Linguistic Resources for Meeting Recognition

Meghan Glenn, Stephanie Strassel
Linguistic Data Consortium

{mlglenn, strassel@ldc.upenn.edu}

Overview

- Conference room test data
  - Careful transcription
    - Speaker noise
- Unique Challenges
- Infrastructure
  - XTrans Toolkit
    - Existing features for meetings
    - Future features for meetings
- Open questions

RT-06S Evaluation Data

transcribed by LDC

- Conference room data
  - Nine meeting sessions, eleven excerpts
  - Contributed by five sites (CMU, EDI, NIST, TNO, VT)
  - Multiple recording conditions for each session
  - Excerpts between 8 and 18 minutes long
  - Between 4 – 9 speakers per meeting
- Primarily business meeting content
  - Transcribers report it was easier to transcribe than previous years' data
- All data carefully transcribed (CTR)

Careful Transcription (CTR) Process

- Using IHM channels
  - Chopped audio files
- 1st pass: manual segmentation
  - Turns → breath groups
  - 3-8 seconds per segment, designed for ease of transcription only
    - ~10 ms padding around each segment boundary
    - Segmentation and transcription of isolated speaker noise such as {breath}
- 2nd pass: initial verbatim transcription
  - No time limit
  - Goal is to “get everything right”
- 3rd pass: verify existing transcription and timestamps, add additional markup
  - Indicate proper names, filled pauses, noise, etc.
  - Revisit difficult sections

Careful Transcription

- Additional QC pass by lead transcriber
  - Using mixed IHM recordings and/or SDM
  - Merge individual transcripts
    - Speaker ID consistency
    - Transcription accuracy, completeness
    - Markup consistency
    - Spell check
    - Check consistency, accuracy of names, acronyms, terminology
  - Check silence (untranscribed) regions for missed speech using customized tool
  - Expand contractions
  - Syntax (format) check
    - Badly formatted <foreign> speech regions
    - Misspelled words
    - Conflicting markup
    - File format errors
- Final check on merged, reformatted transcripts for consistency across meetings

