

Note: you may use slides/images from this presentation in your own work, but make sure to attribute the source as: S. Basu, S. Gupta, M. Mahajan, P. Nguyen, and J. Platt. “Scalable Summaries of Spoken Conversations.” *In Proceedings of Intelligent User Interfaces 2008 (IUI’08)*.

SCALABLE SUMMARIES OF SPOKEN CONVERSATIONS

Sumit Basu, Surabhi Gupta, Milind Mahajan,
Patrick Nguyen, and John Platt

Microsoft Research

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



DIASUMM

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

6:00 just second period it's just the second period and the this special chop the parents that
7:00 like the morning of the before noon on around noon or
8:00 um interested in studying and you very motivated group depressing second year students yeah she talked to him but
9:00 the first year so it the first year brides or more motivated sorry two second year students in their haven't talked to them yet
10:00 they're interested in studying actually studying abroad they're in the morning and help
certain languages mhm what american english you um on languages um you can call any place in the
they can call any place in the an arabic church that american english and japanese um arabic injection arabic not arabic korean
they're trying they're activated computers and translations and

Zoomed out: five minutes

8:00 second year students yeah i a my first be the first year we have about seventeen or eighteen in this study abroad class me the first year so it it and they
used and it's the first year brides or different mean they um they seem to be more motivated sorry could it's true that i'm not sure how many there will
be in a second although there are two second year students in their class this year um-hum so um but i don't know where i i'm not i haven't talked to
them yet to see where they would be warehouse kids are interest that are just what hm um well you can can uh talk with them and see if their uh
whether they're interested in studying actually studying abroad they're planning on it or anything i don't know little better whether i need to have two
days there yeah more

Zoomed in: less than one minute

“Why Does the Text Look All Crazy?”

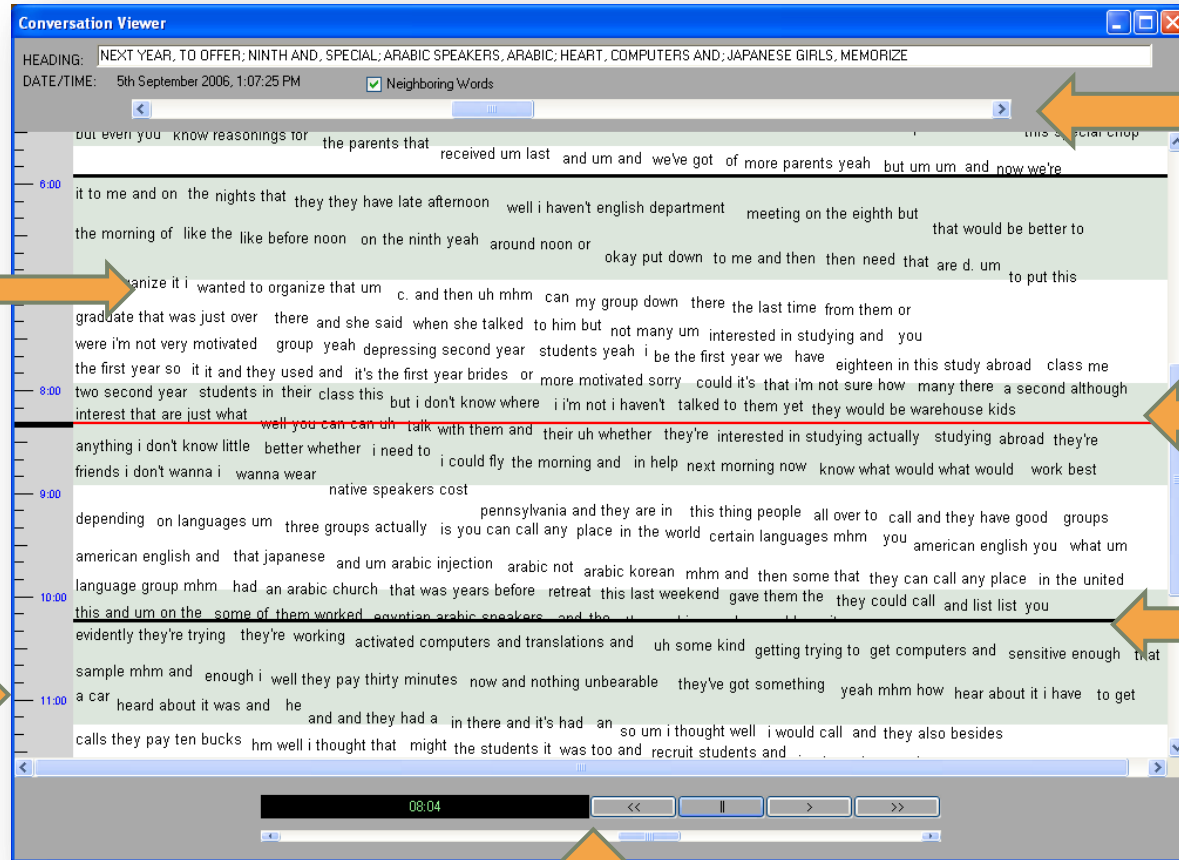
the vertical position is the time of the phrase

