The Mixer 4 and 5 Corpora supporting SRE08

Linda Brandschain, Chris Caruso, Christopher Cieri, David Graff, Abby Neely, Kevin Walker

{brndschn|carusocr|ccieri|graff|aneely|walkerk}@ldc.upenn.edu

University of Pennsylvania
Linguistic Data Consortium

Acknowledgements

Thanks to the following who have supported the Mixer 4 and 5 projects via sponsorship and/or consultation.

- Joe Campbell (MIT-LL)
- George Doddington (SRI)
- Jack Godfrey (DoD)
- Fred Goodman (MITRE)
- Craig Greenburg (NIST)
- Mike King (ITIC)
- Tina Kohler (DoD)
- Alvin Martin (NIST)
- Nikki Mirghafori (ICSI)
- Nelson Morgan (ICSI)
- Katie McGuire (ICSI)
- Christian Mueller (ICSI)
- Doug Reynolds (MIT-LL)
- Wade Shen (MIT-LL)

Mixer Data

- Some properties of robust Speaker Recognition systems
  - text, channel, and language independence
- Data for system development and evaluation should support those requirements
  - multiple, variable samples per speaker
    - conversational speech with variation of topics
    - variation in speech genre/act
  - collection channels also vary across or even within sessions
    - subjects use multiple telephone handsets
    - some sessions recorded via multiple channels
  - multiple languages sampled
    - collections in multiple languages and dialects
    - Bilingual collections in which subjects use at least two target languages (one per session)

Motivation

- Mixer supports research and development of speaker recognition systems robust to variation in:
  - language and dialect: more than 35 languages and multiple dialects of English, Chinese, Arabic and Spanish
  - channel: telephone + 8 to 14 microphones
  - conversational situation: telephone conversation, interviews, reading; words, phrases, sentences, transcripts, stories
  - Demographics: gender, age, location

