

ICSI and the Swiss Connection

Chuck Wooters

Senior Research Engineer

International Computer Science Institute, Berkeley, California, USA



Outline

- Introduction to ICSI
- The Meeting Recorder Project at ICSI
- The ICSI-IM2 Visitor Program



Introduction to ICSI

- Nonprofit, affiliated with UC Berkeley
- Inaugurated in 1988
- Focused on fundamental, open research in computer science
- Current international program participants: Spain, Germany, Switzerland, Finland



Introduction to ICSI (cont.)

- Strong domestic industrial support: Intel, Qualcomm, others
- Strong domestic government support: NSF, DARPA
- Potential ICSI/EPFL/IDIAP ties:
 - IM2: human-computer and human-human interaction
 - ICSI: natural meetings
 - Both: speech research



ICSI Ties to UC Berkeley

- UC Berkeley and affiliated labs (e.g., ICSI) are key centers for EECS research.
- UCB Professors
- UCB Students (PhD and others)
- Board of Trustees includes UCB faculty and admin
- Joint research projects
- Infrastructure: 200 Mb/s connection to campus, “berkeley.edu” domain

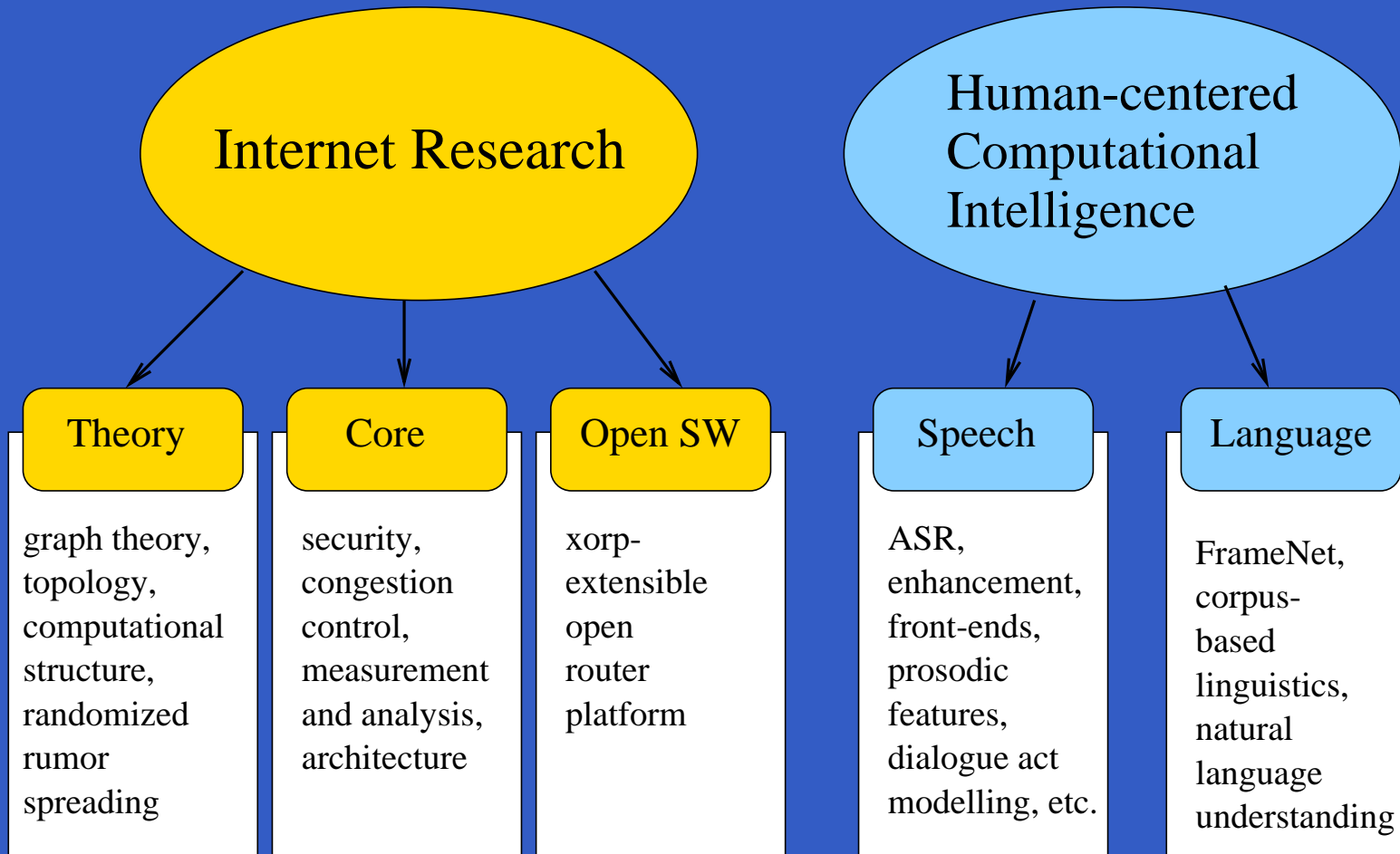


Research at ICSI

- Investigator-driven: 13 senior researchers
- Strong UCB student/faculty involvement
- Visitors key in several areas
- Involvement in standards processes (IETF in networks, ETSI in speech)
- Most work clustered in two areas



