

The Meeting Recorder Project at ICSI

Jane Edwards, Adam Janin, Barbara Peskin and **Chuck Wooters**

Other project participants:

Jeremy Ang, Don Baron, Dan Ellis, David Gelbart, Nelson Morgan,

Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke

International Computer Science Institute

Berkeley, CA

USA



Outline

- Background
- Current Status
- Data Collection Process
- Transcription Effort
- Ability to contribute data
- Open Issues
- References



Background

Goal: to develop technology to “process” spoken language from meetings

- Speaker change detection, speaker tracking.
- SpeechCorder handheld portable device
- Information retrieval
- Dialog analysis/modeling
- Speech recognition
 - Far-field, digits, conversations, etc.



Background (cont.)

- Collaboration with UW, OGI, IBM, SRI, Columbia U.
- Collecting data since Feb 2000
- Goal is 100 meeting-hours



Background (cont.)

Types of meetings we are collecting:

- Regular, weekly group meetings
- “Natural” data (meetings that would happen even if we weren’t recording)
- Close-talking and far-field microphones
- Digits (a few simultaneous sessions)
- Up to 10 speakers per meeting (averaging 6)
- Few meeting types, but many tokens



Current Status

- 1017 total channel-hours recorded
- 529 close-talking hours (3-10 channels per meeting)
- 487 far-field hours (6 channels per meeting)
- 81 meeting-hours (85 meetings)
- about 60 unique speakers

