A Thought Experiment

- What if we were to record all our conversations?
  - The technology is there!

- The usual objections…
  - Privacy, privacy, privacy
  - “Not everything you say is that important”
  - But really: how would you navigate all that audio?
We’re Not the First to Raise This Issue

- Meeting Summarization Task
  - Premise: Meetings are Important
  - Controlled Environment
    - Did you clip on your microphone?
  - DIASUMM system (CMU)
    - Turn segmentation (one mic per person)
    - Topic segmentation
Why Not Just Record Meetings?

- Because We Talk All the Time
  - Many important discussions are at lunch, in the hallway, at the water cooler, etc.
  - We can’t predict when the important idea or reference will occur, and we may not have a way of jotting it down
Is This So Crazy After All?

- Some people have this problem every day
  - Doctors, Lawyers, Journalists, Ethnographers
  - Current solutions are expensive and don’t scale
- Many of the rest of us do too!
Other Speech Summarization Work

- **Speech Summarization**
  - Christensen – use opening sentences
  - Koumpis and Renals – per-word classifier
  - He et al. – involve usage data
  - Maskey and Hirschberg – summary from audio

- **Meeting Summarization**
  - DIASUMM (from earlier slide)
  - Diarization (many groups)

- **Meeting Understanding**
  - Patrick’s talk (next!)
Our Goals

- **This Work: Browse an Individual Conversation**
  - Where the conversation is long
  - Where the user was a participant
  - Where the speech recognition is noisy
  - Where turn segmentation is not available or too noisy

- **Enable the User To:**
  - See and hear the whole conversation at once
  - See portions in details when necessary
  - Quickly get a sense for topics and regions
Our Approach

- Inspiration: Zoomable User Interfaces (ZUIs)
  - Pioneered by Ben Bederson and colleagues
  - Concept: navigate large bodies of information by using multiple scales

- Goals of our interface
  - Continuous zooming from the entire conversation down to entire transcript
  - Mantra: “The Audio is the Document” - use text for scanning but audio for content
  - Make it clickable, playable, movable, draggable

- But what does zooming mean for audio?
Zooming Out and Zooming In

Zoomed out: five minutes

- Second period it’s just the second period and the special chop. The parents that like the morning of the first year or the first year's brides or more motivated sorry, two second year students in their study abroad class. They're interested in studying actually. They're talking they're activated computers and translations and... thing like that. They're interested in studying actually. They're talking they're activated computers and translations and. They're interested in studying actually. They're talking they're activated computers and translations and... thing like that.

Zoomed in: less than one minute

- It's just the second period and the special chop. The parents that like the morning of the first year or the first year's brides or more. They're interested in studying actually. They're talking they're activated computers and translations and. They're interested in studying actually. They're talking they're activated computers and translations and... thing like that.
"Why Does the Text Look All Crazy?"

the vertical position is the time of the phrase

the horizontal position is the order of appearance in the conversation

this approach strikes a balance between showing the separation of key phrase occurrences in time and making things readable when completely zoomed in.
A Look at the Whole Interface

Text pane (draggable, scrollable)

Timeline

Zoom bar

Play cursor

Topic Boundary

Audio transport controls
Why Not Just Show the Transcript?

- No turn information
  - One block of text
- Reading noisy recognition makes your brain hurt
  - One block of hard-to-read text
- A one-hour conversation is a very long transcript
  - One very long block of hard-to-read text
How We Do What We Do Do

Audio

Speech Recognizer

Topic Segmenter

Keyword Extraction and Ranking

Rendering the User Interface
Speech Recognition

- How Noisy Is It?
  - 75% for conversational speech
  - An example:
    - Announcer: “The Buick Enclave, the finest luxury crossover ever. Visit Buick.GM.ca to see that where there’s passion, there’s beauty.”
    - Recognizer: “the buick on clay the finest luxury crossover ever visited you like donkey and don't see any winners pension there's beauty”
  - Confidence is of limited use
Dealing with the Noise

- Typical Problems
  - Misrecognized/misgrouped words
  - No punctuation or turns
  - Natural Language tools break down!

