
Meeting Recorder: Audio Processing

Dan Ellis
<dpwe@ee.columbia.edu>

Lab**ROSA**, 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

1

ICSI Meeting Recorder data

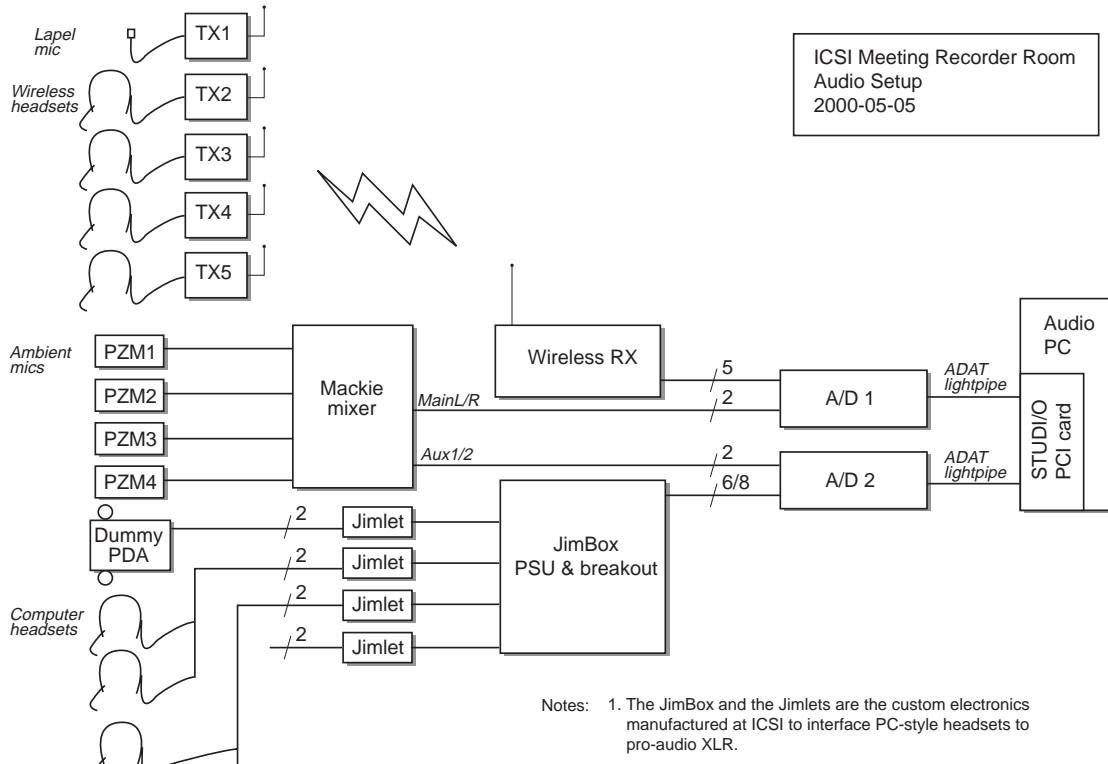
(with UW, SRI, IBM, Columbia)

- **Microphones in conventional meetings**
 - for summarization/retrieval/behavior analysis
 - informal, overlapped speech
- **Data collection (ICSI, UW, ...):**

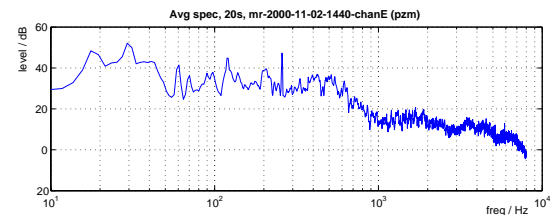


- 100 hours collected, ongoing transcription
- **NSF ‘Mapping Meetings’ project**
 - also interest from NIST, DARPA