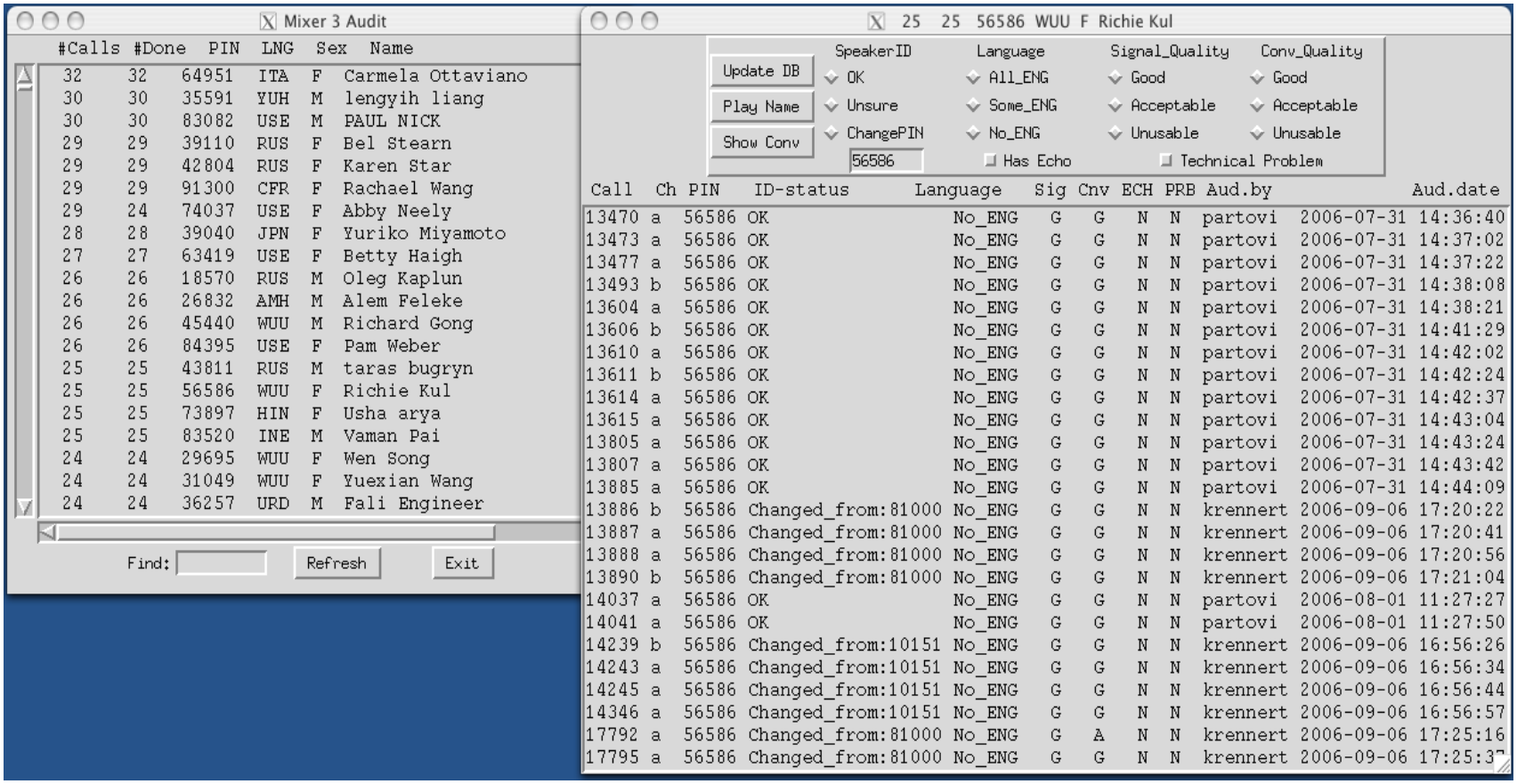
Mixer Platform Design

- Mixer platform designed to address changing telephony
  - Issues Encountered
    - increased cell phone use
    - inexpensive domestic and international calling rates
    - rise in use of call forwarding and call-screening
  - Solutions
    - reduce hours of the study
    - exploit all lines available to robot operator
    - reduce impediments to matching subjects
    - allow any pairing, including duplicates
    - over recruit
    - set goals 20 – 25% higher than required by project sponsors
    - lower per call payment; large completion bonuses
    - encourage subjects to give true, narrow availability schedule
    - increase robot activity to combat increased miss ratio

Mixer Call Platform

- Mixer style studies begin when participant pool >= 200
- Mixer 4 & 5 conducted simultaneously
- > 40 topics cycled
  - current political and social issues, religion, hobbies, sports, etc
  - no penalty for speaking “off topic” so long as conversation is topical
  - participants could refuse call after hearing the topic of the day
- Auditing
  - calls audited for length, sound quality, quantity/suitability of speech.
  - participants who reached their goal were deactivated