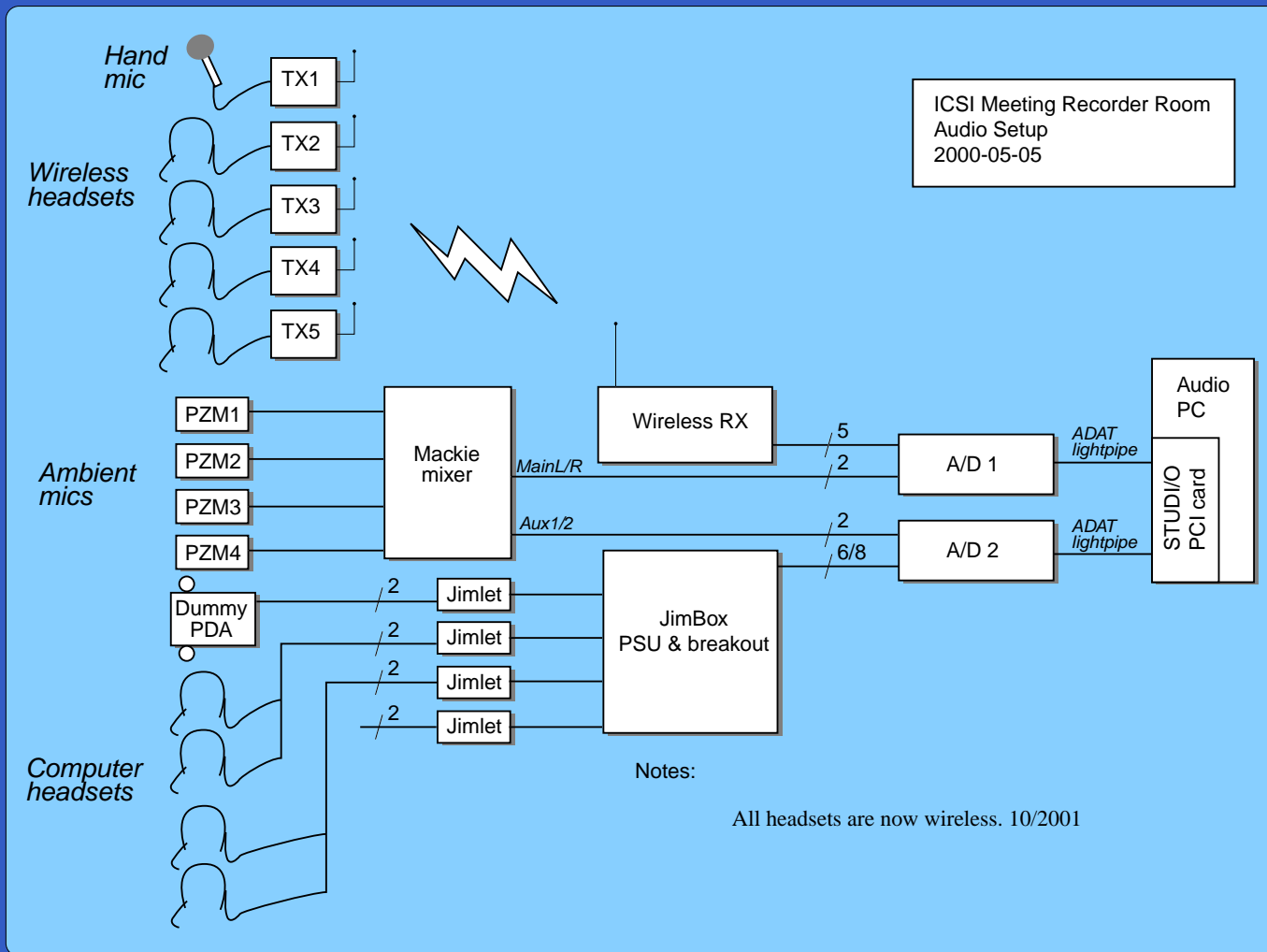


Data Collection Process

- Audio format: NIST Sphere, shortened (compressed), 16 KHz, 16 bit
- Up to 16 Channels (each in its own file):
 - 2 “PDA” mics
 - 4 PZM omni-directional (table-top) mics
 - 10 (max) close-talking (Sony[®] and Crown[®], mostly radio)
- Skew (fixed by including offset info. for each channel)



Meeting Room Hardware



Transcription Effort

- Transcription File Format
- Transcription Tools
- Transcription Process
- What do we transcribe?
- Transcript “Transformations”
- Current Transcription Status



Transcription File Format

XML based on the following:

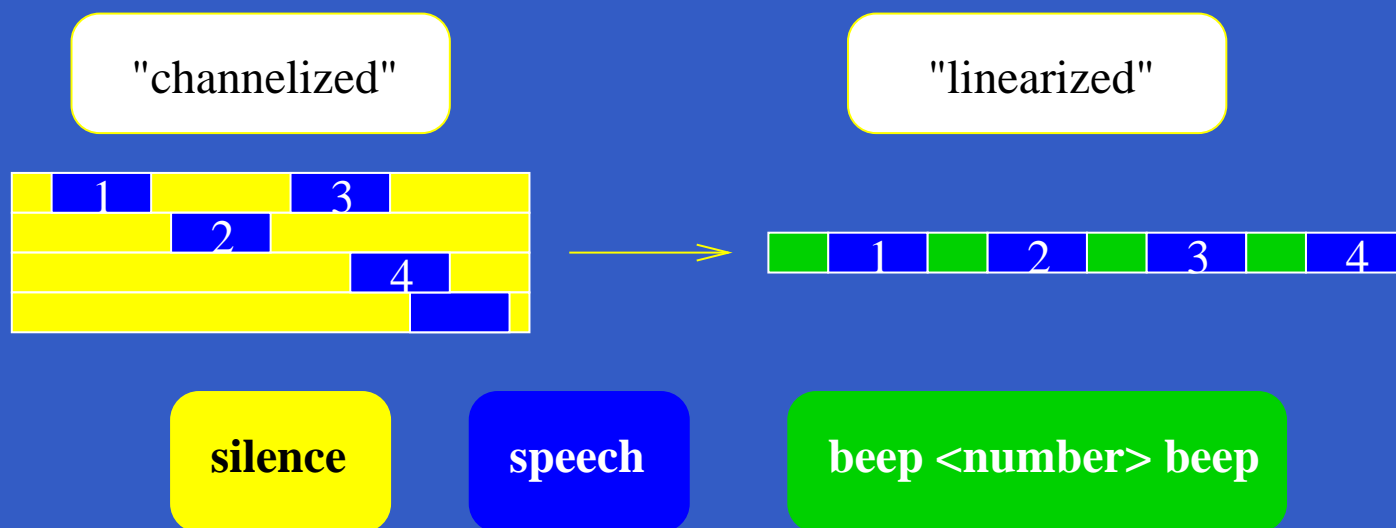
- ETCA “Transcriber” tool.
- TEI (Text Encoding Initiative), especially for time concepts.
- Annotated Transcription Graphs of Liberman, Bird et. al. (ATLAS).

See “References” slides below for more details on these items.