Data from the ICSI project

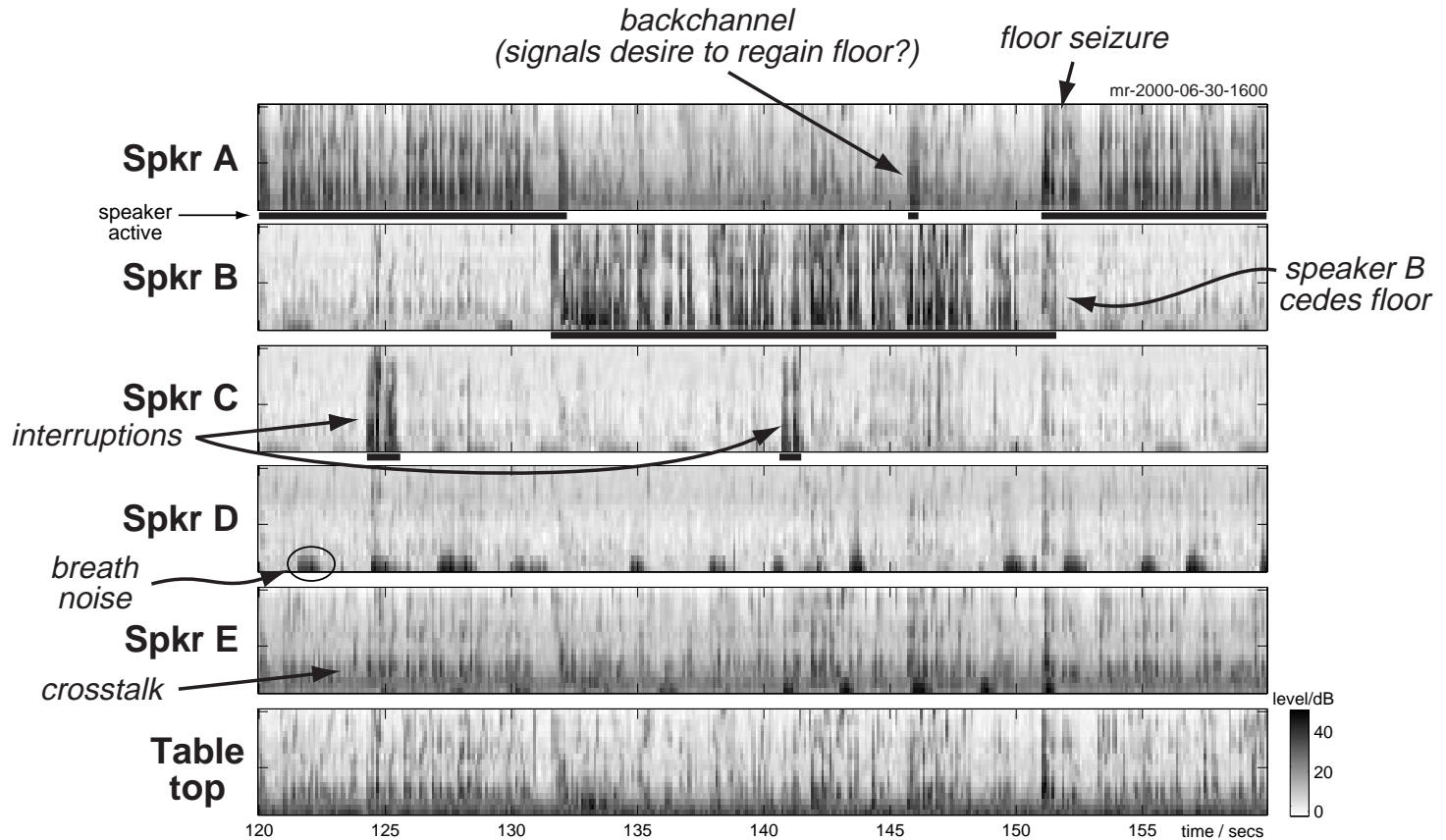


- **16 channels @ 16 kHz, 16 bit**
- **Preprocessing**
 - high-pass filter!
 - 64 sample skew!