Auditing Tool



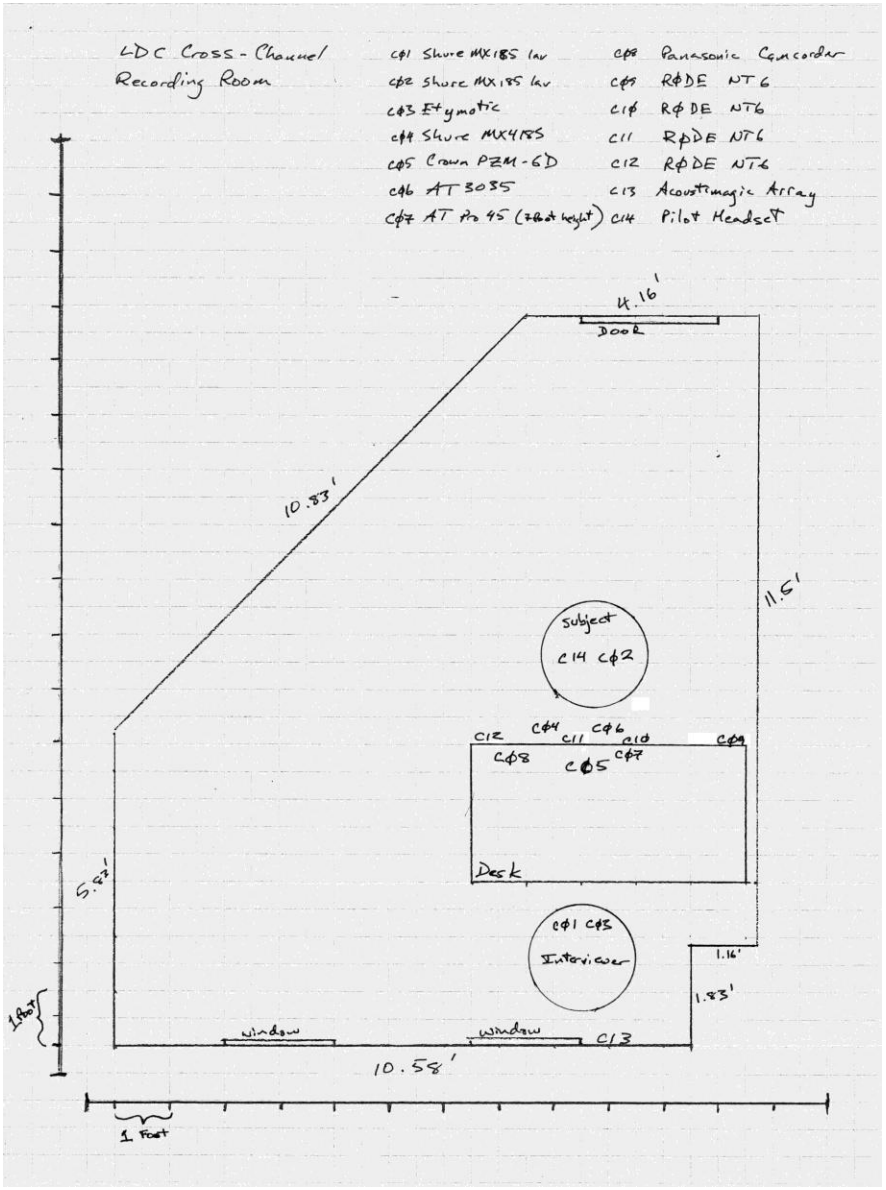
Telephone Speech Collection System

- Windows XP computer with Dialogic T-1 interface, 2GB RAM, 500GB RAID storage, 16GB Solid-State storage
- Fisher style application, built upon Intel CT-ADE SDK, Perl, and MySQL DBMS
- 24 line T-1 circuit from Verizon, three Toll-Free inbound numbers
- Realtime call progress interactions processed on solid state drive
- Call participants are paired dynamically

Cross Channel Collection System

- Fourteen microphones in Mixer 4 & 5 with diverse target applications (Mixer 1 & 2 used eight microphones)
- Windows XP computer, 1GB RAM, 300GB storage
- MOTU 896HD and 8pre audio interfaces
  - balanced low-impedance inputs
  - connected to host computer via firewire
- mchan_rec multi-channel audio recording applet
  - simple interface customized for interview recording
  - automatically identifies sound hardware
  - remote controllable via TCP/IP