Transcription Tools

- trans → channeltrans
- “linearizing” transcripts (for fast first-pass transcription)

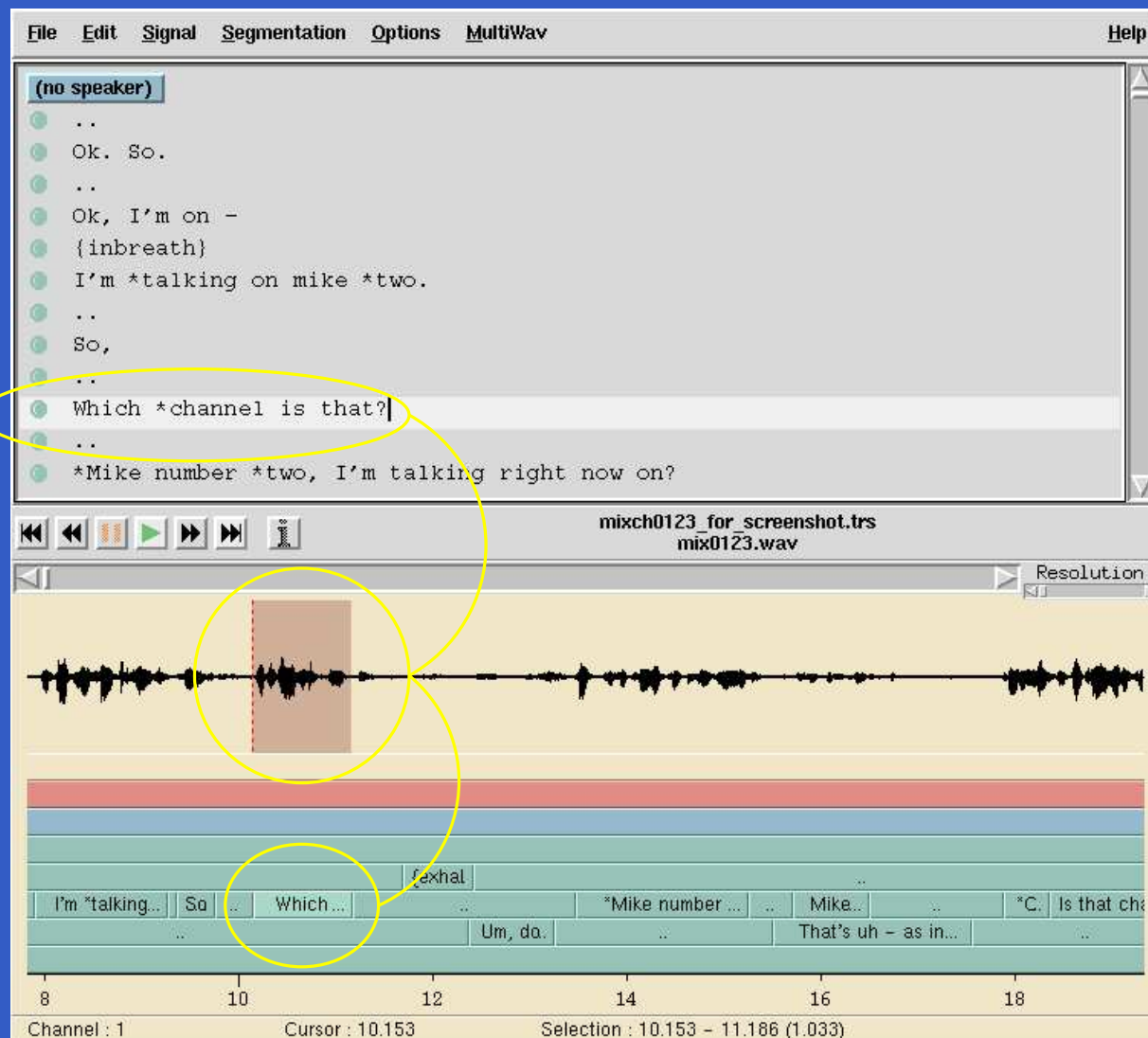


Transcription Tools (cont.)