2

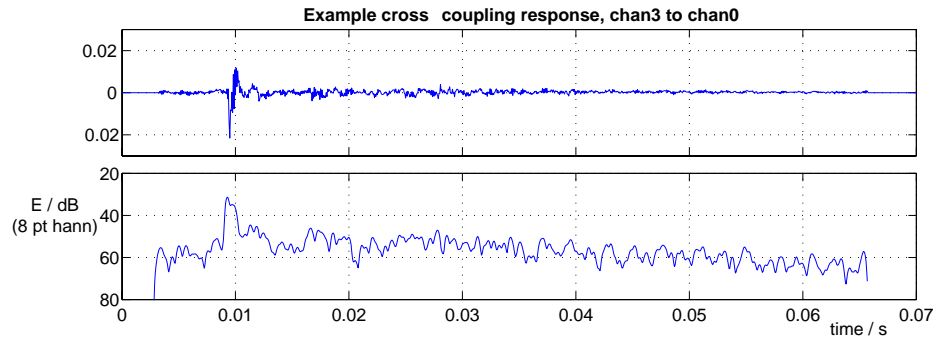
Close-mic channels



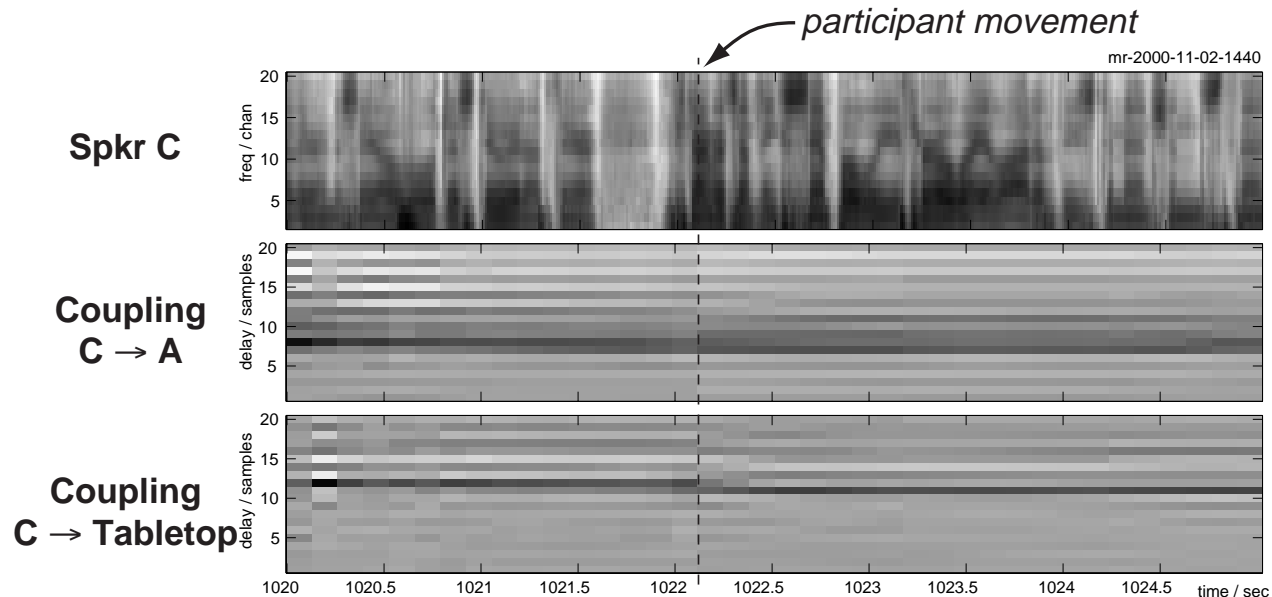
- **Crosstalk**
- **Speaker activity detection**

Impulse response coupling

- **Cross-correlation recovers impulse response**



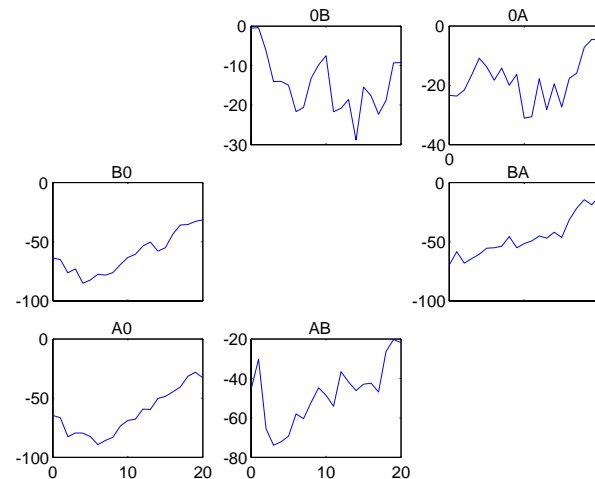
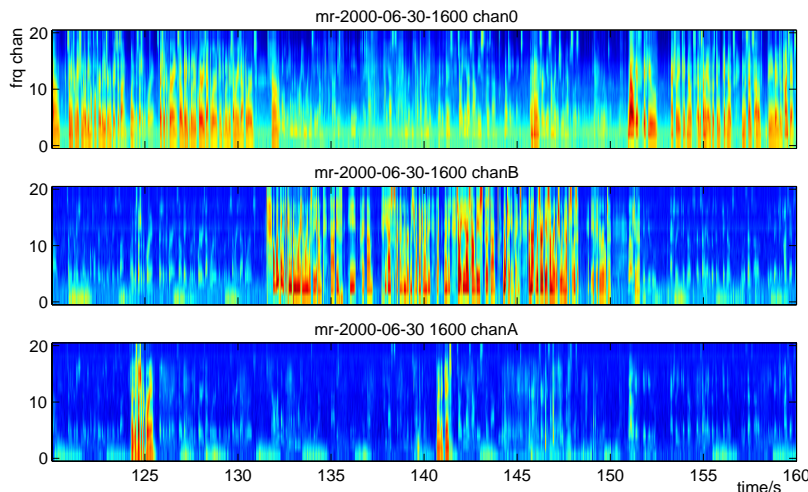
- **Coupling to each mic gives motion**



