Overview of CTS Collection at LDC

David Graff, Shudong Huang,
Chris Cieri, Kevin Walker
(graff, ccieri, shudong, walkerk)@ldc.upenn.edu

ftp://ftp.ldc.upenn.edu/pub/ldc/MMSR/LDC_sre2006.ppt

How Mixer3 Got Started

- Discussion of new SRE collection began in Nov. 2006
- Two other collections were already in progress at LDC (LVDID, Fisher Spanish)
- Multi-language LVDID was stalling due to lack of incentives for participants
- Multi-language SRE could provide incentives; both projects could use same speaker pool
- SRE collection began in earnest Dec. 20
- Auditing began Jan. 23
- 4000+ 2-ch. calls delivered to NIST Mar. 1

March Delivery to NIST

- 4393 calls delivered
- 1108 distinct speakers represented
  - 495 speakers had 12 or more call sides
  - 323 of those had used 4 “unique” handsets
  - Adjustments made for 80 duplicate enrollments
- 7538 sides audited (1248 pending as of 3/1)
  - All calls had at least one side audited
  - All high-count speakers done first
- 2838 English calls, 1555 non-English calls
  - 331 “expected ENG, heard some/no_ENG”
  - 2481 “All_ENG” calls had at least one non-native
- 275 non-Eng calls held in reserve for LVDID

Current Collection Status

- Over 8850 calls collected (~4500 new)
- 1932 speakers represented (824 new)
  - 658 have 15 or more calls (288 “new”)
  - 375 of them have 5 handsets (155 “new”)
    - Plus 135 with 5 handsets, but < 15 calls
  - 103 duplicate enrollments identified
- Steps in progress to increase collection focus on LVDID languages

How Call Collection Works

- Database tracks speakers, phone #'s, availability, languages, call activity
- System dials out during availability hours
  - Subjects can call in at any time
- Caller is asked to speak his/her name, and to describe phone set (via keypad menu)
- Active lines kept on hold until they can be "bridged"
  - Bridging is quasi-random
- Topic is announced, ulaw streams from each channel are stored in separate files

Call Collection (cont.)

- After 9m30s, announce that call is done
  - Terminate at 10 minutes if lines are active
- Each morning, calls from previous day are uploaded, multiplexed, and fed through automatic speech endpoint detection
- If "speech" on both sides, and sum >=5min, call is made available for manual audit
  - Calls with empty sides stop here
- Web site allows subjects to track their status
  - They can ask/complain about call rejection

How Auditing is Done

- Subjects are presented to auditor in order of #calls completed (singletons last)
- Auditor works on one subject at a time
- For each call:
  - Hear response to “speak your name” prompt
  - Visually review full waveform of single side, play back ~2min of snippets (beginning, middle, end)
  - Judge Spkr-ID, noise, echo, conversation, language (All_Eng, Some_Eng, No_Eng)
  - Where necessary, mark signal or conversation quality as “Unusable”, or mark “technical problem”

Auditing Process (cont.)

- “Wrong voice” flagged in audit table as “change of subj_id”
- Auditor can have two or more subjects on screen at once, to compare different subj_id’s
- Separate GUI available for reviewing enrollment data, to search for duplicates (one person enrolled with multiple subj_ids)
- Subj_id corrections are propagated across call data, with history info kept in “remarks” table (keyed by subj_id)

Comparing Mixer 2 vs. 3

- Both collections used same hardware, same basic call collection system, T-1
- Mixer3 added spoken-name recording, used different SQL schema
- Mixer3 recruits a wider range of languages (for LVDID)
- Auditing focus/task is basically same, but Mixer3 audit interface is a big improvement over the Mixer2 process

Comparing Mixers (cont.)

- Phone/Mic Type statistics

Mixer 1-2 Phone-type/Mic-type Distributions

	spkr-phone	headset	earbud	handheld	
cell	282	417	319	2307	24.4%
cordless	190	762	130	3278	32.0%
landline	234	484	224	5007	43.6%
	5.2%	12.2%	4.9%	77.7%	13634

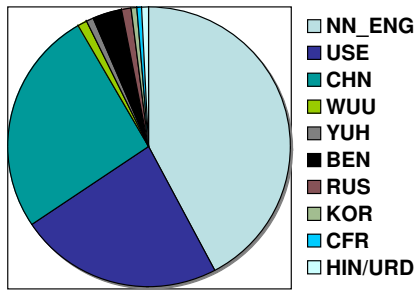
Mixer 3 Phone-type/Mic-type Distributions

	spkr-phone	headset	earbud	handheld	
cell	1024	758	260	1797	46.6%
cordless	236	387	34	1124	21.6%
landline	521	494	119	1477	31.7%
	21.6%	19.9%	5.0%	53.4%	8231

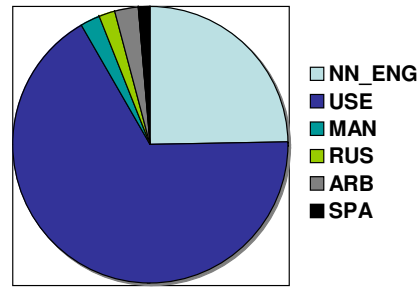
(missing sides = subjects did not respond)

Language characteristics

Mixer3 sides by lang.



Mixer1-2 sides by lang



"Simple" mistakes…

Systems are more sensitive to "corpus noise"

- Not all "gamers" were found in time.
  - LDC could use a good SRE system (or a fusion?) to assist auditors, and time to apply it
- NIST lacked audit data on languages used
  - Need to improve joint LDC/NIST planning, create a more explicit delivery spec
- "Empty" segments in model/test data
  - Extra QC after selection, loop-back to LDC, and more time budgeted for data prep