it to me and on the nights that they they have late afternoon well i haven't
the morning of like the like before noon on the ninth year around noon or
is to organize it i wanted to organize that um c. and then uh mhm can my
graduate that was just over there and she said when she talked to him but
were i'm not very motivated group yeah depressing second year students
the first year so it it and they used and it's the first year brides or more m
two second year students in their class this but i don't know where i i'm n
interest that are just what

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



Zoom bar

Play cursor

Topic Boundary

Audio transport controls

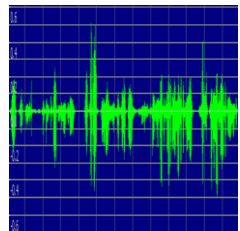
Text pane
(draggable,
scrollable)

timeline

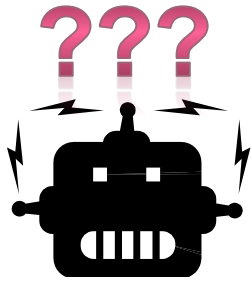
Why Not Just Show the Transcript?

- 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



Audio



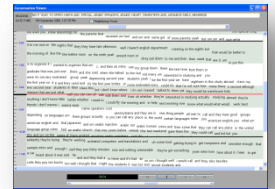
Speech
Recognizer



Topic
Segmenter



Keyword
Extraction
and Ranking

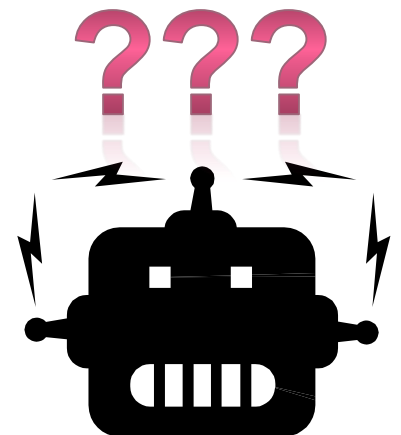


Rendering
the User
Interface

Speech Recognition

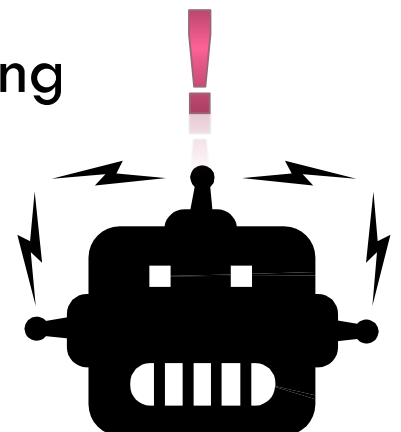
□ How Noisy Is It?