Speaker Activity Detection

(with Sam Keene)

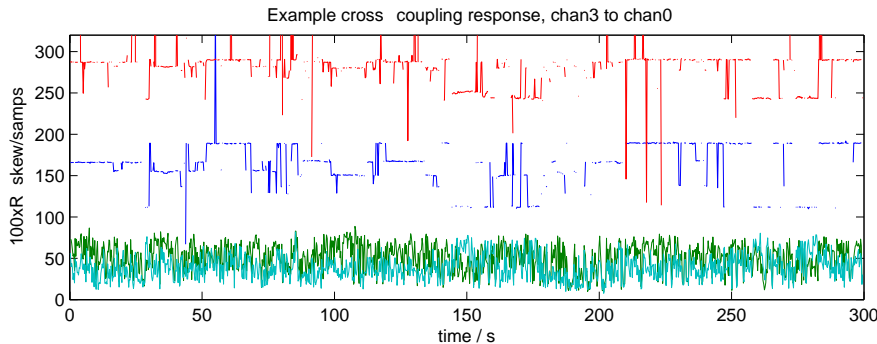
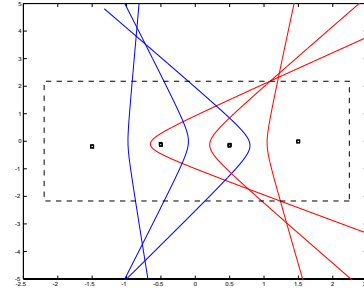
- **Noisy crosstalk model: $\mathbf{m} = \mathbf{C} \cdot \mathbf{s} + \mathbf{n}$**
- **Estimate subband \mathbf{C}_{xA} from A's peak energy**
 - i.e. 'sparsity' assumption
 - ... then linear inversion to recover speaker act.
- **20 subband crosstalk gains for each spkr x mic**



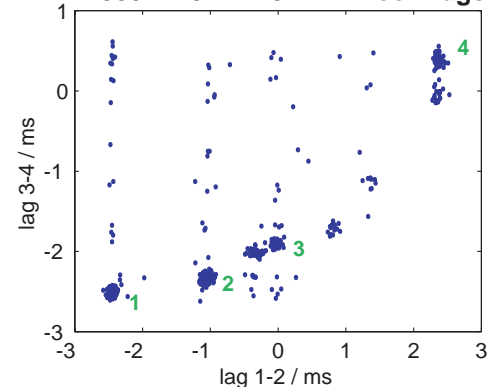
3

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



mr-2000-11-02-1440: PZM xcorr lags

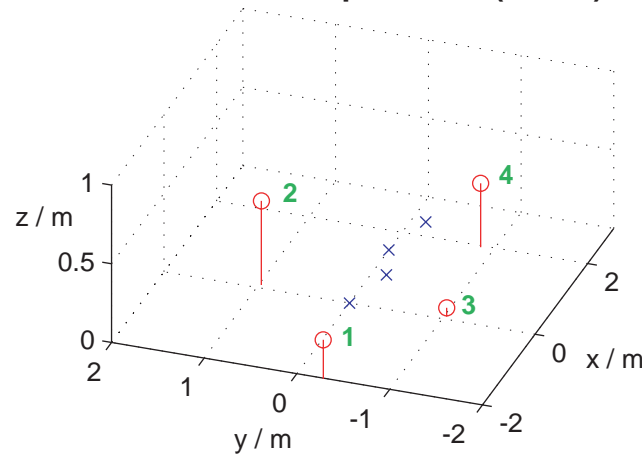


Speaker localization

(with Huan Wei Hee)