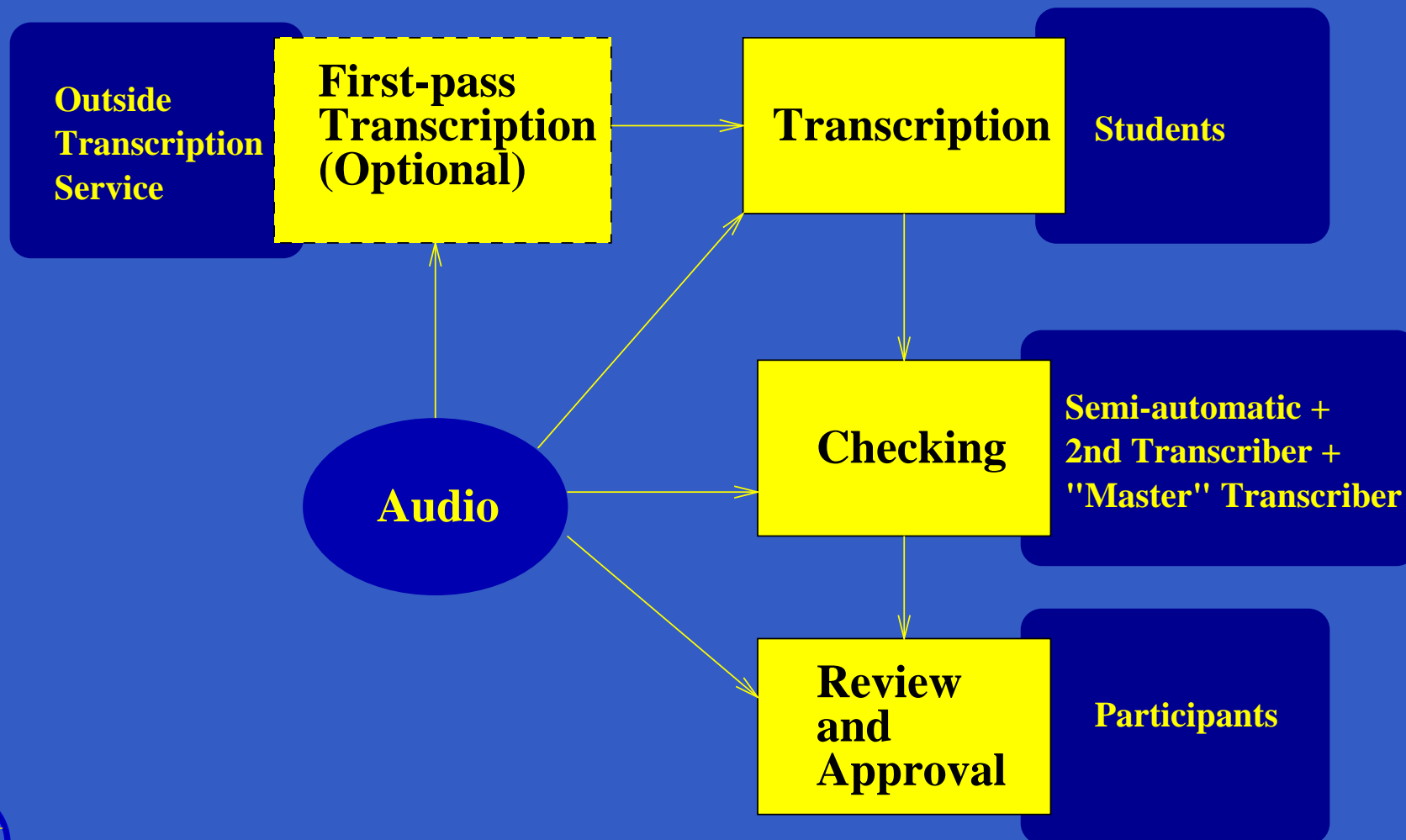
Transcription
for the current
channel

Current chan

All channels



Transcription Process



What do we transcribe? (Part I)

- Speakers, channels
- Words (plus abbreviations, acronyms, etc.)
- Overlaps (recoverable from time marks)
- Disfluencies (e.g. um, eh & interruptions)
- Backchannels (e.g. uh-huh)
- Non-canonical pronunciations
- False-starts, interruptions, etc.



