ICSI Meeting Data

Project Participants:

J. Ang, S. Bhagat, D. Baron, B. Chen, R. Dhillon,

J. Edwards, D. Ellis, D. Gelbart, A. Janin, A. Krupski,

N. Morgan, B. Peskin, T. Pfau, S. Renals,

E. Shriberg, A. Stolcke, **C. Wooters**

Background

- Began collecting data in Feb 2000
- Collaborations with UW, SRI, Columbia U., IBM, OGI; new ones with IM2, M4.
- Goal: development of technology to process spoken language from “natural” meetings

Current Research Using Meeting Data

- Speaker change detection, speaker tracking
- SpeechCorder handheld portable device
- Topic segmentation and summarization
- Automatic metadata extraction
- Dialog analysis/modeling
- Speech recognition
  - Far-field acoustics, conversational speech, speech activity detection, etc.

Types of Meetings

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

Meeting Room

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- lapel mic used in some early meetings)

Transcription File Format

XML based on the following:

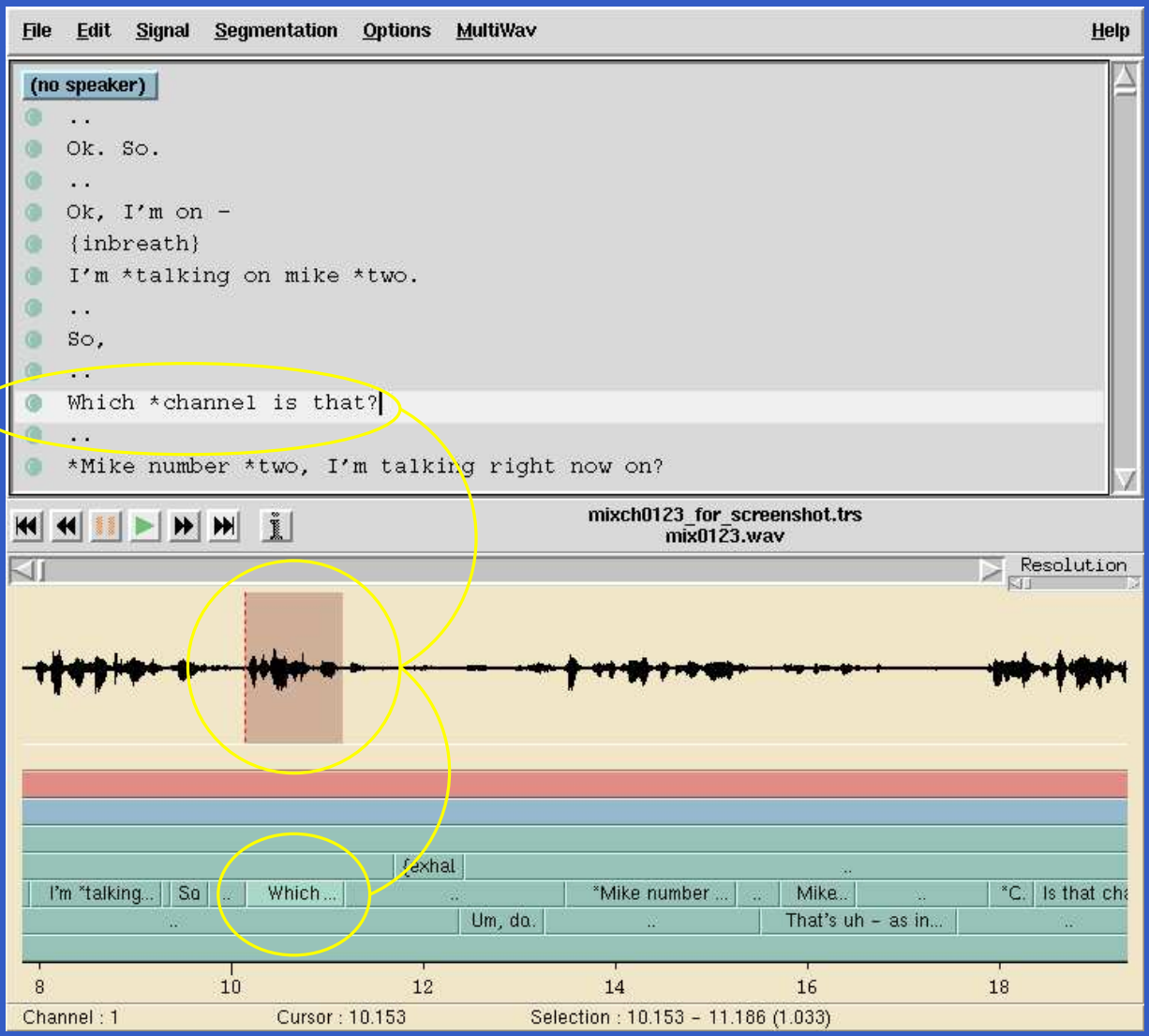
- ETCA “Transcriber” tool.
- Annotated Transcription Graphs of Liberman, Bird et. al. — ATLAS (Architecture and Tools for Linguistic Analysis Systems).