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

Inferred talker positions (x =mic)

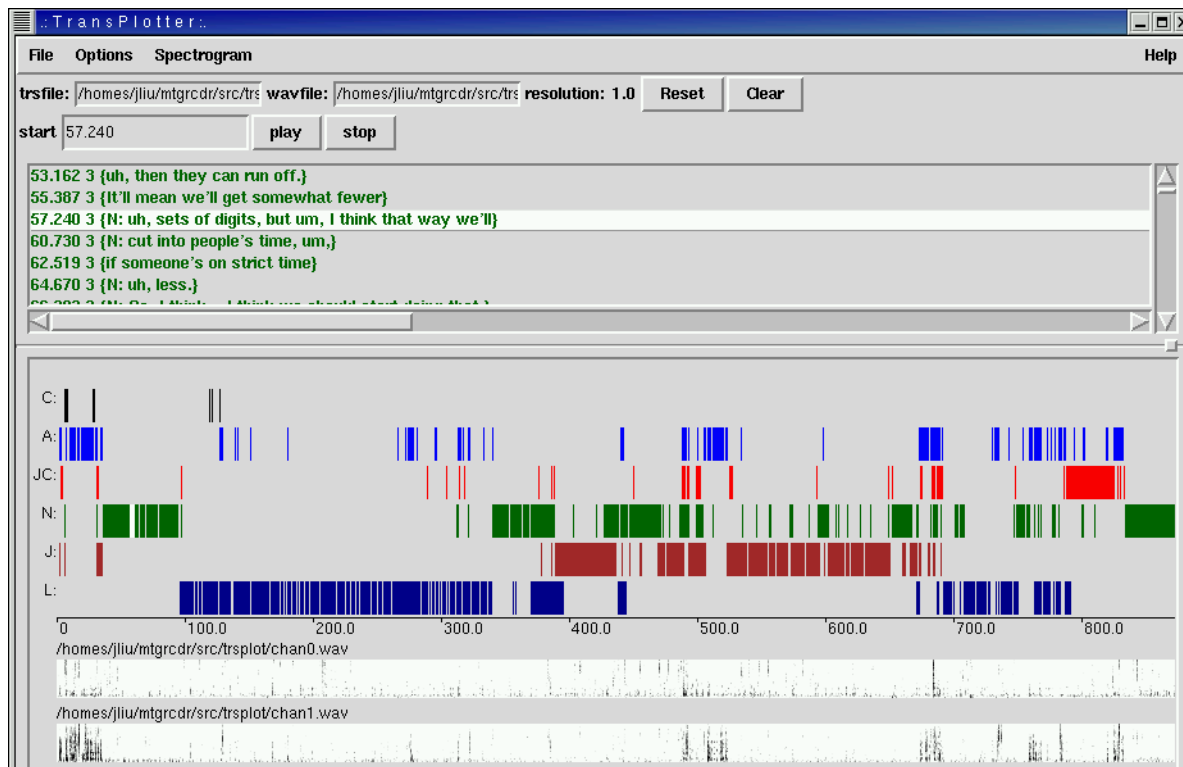


- gradient descent on implied Δt s
- **Ambiguity:**
 - mic positions not fixed
 - speaker motions
- **Iterative estimation of speaker, mic locations**

3

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:

Enter query: Use ASR output History

Results for:

	Meeting	Channel	Date	Offset	Context	
<input type="checkbox"/>	mr2_u	3	2000feb16	11:39	having getting hold of a transcriber i think tl	
<input type="checkbox"/>	mr2_u	2	2000feb16	25:17	because the transcriber's put in intonational	
<input type="checkbox"/>	mr2_u	1	2000feb16	39:06	i'm going to go download transcriber and ja	
<input type="checkbox"/>	mr2_u	3	2000feb16	25:38	one um this is called transcriber um binary	

Meeting: Date: File:

Jane: having -
Jane: getting hold of a transcriber. I think that Cogsci has one.
Adam: Mm-hmm.
Eric: Yeah.
Jane: And I could probably borrow one.
Eric: Well,
Dan: 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