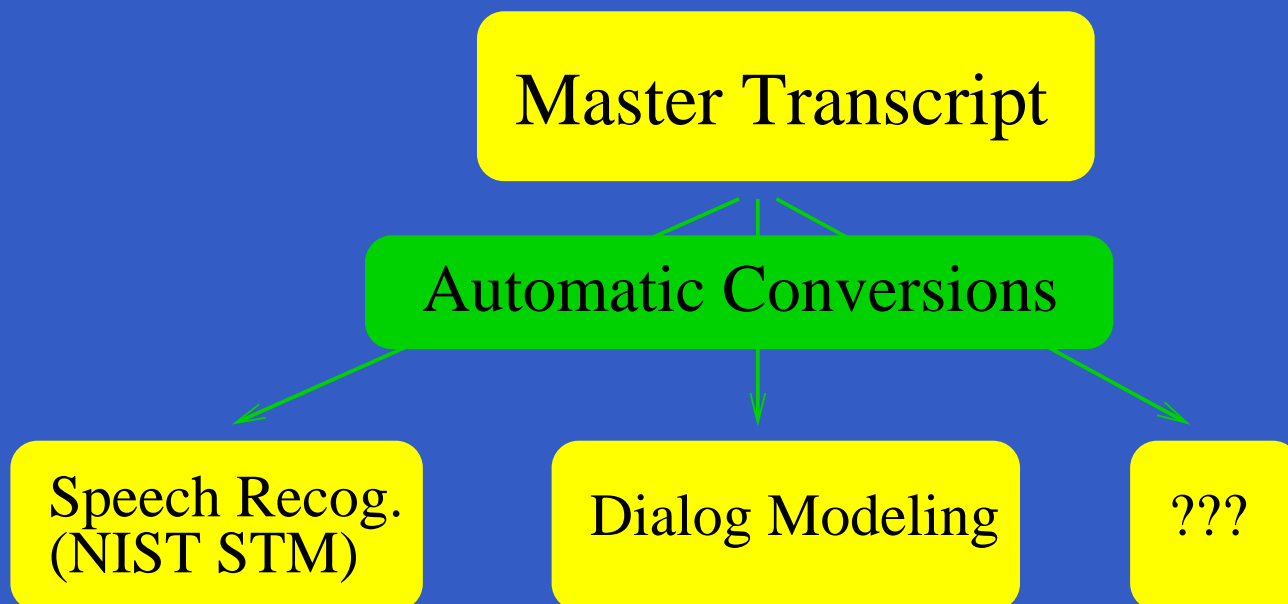
What do we transcribe? (Part II)

- Non-lexical events:
 - vocal: cough, laugh, breath, etc.
 - non-vocal: door slam, paper noise, etc.
- Acoustic uncertainty
- Qualifying information & contextual remarks
- “Bleeps”
- Utterance boundaries (via standard orthographic conventions)

