Meeting Recorder: Audio Processing

Dan Ellis
<dpwe@ee.columbia.edu>

LabROSA, Columbia University
and ICSI, Berkeley

Outline

1  ICSI Meeting Recorder
2  Close-mics: cancellation & turn estimation
3  Tabletop mics: turns & speaker location
4  Visualization tools
5  Future Work
ICSI Meeting Recorder data
(with UW, SRI, IBM, Columbia)

- Microphones in conventional meetings
  - for summarization/retrieval/behavior analysis
  - informal, overlapped speech

- Data collection (ICSI, UW, ...):
  - 100 hours collected, ongoing transcription

- NSF ‘Mapping Meetings’ project
  - also interest from NIST, DARPA
Data from the ICSI project

- 16 channels @ 16 kHz, 16 bit
- Preprocessing
  - high-pass filter!
  - 64 sample skew!
Close-mic channels

- Crosstalk
- Speaker activity detection
Impulse response coupling

- Cross-correlation recovers impulse response
  
  ![Example cross coupling response, chan3 to chan0](image)

- Coupling to each mic gives motion
  
  ![Spkr C](image)
Speaker Activity Detection
(with Sam Keene)

- Noisy crosstalk model: \( \mathbf{m} = \mathbf{C} \cdot \mathbf{s} + \mathbf{n} \)
- Estimate subband \( \mathbf{C}_{\mathbf{xA}} \) from A’s peak energy
  - i.e. ‘sparsity’ assumption
  - ... then linear inversion to recover speaker act.
- 20 subband crosstalk gains for each spkr \( \times \) mic
Tabletop mics: Turn detection

- 4 mics ~ 1m separated along center of table
  - 3 timing differences
  - slight L/R offset to disambiguate

- Hi-res cross-correlation for timings
  - use normalized peak value for confidence
  - cluster results
Speaker localization
(with Huan Wei Hee)

- **Timing differences** \(\rightarrow\) speaker positions \((x,y,z)\)

  - Gradient descent on implied \(\Delta t\)s

- **Ambiguity:**
  - Mic positions not fixed
  - Speaker motions

- **Iterative estimation of speaker, mic locations**
Visualization: transPlotter

- Speaker turn *patterns* are informative

- Browser for ‘high-level’ view, quick examination
  - snack, iwidgets based
  - public release
Meeting IR tool

- IR on (ASR) transcripts from meetings

Meeting IR Tool

IR Status: idle

Enter query: transcriber

Results for: transcriber

<table>
<thead>
<tr>
<th>Meeting</th>
<th>Channel</th>
<th>Date</th>
<th>Offset</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>mr2_u</td>
<td>3</td>
<td>2000feb16</td>
<td>11:39</td>
<td>having getting hold of a transcriber...</td>
</tr>
<tr>
<td>mr2_u</td>
<td>2</td>
<td>2000feb16</td>
<td>25:17</td>
<td>because the transcriber...</td>
</tr>
<tr>
<td>mr2_u</td>
<td>1</td>
<td>2000feb16</td>
<td>36:06</td>
<td>I'm going to go download transcriber...</td>
</tr>
<tr>
<td>mr2_u</td>
<td>3</td>
<td>2000feb16</td>
<td>25:30</td>
<td>one um this is called transcriber um binary</td>
</tr>
</tbody>
</table>

Meeting: mr2_u  
Date: 2000feb16  
File: mr2_u-c3

Jane: having -
Jane: getting hold of a transcriber. I think that Oogosi has one.
Adam: Hm-hum.
Eric: Yeah.
Jane: And I could probably borrow one.
Eric: Well.
Don: I-I-I actually have one.

- repurposed from Thisl project
Future work

- Speaker turns
  - evaluation of close-mic system
  - speaker characteristics for tabletop mics

- Nonspeech events
  - unsupervised clustering of audio
  - finding the feature space...

- Speech fragment recognition
  - missing-data recognition based on ‘good’ signal
  - recognition of overlapping voices

- High-level browsing
  - the ‘meeting map’ concept
  - summarization