LDC Interview Room Diagram

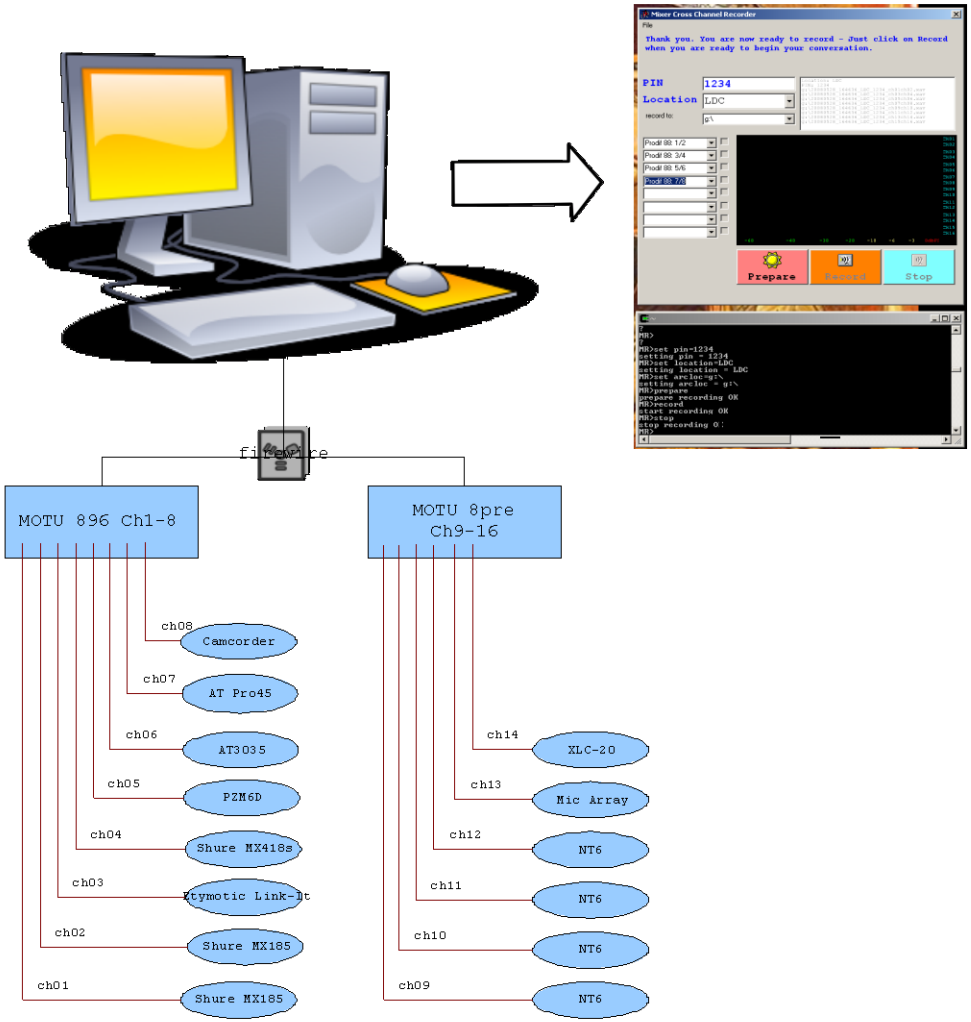


Multi-Channel Set-Up

Ch	Microphone	Placement	Subject/Reference
1	Shure MX185 Lavalier	Interviewer	
2	Shure MX185 Lavalier	Subject	
3	Etymotic Micro-array	Interviewer	
4	Shure MX418X Podium	Desk Front	Center
5	Crown PZM-6D	Desk Top	Center
6	Audio Technica AT3035	Desk Front	Right
7	Audio Technica Pro45	Hanging	Center
8	Panasonic Camcorder	Desk Top	Right
9	Rode NT6	Desk Front	Far Left
10	RODE NT6	Desk Front	Center Left
11	RODE NT6	Desk Front	Center Right
12	RODE NT6	Desk Front	Center Far Right
13	AcoustiMagic Array	Wall Mounted	Center
14	Lightspeed Headset	Subject	

Cross Channel Recording Room

Collection System Diagram



Comparison of Phases

	SB	M1	M2	M3	M4	M5
Core Calls (8+)	✓	✓		✓	✓	✓
Variable Environments	✓					
Unique Handset (4+)	✓	✓	✓	✓		✓
Extended Data (20+)		✓	✓	✓	✓	
Multilingual (4+)		✓		✓		
Cross Channel (2 or 4)		✓	✓		✓	
Transcript Reading (2+)		✓				✓
Interviews (6)						✓

Mixer 1 & 2 Description

- Mixer 1 and 2: Oct. 2003 - May 2005

- Standard telephone component: Mixer 1
  - 600 participants, 10 calls
  - 100 subjects, 4 calls @ Arabic, Mandarin, Russian, Spanish
  - 100 subjects, 4 calls using unique handsets

- Extended call component: Mixer 1
  - 100 subjects, 20 extended calls

- Extended call component: Mixer 2
  - 450 subjects, 20 extended calls

- Cross-channel component: Mixer 1
  - 100 subjects, 4 cross channel calls

- Cross-channel component: Mixer 2
  - 100 subjects, 4 cross channel calls

- Transcript-reading component: Mixer 1
  - 100 subjects, 4 cross channel calls sessions

Microphones Used In Mixer 1 & 2

Channels	Microphones	Position
1	Audio Technica AT3035 Studio Mic	Mic stand
2	Shure MX418S Gooseneck Mic	Mic stand
3	Crown PZM Soundgrabber II	Desk
4	Audio Technica AT Pro45	Hanging
5	Jabra Cellphone Earwrap mic	Headworn
6	Motorola Cellphone Earbud	Headworn
7	Olympus Pearlcorder	Desk, left
8	Radio Shack Computer desktop mic	desk, center

Mixer 1 & 2 Outcome

- Standard telephone component:
  - ~14,500 calls collected
  - ~2290 subjects recorded, ~50% did 10+ calls