Major Scientific Themes



Human-centered Comp. Intelligence

ICSI emphasis: Natural Speech & Language

- Speech processing (especially recognition) — providing robust voice interfaces for ubiquitous (including portable and wireless) distributed information sources.
- Natural language understanding — use deep and corpus-based linguistic models to enable human interfaces to computers.



Speech Research Projects

- SmartKom — German speech interface project
- Center for Ubiquitous Speech Processing (CUSP) — speech/speaker recognition for cellular phones (ETSI speech standards)
- Rich Transcription — generating readable text from speech
- Meeting Browser — information retrieval and extraction from audio



Meetings Have It All

- Recorded (natural) meetings have nearly every conceivable spoken language research problem, in some portion
- Nonetheless, this is a domain with purposeful interaction (thus somewhat constrained)
- Speech technology for meetings has a wide range of applications
- We've been working on it (for several years)



Meetings: Technical Challenges

- ASR with increased speaker variability (pronunciation, rate, overlaps)
- ASR with increased acoustic variability (far field microphones, speaker movement)
- Other acoustic analysis — segmentation, turn detection, speaker ID, emotion, etc.
- Dialog abstraction, analysis
- Meeting summarization, IR
- Portable devices (energy, memory, ...)



ICSI Work On Meetings

- Primary collaborations with UW, SRI; also with OGI, Columbia U, IBM; new ones with IM2, M4.
- Fundamental Goal: technology to process spoken language from “natural” meetings
- Near-term goals: data collection, transcriptions, preliminary systems for segmentation, ASR
- Long-term goals: understanding, “mapping” meetings



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 6)
- Few meeting types, but many tokens