- 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



up here uh but hm well **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

Conversation Viewer

HEADING: NEXT YEAR, TO OFFER; NINTH AND, SPECIAL: ARABIC SPEAKERS, ARABIC; HEART, COMPUTERS AND; JAPANESE GIRLS, MEMORIZE

DATE/TIME: 5th September 2006, 1:07:25 PM

Neighboring Words

6:00 but even you know reasonings for the parents that received um last and um and we've got of more parents yeah but um um and now we're

7:00 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 are d. urn to put this

8:00 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 the first year so it it and they used and it's the first year brides or more motivated sorry could it's that i'm not sure how many there a second although two second year students in their class this but i don't know where i i'm not i haven't talked to them yet they would be warehouse kids interest that are just what well you can can uh talk with them and their uh whether they're interested in studying actually studying abroad they're 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

9:00 friends i don't wanna i wanna wear native speakers cost

10:00 depending on languages um three groups actually pennsylvania and they are in this thing people all over to call and they have good groups 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 list you this and um on the some of them worked, evation arabic speakers, and the

11:00 evidently they're trying they're working activated computers and translations and uh some kind getting trying to get computers and sensitive enough that sample mhm and enough i well they pay thirty minutes now and nothing unbearable they've got something yeah mhm how hear about it i have to get a car heard about it was and he and and they had a in there and it's had an so um i thought well i would call and they also besides calls they pay ten bucks hm well i thought that might the students it was too and recruit students and

08:04

Demo

User Study: Goals

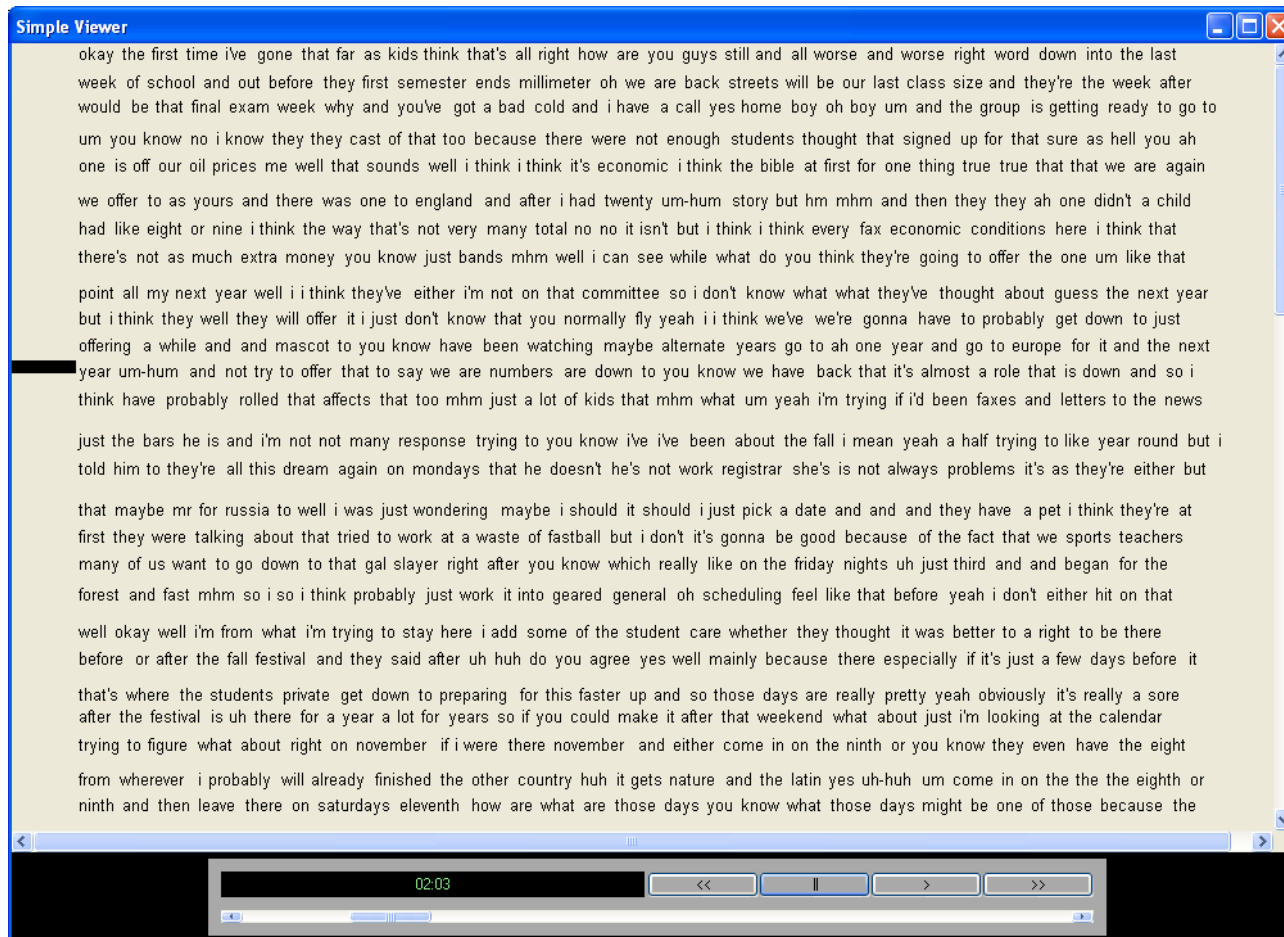
- Goal: Test Effectiveness for Browsing
 - Unfortunately, testing for browsing is difficult
 - Typical approach: information retrieval task
 - Our approach: IR, but with a few twists to encourage browsing:
 - We asked our users to pretend to be reporters looking for a quote
 - We had the users listen 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