- Cross-channel component:
  - 8-channel in-office recordings of phone caller
  - 1035 recordings made at 3 sites (LDC, ICSI, ISIP)
  - 240 subjects, 199 did 4 or more XC-calls

- Transcript-reading component:
  - 940+ conversations transcribed for reading samples
  - 200+ reading sessions recorded
  - 63 subjects, most did between 1 and 4 sessions

Mixer 3 Description

- Registration open to adult, fluent speakers of more than 30 languages
- Demographics collected for all participants including date of birth, gender, native language, other languages spoken and age at which English proficiency acquired, country, state and city where born and where raised, years of education, height, weight and ethnicity
- Participants asked to make 15, 10 minute calls using the Mixer platform
- Robot operator placed outgoing calls according to participants schedules, attempting to match native speakers
- Bonuses were given for verified calls in languages other than English or calls made using unique handsets
- Repeat pairings permitted but minimized due to the size of subject pool
- To increase the yield, some hours/days were devoted to specific languages. Native speaking participants of the designated language were notified via email and the Language Hour schedule was available on the Mixer 3 website.

Mixer 4

- Mixer 4 was designed to support speaker recognition research and technology evaluations

- Demographics of Subject Pool
  - Native Speakers of American English
  - 25% from Philadelphia
  - 25% from Berkeley
  - 50% from the entire US , however we recruited heavily in Georgia, Texas, Illinois, and New York

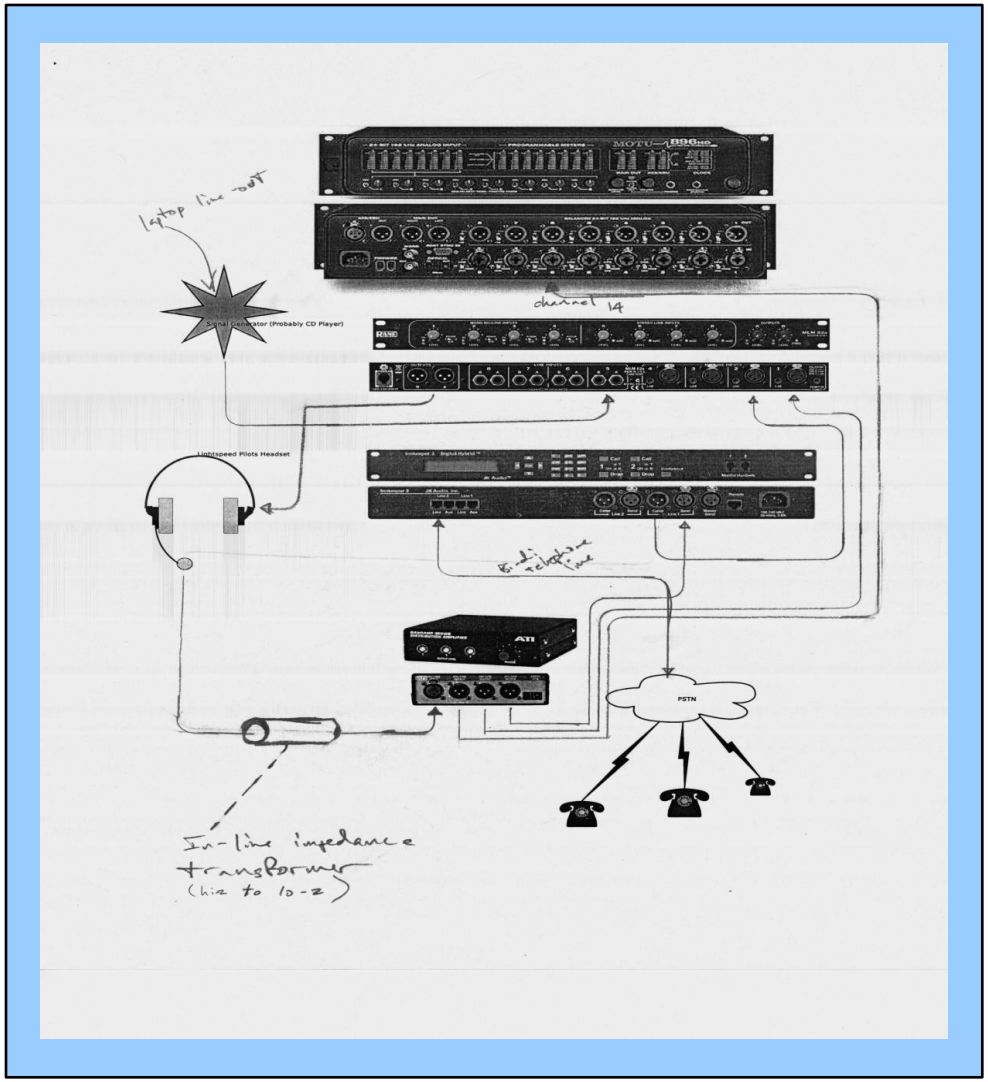
- Original Goals for Mixer 4
  - 400 Subjects that made 10, 10 minute phone calls
  - 200 Visited one of our two sites where they completed 2 cross-channel call
  - 100 Participants were asked to complete extended data calls (20 x 10-minute phone calls)