Transcription Tools (Channeltrans)

Transcription for the current channel

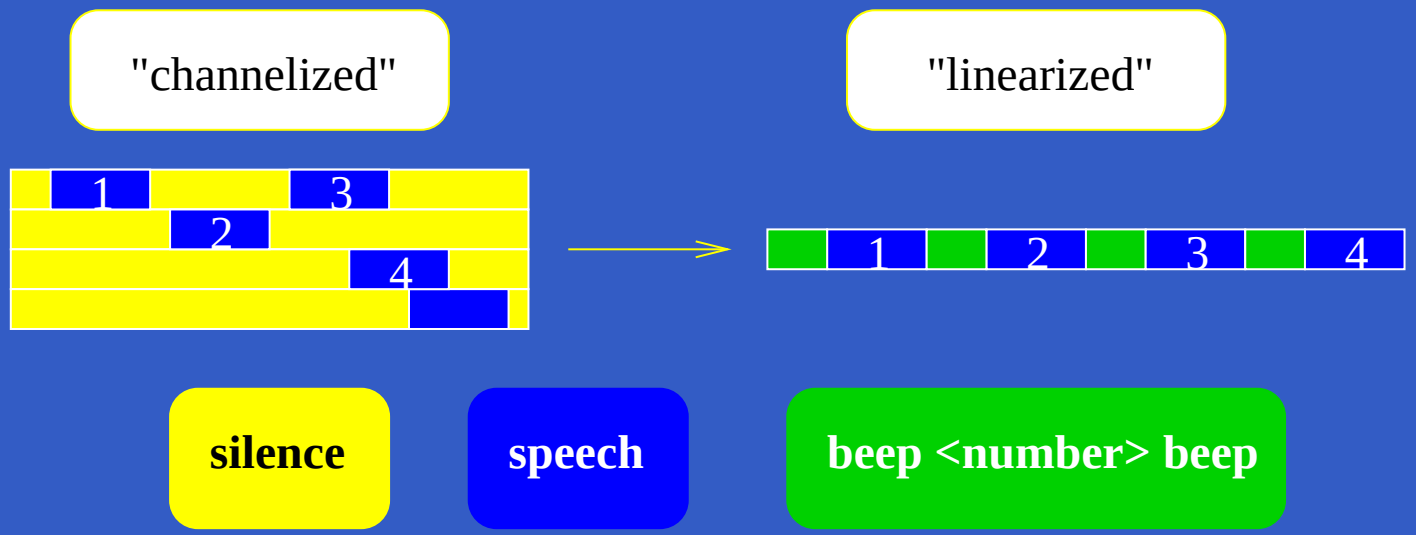
Current chan

All channels

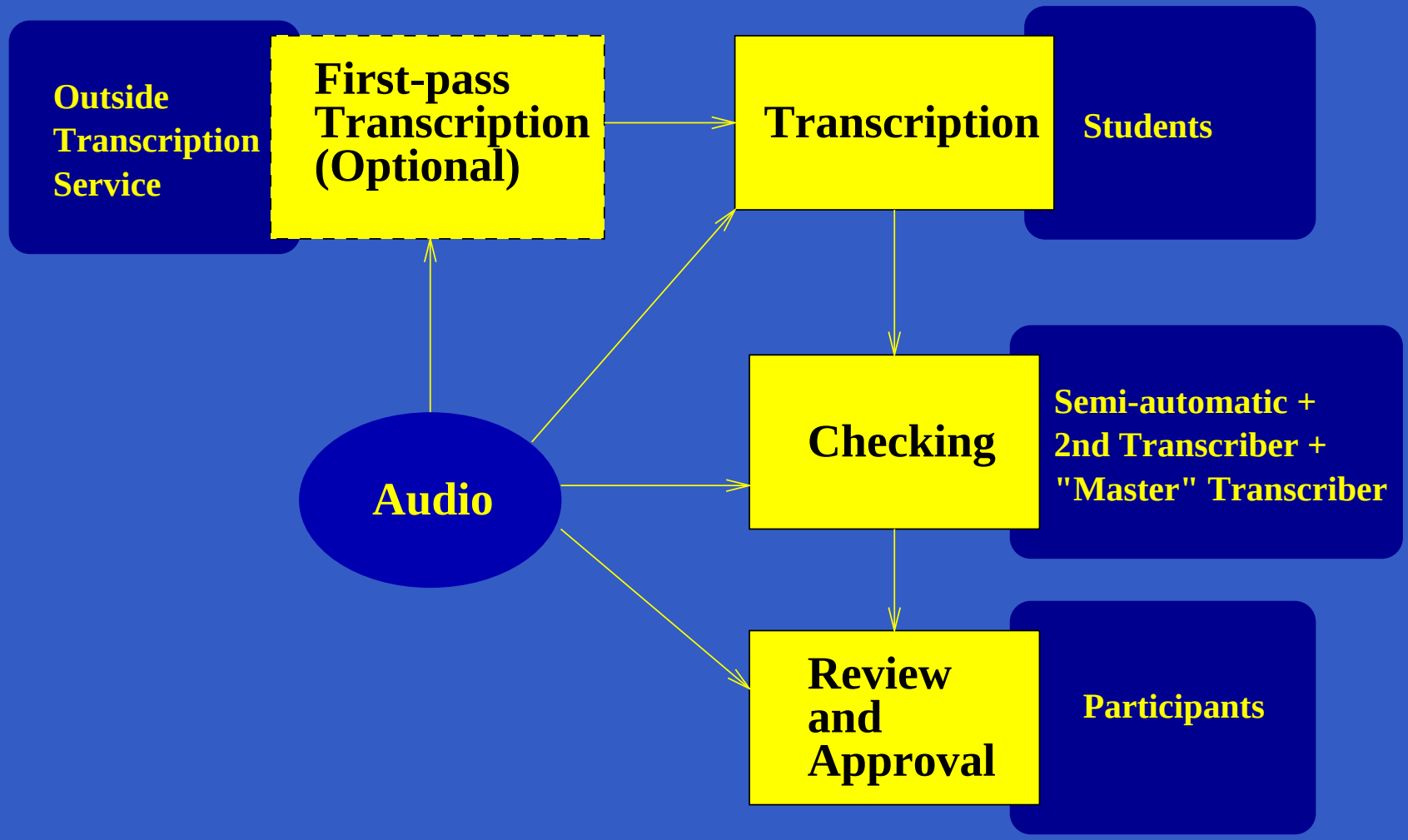


Transcription Tools (Linearization)

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



Transcription Process



What do we transcribe? (Part I)