Simple Viewer

okay the first time i've gone that far as kids think that's all right how are you guys still and all worse and worse right word down into the last week of school and out before they first semester ends millimeter oh we are back streets will be our last class size and they're the week after would be that final exam week why and you've got a bad cold and i have a call yes home boy oh boy um and the group is getting ready to go to um you know no i know they they cast of that too because there were not enough students thought that signed up for that sure as hell you ah one is off our oil prices me well that sounds well i think i think it's economic i think the bible at first for one thing true true that that we are again we offer to as yours and there was one to england and after i had twenty um-hum story but hm mhm and then they they ah one didn't a child had like eight or nine i think the way that's not very many total no no it isn't but i think i think every fax economic conditions here i think that there's not as much extra money you know just bands mhm well i can see while what do you think they're going to offer the one um like that point all my next year well i i think they've either i'm not on that committee so i don't know what what they've thought about guess the next year but i think they well they will offer it i just don't know that you normally fly yeah i i think we've we're gonna have to probably get down to just offering a while and and mascot to you know have been watching maybe alternate years go to ah one year and go to europe for it and the next year um-hum and not try to offer that to say we are numbers are down to you know we have back that it's almost a role that is down and so i think have probably rolled that affects that too mhm just a lot of kids that mhm what um yeah i'm trying if i'd been faxes and letters to the news just the bars he is and i'm not not many response trying to you know i've i've been about the fall i mean yeah a half trying to like year round but i told him to they're all this dream again on mondays that he doesn't he's not work registrar she's is not always problems it's as they're either but that maybe mr for russia to well i was just wondering maybe i should it should i just pick a date and and and they have a pet i think they're at first they were talking about that tried to work at a waste of fastball but i don't it's gonna be good because of the fact that we sports teachers many of us want to go down to that gal slayer right after you know which really like on the friday nights uh just third and and began for the forest and fast mhm so i so i think probably just work it into geared general oh scheduling feel like that before yeah i don't either hit on that well okay well i'm from what i'm trying to stay here i add some of the student care whether they thought it was better to a right to be there before or after the fall festival and they said after uh huh do you agree yes well mainly because there especially if it's just a few days before it that's where the students private get down to preparing for this faster up and so those days are really pretty yeah obviously it's really a sore after the festival is uh there for a year a lot for years so if you could make it after that weekend what about just i'm looking at the calendar trying to figure what about right on november if i were there november and either come in on the ninth or you know they even have the eight from wherever i probably will already finished the other country huh it gets nature and the latin yes uh-huh um come in on the the the eighth or ninth and then leave there on saturdays eleventh how are what are those days you know what those days might be one of those because the

