The Rich Transcription 2007 Evaluation Overview

$\mathcal{R}_{ich}$
$\mathcal{T}_{ranscription}$

http://www.nist.gov/speech/tests/rt/rt2007

Jonathan Fiscus and Jerome Ajot
May 10-11, 2007

Rich Transcription 2007
Meeting Recognition Workshop

Fostering Multi-modal Technologies
CLEAR / RT

- Addressing the evaluations to drive research in technologies that cross disciplinary boundaries (Vision, Audio, Text, Gesture, Cognitive Psychology, etc.)
- Two Evaluation Series
  - Rich Transcription
    - Evaluation of technologies for language transcription
  - Classification of Events, Activities, and Relationships Evaluation Program and Workshop series
    - Evaluations of technologies for human activity and interaction analysis
- Supported by several programs
  - AMI, CHIL, NIST, VACE

NIST
National Institute of
Standards and Technology

Rich Transcription Conceptualization
(Broadcast News Domain)

Traditional Speech-To-Text Output	Enriched Output	Readable Transcript
from a. b. c. news world headquarters in new york this is world news tonight with peter jennings good evening every one we're going to begin tonight on the presidential campaign trail littered as it is today with democratic candidates who have suddenly been obliged to reassess their futures when former vice president al gore formally endorsed howard dean as the presidential candidate today	`<speaker name="Announcer"> <SU type="stmt">` from `<prop_noun>` a. b. c. news world headquarters `</prop_noun>` in `<prop_noun>` new york `</prop_noun> <phrase_bound>` this is `<prop_noun>` world news tonight `</prop_noun>` with `<prop_noun>` peter jennings `</prop_noun> </phrase_bound> </SU> </speaker>` `<speaker name="Peter Jennings"> <SU type="stmt">` good evening every one `</SU> <SU type="stmt">` we're going to begin tonight on the presidential campaign trail	**Announcer:** From ABC News World Headquarters in New York, this is World News Tonight with Peter Jennings. **Peter Jennings:** Good evening every one. We're going to begin tonight on the presidential campaign trail. Littered as it is today with Democratic candidates who have suddenly been obliged to reassess their futures when former Vice President Al Gore formally endorsed Howard Dean as the presidential candidate today.

Other language processing →

Translate
Summarize
Parse
Extract Info
Search

Rich Transcription Evaluation Series

- Goal:
  - Develop recognition technologies that produce language content representations (transcripts) which are understandable by humans and useful for downstream processes.
- Domains:
  - Meeting Room speech
  - Broadcast News (BN)
  - Conversational Telephone Speech (CTS)
- Parameterized "Black Box" evaluations
  - Evaluations control input conditions to investigate weaknesses/strengths
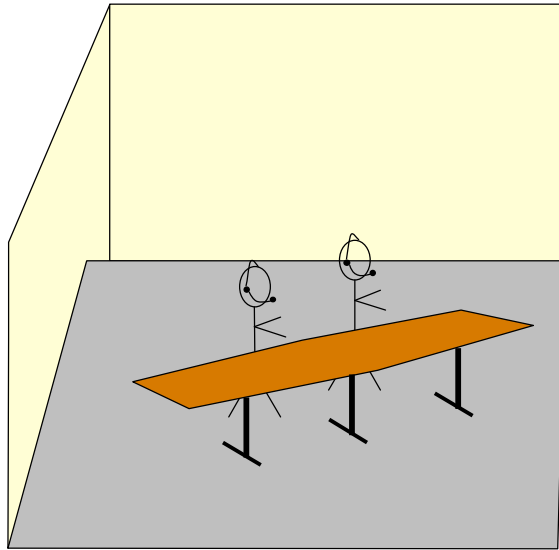 - Sub-test scoring provides finer-grained diagnostics

RT-07 Evaluation Tasks

- Focus on core speech transcription technologies
  - Speech-To-Text Transcription (STT)
    - Transcribe the spoken words
  - Diarization "Who Spoke When" (SPKR)
    - Detect segments of speech an cluster them by speaker
  - Speaker Attributed Speech-To-Text (SASTT)
    - Transcribe the spoken words and associate them with a speaker
    - Experiment