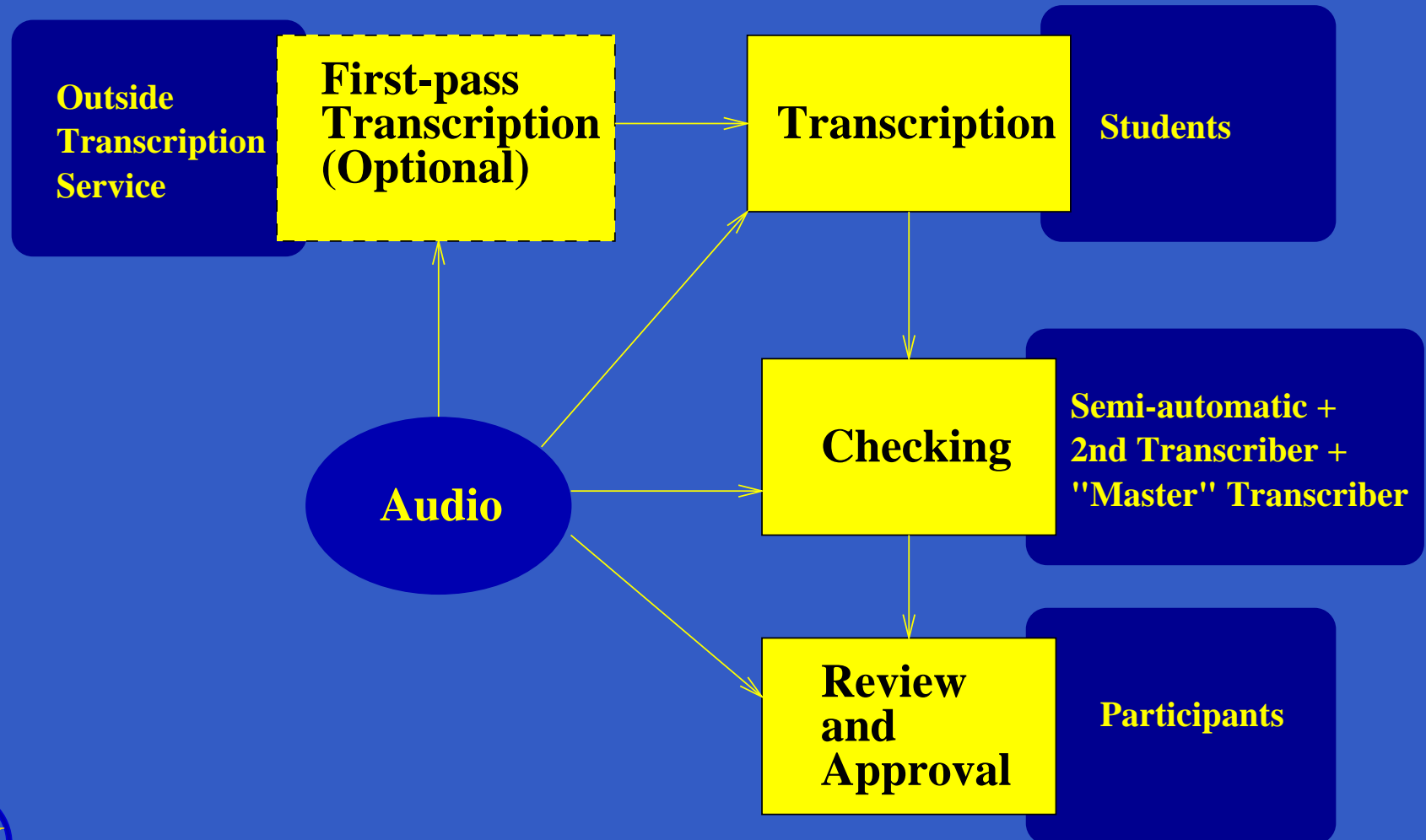


Data Collection Process

- 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)



Transcription Process



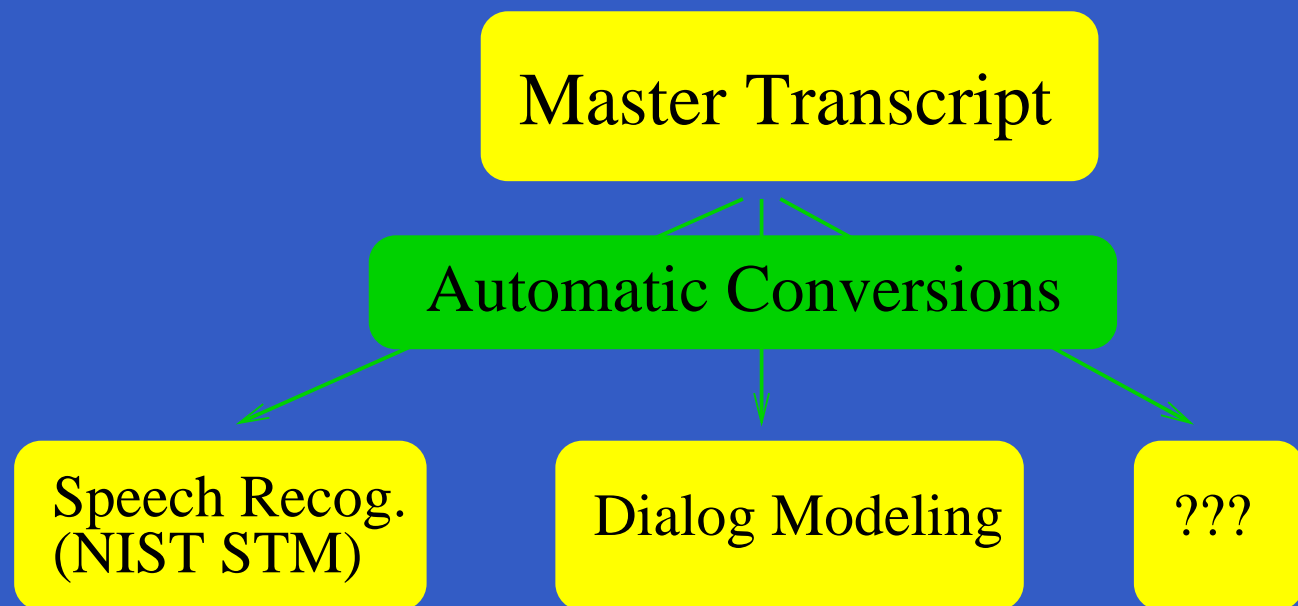
What do we transcribe?

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



Transcript “Transformations”

Master XML transcript is transformed to application specific versions.



Corpus Status

- Over 80 hours x 8-15 channels recorded at ICSI
- Mostly transcribed, but going through final checking and approval stages
- Connection to NIST efforts: transcription standards for NIST recordings, RT-02 dev and eval data
- Planning to distribute corpus via LDC



Current ICSI Meeting Research

- Speech Activity Detection/Segmentation
- Enhancement for reverberated (far-field) speech recognition
- Portable device issues
- Prosodic features
- Overlaps, modeling for punctuation, ASR
- Adjacency pairs, pragmatics
- Information retrieval and summarization



ICSI-IM2 Visitor Program

- Pre- and post-doctoral visitors
- Six-month or one-year (preferred) research stay in Berkeley
- Visitors will be involved in ICSI projects relevant to IM2
- Initially a two-year program, but if successful, we expect to keep renewing it.
- Swiss contact: Hervé Bourlard