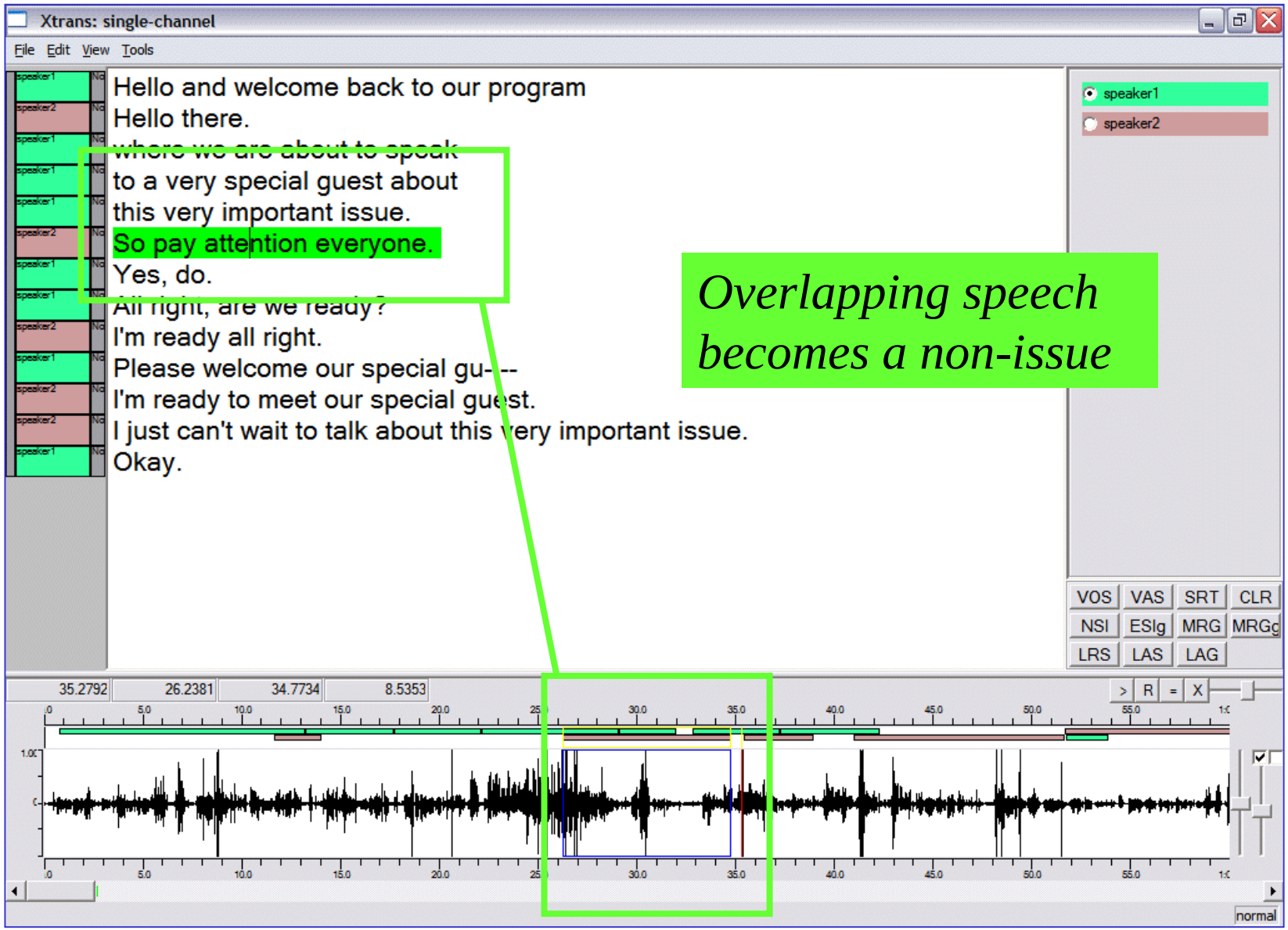
Unique Challenges

- Many speakers = longer to transcribe
- Varying levels of speaker participation
  - Often no speech but other speaker/background noise or loud breaths/sighs
- Meeting content
  - Primarily project discussion groups, technical meetings
- Access to video would also enrich transcription process