RT-07 Evaluation Participants

Site ID	Site Name	Evaluation Task		
		SPKR	STT	SASTT
AMIDA	Augmented Multi-party Interaction with Distance Access	X	X	X
I2R/NTU	Infocomm Research Site and Nanyang Technological University	X		
IBM	IBM	X	X	X
ICSI	International Computer Science Institute	X		
LIA	Laboratoire Informatique d'Avignon	X		
LIMSI	Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur	X		X
SRI/ICSI	SRI International and International Computer Science Institute		X	X
UKA	Karlsruhe University		X	
UPC	Universitat Politècnica de Catalunya	X		

Six System Input Conditions

Distant Mic. Conditions
- Mult. Distant Mics. (MDM)
- Mult. Source Localization Arrays (MSLA)
- Mult. Mark III Arrays (MM3A)
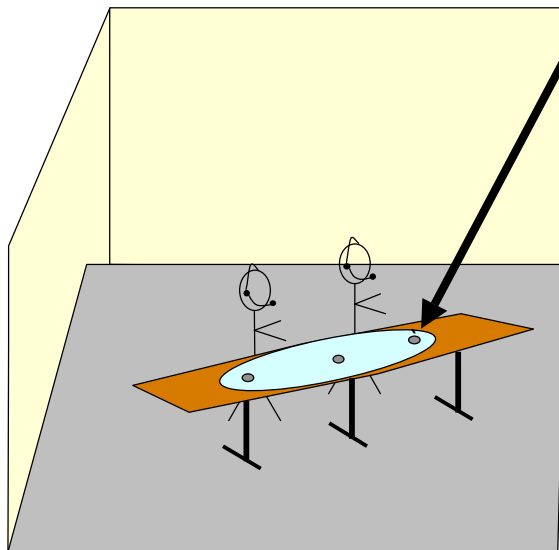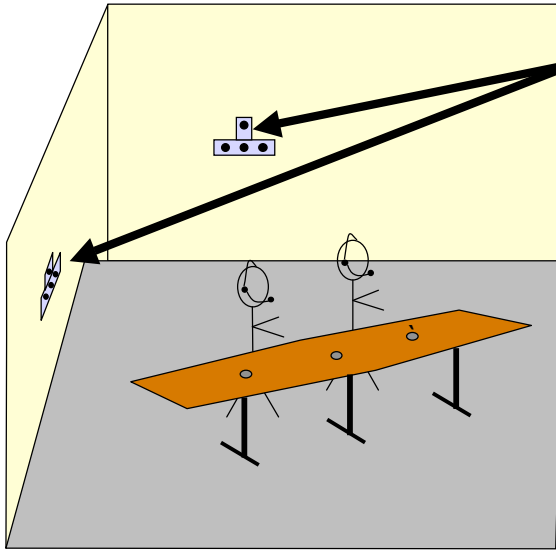- All Distant Mics. (ADM)

Contrast Conditions
- Individual Head Mics. (IHM)
- Single Distant Mic. (SDM)

Laboratory Conditions
- IHM with reference segmentations

Distant Mic. Conditions
- Mult. Distant Mics. (MDM)
- Mult. Source Localization Arrays (MSLA)
- Mult. Mark III Arrays (MM3A)
- All Distant Mics. (ADM)

Contrast Conditions
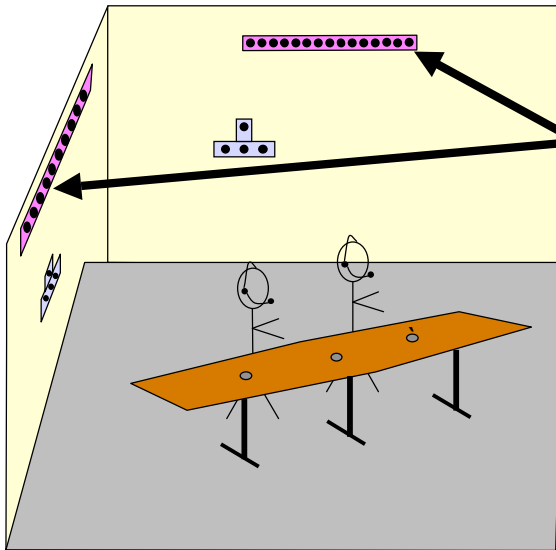- Individual Head Mics. (IHM)
- Single Distant Mic. (SDM)