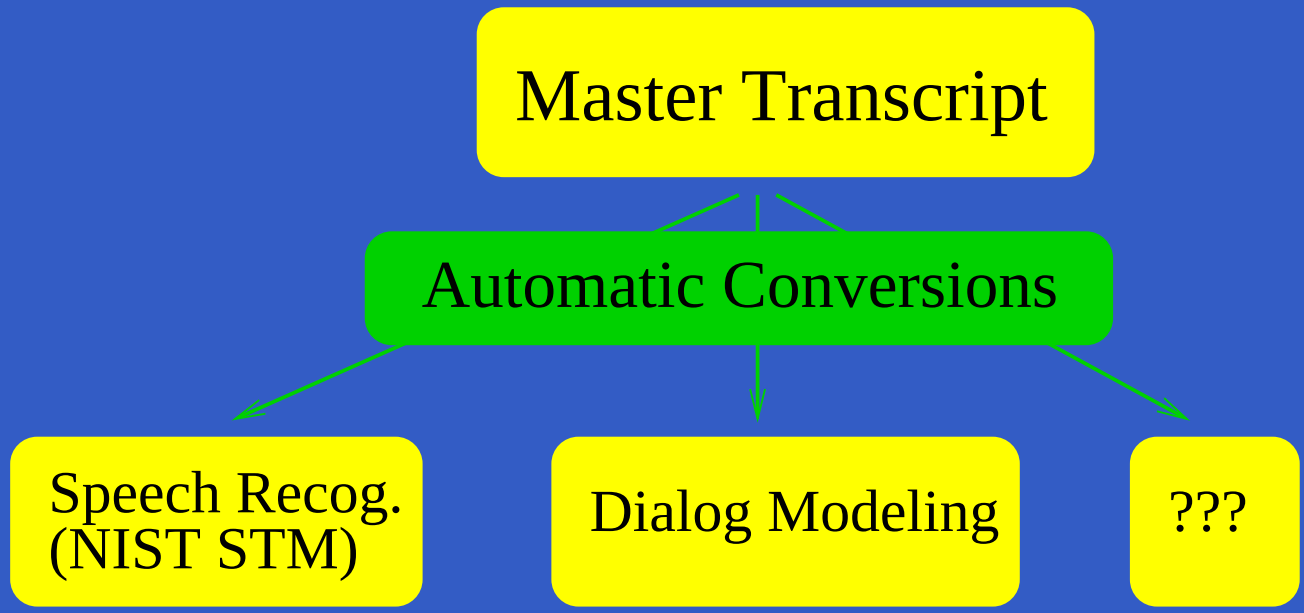
- Speakers, channels
- Words (plus abbreviations, acronyms, etc.)
- Overlaps (recoverable from time marks)
- Disfluencies (e.g. um, eh & interruptions)
- Backchannels (e.g. uh-huh)
- Non-canonical pronunciations
- False-starts
- Emphasis

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 segmentation (via standard orthographic conventions)

Transcript “Transformations”

Master XML transcript is transformed to application specific versions.



Corpus Status

- 87.8 meeting-hours (91 meetings)
  - transcribed: 76.5 hrs
  - checked: 43.4 hrs
  - approved: 40.46 hrs
- 1106.2 total channel-hours recorded
- 579.5 close-talking hours (3-10 channels per meeting)
- 526.7 far-field hours (6 channels per meeting)
- 72 unique speakers

Meeting Collection Issues

- How do we distribute the data?
  - Estimating 50 Gigs of data for 100 hours
  - Solution: LDC
- “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

- ICSI Meeting Recorder Project:
  http://www.icsi.berkeley.edu/Speech/mr/
- ETCA “Transcriber” tool:
  http://www.etca.fr/CTA/gip/Projects/Transcriber/
- ATLAS: http://www.nist.gov/speech/atlas/
- Annotation graphs:
  http://morph.ldc.upenn.edu/AG/