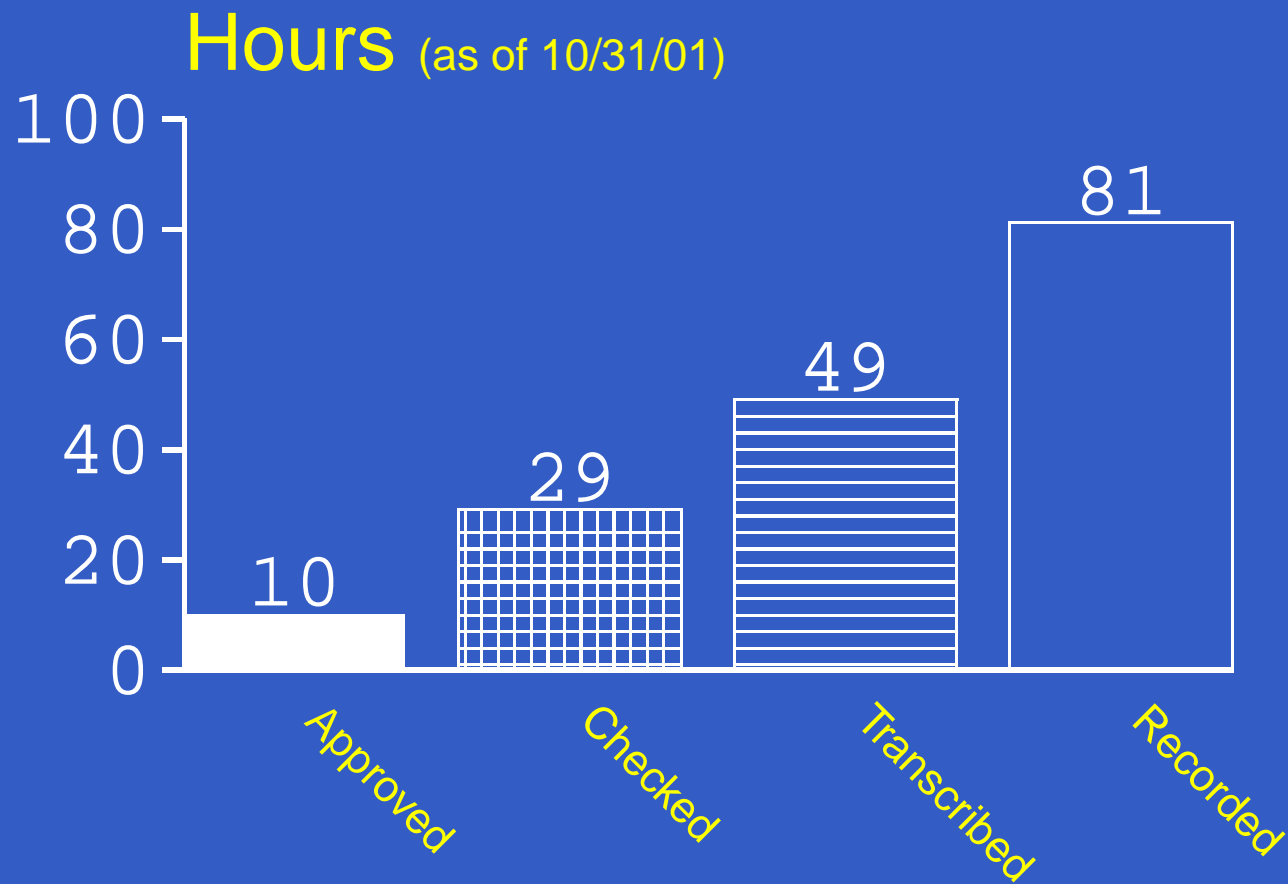


Transcript “Transformations”

Master transcript to be transformed to application specific versions.



Current Transcription Status



Ability to Contribute

- Our intention is to make the meeting corpus available to the community.
- Consent forms and Human Subjects approval
- Planning to have all 100 meetings collected by the end of 2001. Data to be transcribed and approved for release “soon” after that.



Open Issues

- How do we distribute the data?
 - Estimating 50 Gigs of data for 100 hours
- “Bleeping” vs. discarding entire meeting
- What gets transcribed? (Can’t anticipate all desired levels of annotation nor all potential applications.)
- Legal “responsibility” of organization collecting the data.



References

WWW:

- ETCA “Transcriber” tool:
<http://www.etca.fr/CTA/gip/Projects/Transcriber/>
- TEI (Text Encoding Initiative):
Main site: <http://www.tei-c.org/>
XML for TEI Lite:
<http://www.oasis-open.org/cover/tei.html>
- ATLAS (Architecture and Tools for Linguistic Analysis Systems)
<http://www.nist.gov/speech/atlas/>



References

WWW: (cont.)

- Annotation graphs (which are a special case of ATLAS):
<http://morph.ldc.upenn.edu/AG/>
- EAGLES (Expert Advisory Group on Language Engineering Standards):
<http://www.ling.lancs.ac.uk/eagles/delivera/wp4aug1.html>
- MATE (Multilevel Annotation, Tools Engineering):
<http://mate.nis.sdu.dk/>



References (cont.)

Tools:

- Barras, C. Geoffrois, E., Wu, Z., and Liberman, M. (2001). Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication*, 33, 5-22. ("Transcriber" interface)

Markup:

- Bird, S., Day, D., Garofolo, J., Henderson, J., Laprun, C., and Liberman, M. (2000) "ATLAS: A flexible and extensible architecture for linguistic annotation." <http://morph ldc.upenn.edu/AG/>
- Johansson, S. (1995). The approach of the Text Encoding Initiative to the encoding of spoken discourse. In G. N. Leech, G. Myers, J. Thomas (eds), *Spoken English on Computer: Transcript, Mark-up and Application* (pp. 82–98). NY: Longman Publishing.



References (cont.)

Transcription:

- Edwards, J. A. (1993). Principles and Contrasting Systems of Discourse Transcription. In J. A. Edwards, and M. D. Lampert (eds). Talking Data: Transcription and Coding in Discourse Research (pp. 3–31). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Edwards, J. A. (2001). The transcription of discourse. In The Handbook of Discourse Analysis. Edited by Deborah Tannen, Deborah Schiffrin, and Heidi Hamilton. New York: Blackwell.
- Leech, G., Weisser, M., Grice, M. and Wilson, A. (1998). "Survey and guidelines for the representation and annotation of dialogues." In: Gibbon, D., Moore, R. and Winski, R. (2nd edition). Handbook of standards and resources for spoken language systems. Berlin: Mouton de Gruyter.
(<http://www.ling.lancs.ac.uk/eagles/delivera/wp4aug1.html>)