Laboratory Conditions
- IHM with referencesegmentations

Six System Input Conditions

Distant Mic. Conditions
- Mult. Distant Mics. (MDM)
- Mult. Source Localization Arrays (MSLA)
- Mult. Mark III Arrays (MM3A)
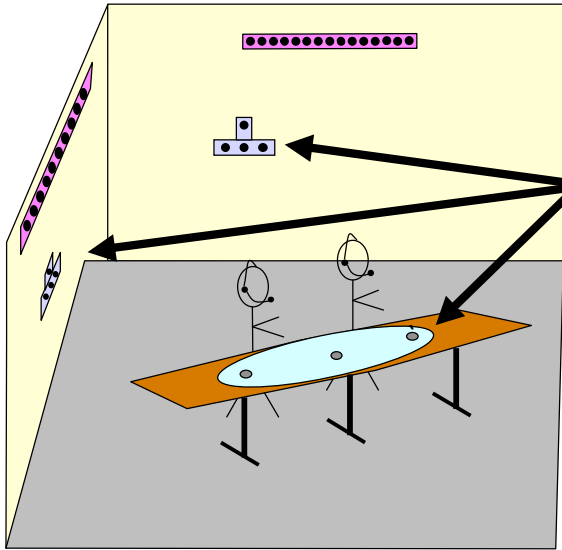- All Distant Mics. (ADM)

Contrast Conditions
- Individual Head Mics. (IHM)
- Single Distant Mic. (SDM)

Laboratory Conditions
- IHM with reference segmentations

Six System Input Conditions

Distant Mic. Conditions
- Mult. Distant Mics. (MDM)
- Mult. Source Localization Arrays (MSLA)
- Mult. Mark III Arrays (MM3A)
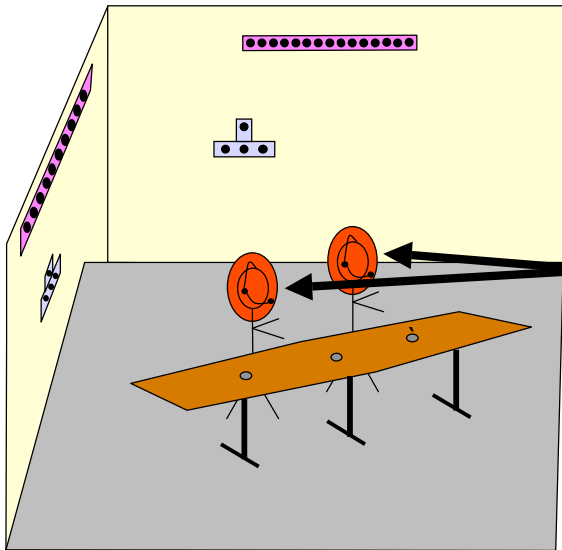- All Distant Mics. (ADM)

Contrast Conditions
- Individual Head Mics. (IHM)
- Single Distant Mic. (SDM)

Laboratory Conditions
- IHM with reference segmentations

Six System Input Conditions

Distant Mic. Conditions
- Mult. Distant Mics. (MDM)
- Mult. Source Localization Arrays (MSLA)
- Mult. Mark III Arrays (MM3A)
- All Distant Mics. (ADM)

Contrast Conditions
- Individual Head Mics. (IHM)
- Single Distant Mic. (SDM)

