Rich Transcription 2003

Spring Evaluation and Workshop

May 19, 2003

Welcome

• General Information sheet in notebook
  – meeting rooms, meals, map, transportation

• Conference Coordinators
  – Patrice Boulanger, NIST
  – Loeetti “Lo” Alexander, CACI

• Evaluation Form
  – Please fill it out and return to registration desk by end of workshop
Motivation

• Past Speech-To-Text (STT) output: Play
  – “ew very nice yes that’s that’s the ah first car uh well my first ownership of something major that’s cool”
  ➢ plain STT token output is difficult for humans to read/understand, difficult for machines to process (beyond “bag-of-words” approaches)

• What humans (and machines) would rather see:
  Very nice .
  Yes .
  That’s my first ownership of something major .
  That’s cool .

• How can this be produced?
  – First enrich STT output stream with syntactic/semantic Metadata Extraction (MDE):
    • structural information needed for rendering readable transcripts for humans
    • linguistic/semantic information for downstream language processing applications

  Rich Transcription (RT) = STT + MDE

  – Then use MDE markup to render transcript into readable form
    • Transformed Transcription (XT) = f(RT)

Useful Metadata for Readable Rendering

- Lexical tokens
- Word fragment detection
- Sentence-like units
- Non-essential clauses
- List structures
- Aside comments
- Pronominal co-reference
- Verbal edits (restarts and repetitions)
- Pause fillers
- Filler disfluencies
- Speaker Information
  • Named entities
  • Numeric expressions
  • Proper adjectives
  • Adjectival phrases
  • Acronyms
  • Background acoustics
  • Direct quotations

= Currently addressed in EARS
Rich Transcription Series

• Evaluation/workshop series focused on creating/improving rich transcription technologies
  – different component technologies have various levels of maturity

• Began with RT-02 last year
  – STT and Speaker Segmentation Tasks

• During past year
  – STT community has been working hard to drive down error rates.
  – MDE community has been working hard to define the sentence and disfluency tasks and evaluation metrics.

RT-03 Evaluations/Workshops

• RT-03S Spring Evaluations:
  – BNews and Conversational Telephone Speech Recognition (transcribe words)
  – Speaker Diarization (cluster speech by speaker and classify speakers by gender)

• RT-03F Proposed Fall Evaluations:
  – SU Detection and Recognition (tag and type sentences)
  – Disfluency Detection and Recognition (tag and type disfluent words [Filler,Edit,IP] )
  – Meeting Room Recognition (transcribe words)
  – “Spkr What” Detection (tag recognized words with speakers/turns)
  – RT (transcribe words plus metadata: \( RT = STT+MDE \))
  – XT (produce human-rendered text from RT: \( XT = f(RT) \))
RT03S Speech-to-Text (STT) Tests (Words)

- Goal: Transcribe word tokens spoken
- Many dimensions explored:
  - Languages:
    - English, Mandarin-Chinese, Egyptian-Arabic
  - Domains:
    - Broadcast News and Telephone Conversations
  - Processing Speeds:
    - Realtime, 10X, “unlimited” processing speeds
  - Test Set Type:
    - Fixed (Progress) vs. evolving (Current) test sets
  - System Type:
    - Primary vs. contrastive systems

RT-03 Speaker Diarization Metadata Tests (Speakers/Turns)

- Goal: Identify segments of speech and group them by speaker, identify gender of each speaker
- Dimensions:
  - Domains:
    - English broadcast news and telephone conversations
  - Control conditions:
    - Reference words known
RT03-S Test Corpora

• Current Test Sets – Fresh data for each new evaluation
  – English Current Test Set
    • First half used for Speaker Diarization Tests
    • Second half to be used for Fall MDE Tests
  – Arabic Current Test Set
  – Chinese Current Test Set

• Progress Test Sets – Reusable data
  *(for EARS participants only)*
  – English

• All test sets contain both broadcast news and conversational telephone speech subsets

Thanks!

• Thanks to the Stephanie Strassel and the LDC for preparing the test reference data
• Thanks to Patricia, Lo, Adam Cushing, and Alvin Martin for organizing the workshop logistics
• Thanks to DARPA for sponsoring the evaluation and workshop
• A special thanks to the STT and Speaker Diarization Evaluation Participants