Mixer 5

- Mixer 5 focused on cross-channel recordings of face to face interviews where the goal is to elicit speech within a variety of situations.
- Demographics of Subject Pool
  - Native language undefined, however participants had to be fluent in English
  - Approximately 50% recruited from Philadelphia, PA
  - Approximately 50% recruited from Berkeley, CA
- Goals for Mixer 5
  - 300 Participants
  - Each Participant must complete 6 half hour sessions completed in no less than 6 days. Each session had a mandatory 30 minute break between sessions.
  - Each of the 300 Participants must also complete 10 ten-minute phone calls
  - Foreign language calls were encouraged but not required
  - Bonuses were issued for the completion of 4 unique phone calls
- High/Low Vocal Effort Phone Calls
  - ~1/3 of Mixer 5 Participants completed these calls
  - Lightspeed XLC-20 headphones provide 40db passive acoustic isolation
  - High Vocal Effort: Input audio is 65dB and relative levels of the mix components are 30% side-tone, 40% remote speaker and 30% white noise.
  - Low Vocal Effort: Input audio is 65dB with no white noise.

Equipment into the telephone Configuration for introducing feedback channel

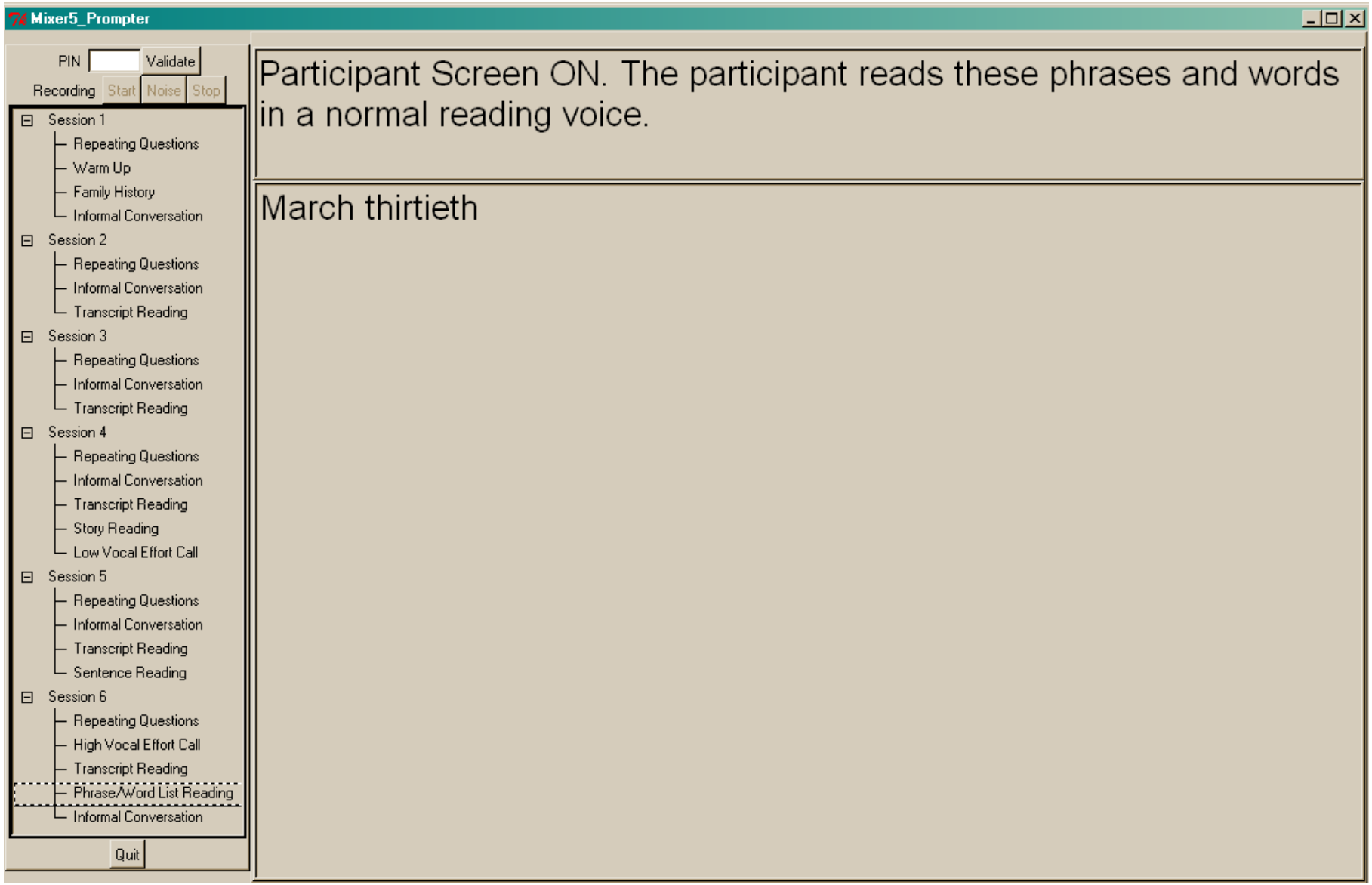
- We wanted to be able to optionally introduce a variable amount of static into the subjects earpiece. The subject wears an acoustically isolating headset with an electret boom mic.
- The subjects headset is fed three signals: their interlocutor's speech, their own microphone output, and optionally a looped recording of white noise.
- The microphone output is also set across the telephone network to the interlocutor.