02:03

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

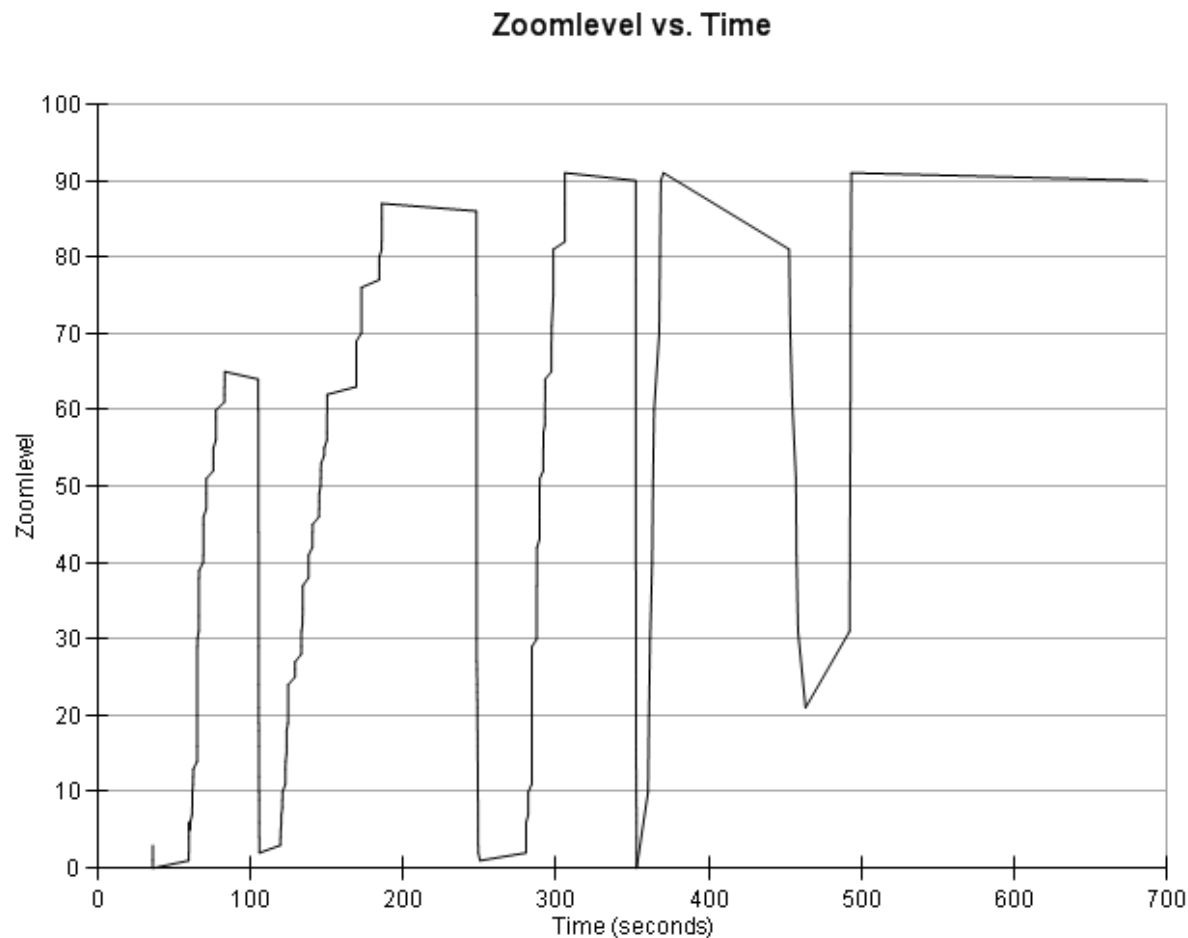
- Users uniformly preferred our interface and would use it again
- Felt “in control” of information; more manageable
- Felt that they would browse their own conversations if they had this interface

□ 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 $p=0.3$

Interface	Time per Answer (sec)
Scalable (our method)	76.1
Non-Scalable	85.7

- ▣ 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