- Overcome this by using **keywords and timing** to index **audio**
  - High TFIDF words are less likely to be wrong
Topic Segmentation

- Notion of “topic” ill-defined for conversational speech
- We tried many text topic detection / segmentation approaches with little success
- We used an approach that worked well on broadcast news, similar to TextTiling
  - Trained on news data with clear topics
  - Classified each point in time as a boundary/non-boundary
  - Features: lexical, prosodic, not news-specific
Choosing Keywords

- We Need a Ranked List of Keywords
- Easy: rank unigrams by TFIDF
  - But, unigrams alone are of limited value
- We can compute bigram TFIDFs too
  - But, they have a different numeric scale
- We have a fancy way of putting them into the same numeric scale
  - But, the details are in the paper and wouldn’t add much to the talk
Deciding What Words to Render

- Default Plan of Showing Highly-Ranked Words Fails
  - Need the context of neighboring words
  - Go down ranked keyword list, mark all occurrences, as well as neighbors
  - Increase word density with an exponential characteristic, to go from scanning to reading

Term List
- English
- Spanish
- mr workshops
- Sunday

we’re working with the and the Sunday were aim to me and his and two teams kids are here mhm and uh going to L.A. first are you doing services Sunday july first in lancaster and the spanish church and so he yeah to be able to switch from english to spanish english mhm without influencing too much and so we’ve been working really hard on both on nancy shun mhm but yeah so it he he just went to nashville so oh two they’re ethnic saying no to that works out mr workshops all the all the typewriter I see oh yeah yeah so that’s where branch was that was so adamant coming back with them I
but even you know reasoning for the parents that received um last and um and we've got of more parents yeah but um um and now we're
to it to me and on the nights that they they have late afternoon well i haven't english department meeting on the eighth but that would be better to
the morning of like the like before noon on the ninth yeah around noon or okay put down to me and then then need that and d. um to put this
is to organize it i wanted to organize that um c. and then uh mhm can my group down there the last time from them or
graduate that was just over there and she said when she talked to him but not many um interested in studying and you were i'm not very motivated group yeah depressing second year students yeah i be the first year we have eighteen in this study abroad class me two second year students in their class this but i don't know where i'm not i haven't talked to them yet they would be warehouse kids
anything i don't know little better whether i need to i could fly the morning and in help next morning now know what would what would work best
native speakers cost
depending on languages um three groups actually is you can call any place in the world certain languages mhm you american english you what um
american english and that japanese and um arabic injection arabic not arabic korean mhm and then some that they can call any place in the united
language group mhm had an arabic church that was years before retreat this last weekend gave them the they could call and list you this and um on the some of them worked american arabic readers and the
and and they had a in there and it's had an they pay thirty minutes now and nothing unbearable they've got something yeah mhm how we're about this i have to get a car heard about it was and he
and it calls they pay ten bucks um well i thought that might the students it was too and recruit students and...
User Study: Goals

- Goal: Test Effectiveness for Browsing
  - Unfortunately, testing for browsing is difficult
  - Typical approach: information retrieval task
  - Our approach: IR, but with a few twists to encourage browsing:
    - We asked our users to pretend to be reporters looking for a quote
    - We had the users listen to the audio a few days beforehand to simulate being involved in the conversation
    - We created questions that didn’t contain content words
User Study: Details

- 10 Subjects
- Two 15-minute Conversations
- Six Questions Each
- Two Interfaces (ours and control)
- All Relevant Factors Randomized
Control Condition

- Scrollable, playable full transcript
Was Our Ploy Successful?

- Sort Of…
  - Users still uniformly begged us for a “search box”
  - Some users only listened to the audio the morning of
  - Even half hour conversations were too long in our pilots
  - Users had no “stake” in the conversations
Qualitative Results

- **Cheers**
  - Users uniformly preferred our interface and would use it again
  - Felt “in control” of information; more manageable
  - Felt that they would browse their own conversations if they had this interface

- **Jeers**
  - Didn’t like listening to irrelevant conversations
  - Didn’t like slanting of text
    - “the text embodiment of mental illness”
  - The “80’s-line-printer” look was not universally loved
How Did Users Use the UI?

- Users zoom in and out to go from context to detail!
Quantitative Results

- Task completion time
  - ANOVA with interfaces, users, and questions
  - Small speedup with low significance \( p=0.3 \)

<table>
<thead>
<tr>
<th>Interface</th>
<th>Time per Answer (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scalable (our method)</td>
<td>76.1</td>
</tr>
<tr>
<td>Non-Scalable</td>
<td>85.7</td>
</tr>
</tbody>
</table>

- Hypothesis: conversations were too short

- UI Instrumentation (next slide)
Conclusions and Future Work

- For the Present
  - Our systems seems to make conversation browsing more manageable
  - Next steps including testing on an audience that needs this for everyday tasks (reporters, ethnographers, etc.)

- For the Future
  - Automatically collect, then browse all conversations
  - Automatic segmentation in poor recording conditions
FIN
UI Pilots and Design Iterations

- If You Think It Looks Bad Now...
  - Putting the beautiful UI you slaved over in front of cruel, cruel users is an important (though painful) process
  - Still room for improvement