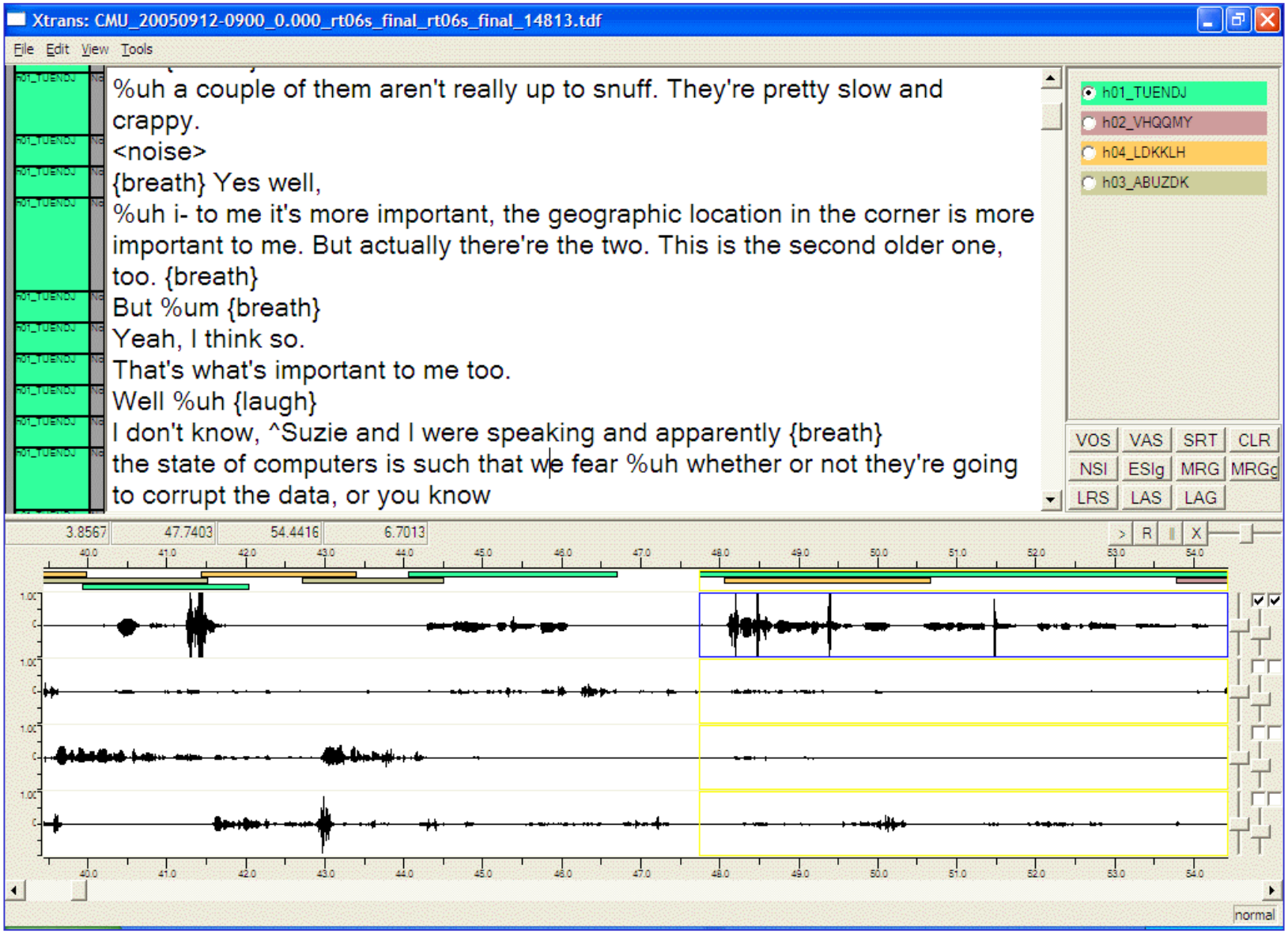
XTrans Annotation Tool

- Multipurpose speech annotation tool
- Multilingual, multi-platform, multi-format
  - Written in Python
  - AGTK infrastructure
  - Import/Export from/to a variety of formats
- Customized task modules
  - Careful transcription
  - Quick transcription
  - “Metadata” annotation
    - Structural features
      - “SU” boundaries
      - Story or topic boundaries
    - Speaker diarization

One Channel View



MultiSpeaker, Single Channel View



MultiSpeaker, MultiChannel

Improvements from XTrans

- Reduces annotation rates
  - Keyboard shortcuts
  - Customized string insertions
    - Vary per file or annotator
  - RT05S annotation rate: 65 X RT average, not broken down by channel
  - RT06S annotation rate: 50 X RT average, not broken down by channel
- Easy to manage
  - Simplified workflow
  - More transcriber options
  - Personalize experience so maximally efficient
- Built-in QC functions
  - Speaker verification
  - Inter-gap listening function

Future features for XTrans

- MP3 and video support
  - Actively pursuing both of these additions
- Adjudication mode
  - Dual transcription and annotation
  - Better train transcribers
  - Experiment with various “gold standard” reference standards

Open Questions

- Segmentation
  - RT05/RT06 turn segmentation rules are to form segments between 3 and 8 seconds long
  - Within GALE program, we've had reasonable success in performing "SU segmentation".
    - Extent of SU (syntactic unit, semantic unit, sentence unit) is extent of segment (regardless of length)
    - Get punctuation "for free"
- Speaker noise annotation
  - Good headphones + IHM channels = lots of speaker noise.
  - How close should transcription be?

Acknowledgements

- Thanks to LDC transcribers
- Kazuaki Maeda and Haejoong Lee
- NIST