Mixer 5 Interview Protocol

Session Number	1	2	3	4	5	6	Min
Repeating Questions	1	1	1	1	1	1	6
Warm Up	4						4
Family Personal	5						5
Informal Conversation	20	9	14	9	9	9	70
Transcript Reading		20	15	10	15	10	70
Story Reading				5			5
Sentence Reading					5		5
Phase/Word List Reading						5	5
Low Vocal/Effort				5			5
High Vocal/Effort						4	4
Total Sessions	30	30	30	30	30	30	180

Mixer 5 Prompter



Mixer 4 &5 Outcomes

- Mixer 4 & 5: June - Dec. 2007 (Phone calls)
  - 560 new speakers (not in Mixer 3), 434 did 8+ calls
  - 256 call sides were cross-channel sessions
  - 138 x-chan speakers, 115 did two x-chan sessions
  - 92 speakers have 2 x-chan and 8+ "normal" phone calls
- Mixer 5: Feb. - Dec. 2007 (Interviews)
  - ~1900 interview sessions of ~30 min each
  - 343 speakers recorded, 289 did 6 sessions (3 hrs/speaker)
  - 300 speakers made at least one phone call
  - 239 speakers have 6 IV sessions and 8+ phone calls

Completed Work

- Mixer 1 & 2
  - in LDC publication pipeline, to be released in 2008
- Mixer 3
  - used in SRE06, SRE08 & LRE07; remainder reserved for future evaluation
- Mixer 4
  - collection completed
  - part used in SRE08 remainder reserved for future evaluation
- Mixer 5
  - interview collection finished
  - phone call collection completed
  - part used in SRE08; remainder reserved for future evaluation

Current Work

- Greybeard – Voice and Aging
  - subjects from previous CTS collections recruited to make new calls
- Potential new studies
  - conduct Mixer 5 style interviews in other languages
  - conduct studies like Mixer 1 & 2 in other languages
- All Mixer data will be published after its use in technology evaluations.

Future Work

- Recording Ideas for new Mixer style studies
  - Unsystematically changing the room acoustics for interviews using movable acoustic "baffles"
  - Systematically varying the distance between the interviewer and the subject
  - Changing the "tone" of the interviews
  - Getting large numbers of speakers from concentrated dialectal regions (South Philly, Brooklyn)
  - Move microphones to more "invisible" places