
Position: General director
Organization: JugoBanka
Approaches

• Trained methods now commonly used
• For names and non-nested entity mentions, tagging models can be used
  – For example: Person-Name-Start, Organization-Name-Continue, Not-A-Name
• Coreference requires combining non-local information
  – Greedy strategies are common, processing each mention against all entities seen so far
  – Graph partitioning algorithms are also being explored
• Relations and events are typically predicted using classifier models based on contextual features:
  – For “the Bush ranch in Texas”, features would include “ranch” as a Facility word, “Texas” as GPE, and “in” as the linking word
Performance

- Best system scores from the ACE-2004 evaluation
Combining Recognition and Extraction

![Graph showing ACE Value Score for Entities and Relations]

- **Entities**: 78.7
  - Transcripts: 47.5
  - ASR: 16.0
- **Relations**: 44.4

Legend:
- **Transcripts**
- **ASR**