Laboratory Conditions
- IHM with reference segmentations

NIST
National Institute of
Standards and Technology

Six System Input Conditions

Distant Mic. Conditions
- Mult. Distant Mics. (MDM)
- Mult. Source Localization Arrays (MSLA)
- Mult. Mark III Arrays (MM3A)
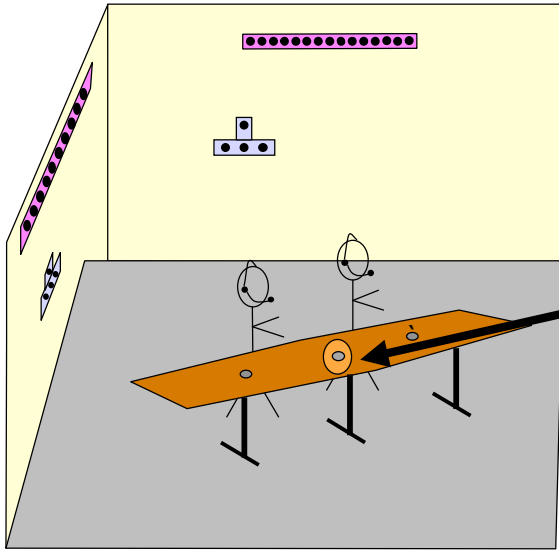- All Distant Mics. (ADM)
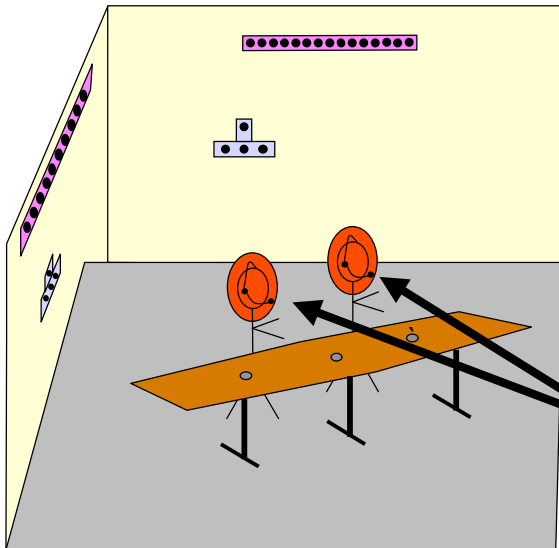
Contrast Conditions
- Individual Head Mics. (IHM)
- Single Distant Mic. (SDM)

Laboratory Conditions
- IHM with reference segmentations

NIST
National Institute of
Standards and Technology

Six System Input Conditions

- Distant microphone conditions
  - Multiple Distant Microphones (MDM)
    - Three or more centrally located table mics
  - Multiple Source Localization Arrays (MSLA)
    - Inverted "T" topology, 4-channel digital microphone array
  - Multiple Mark III digital microphone Arrays (MM3A)
    - Linear topology, 64-channel digital microphone array
  - All Distant Microphones (ADM)
- Contrastive microphone conditions
  - Single Distant Microphone (SDM)
    - Center-most MDM microphone
    - Gauge performance benefit using multiple table mics
  - Individual Head Microphones (IHM)
    - Performance on clean speech
    - Similar to Conversational Telephone Speech
      - One speaker per channel, conversational speech
- Laboratory conditions
  - Individual Head Mics. With reference segmentations (IHM-REFSEG)

Training/Development Corpora

- ICSI Meeting Corpus
- ISL Meeting Corpus
- NIST Meeting Pilot Corpus (including RT-02 test data)
- NIST Phase II Meeting Corpus
- Topic Detection and Tracking Phase 4 (TDT4) corpus
- Fisher English conversational telephone speech corpus
- CHIL '05, '06, and '07 development test sets
- AMI Meeting Corpora
- Rich Transcription 2004 Spring (RT-04S) Development & Evaluation Data
- Rich Transcription 2005 and 2006 Evaluation Data

RT-07 Evaluation Corpora

- Since the meeting domain is so broad, three meeting sub-domains were chosen
  - Conference Room
  - Lecture Room
  - Coffee Break

NIST
National Institute of
Standards and Technology

RT-07 Evaluation Test Corpora: Conference Room Test Set

- Goal-oriented small conference room meetings
  - Group meetings and decision-making exercises
  - 4-6 participants

Duration	180 minutes
Excerpts	8 (22.5 min. per mtg.)
Collection Sites	CMU, Edinburgh Univ., NIST, VT
Selected by	NIST
Transcribed by	LDC

RT-07 Evaluation Test Corpora:
Lecture Room/Coffee Break Test Sets

- Technical lectures in small meeting rooms
  - 3-7 participants
  - All interactive seminars

	Lecture	Coffee Break
Duration (min.)	164	41
Excerpts (5 min.)	32	8
Collection Sites	AIT, IBM, ITC, UKA, UPC	
Selected by	CMU	
Transcribed by	CMU	

Lecture

Coffee Break

Thanks to All Conspirators

- Data Contributors:
  - Augmented Multiparty Interaction (AMI) Program Members: Edinburgh and ICSI
  - Computers in the Human Interaction Loop (CHIL) Partners: AIT, IBM, ITC, UKA, UPC
  - Virginia Tech.
  - NIST
  - CMU
- Test Data Preparation:
  - Selection and Transcription of the Lecture Data: CMU and UKA
  - Transcription of the Conference Data: LDC
  - LIMSI for the use of their forced alignment tools
- Shared system outputs:
  - SRI/ICSI shared head mic. segmentations from last year's system
- Venue arrangements:
  - Rachel Bowers and Teresa Vicente