



# **Speaker Recognition: Building the Mixer 4 and 5 Corpora**

**Linda Brandschain, Christopher Cieri, David Graff,  
Abby Neely, Kevin Walker**

**{brndschn|ccieri|graff|aneely|walker}@ldc.upenn.edu**

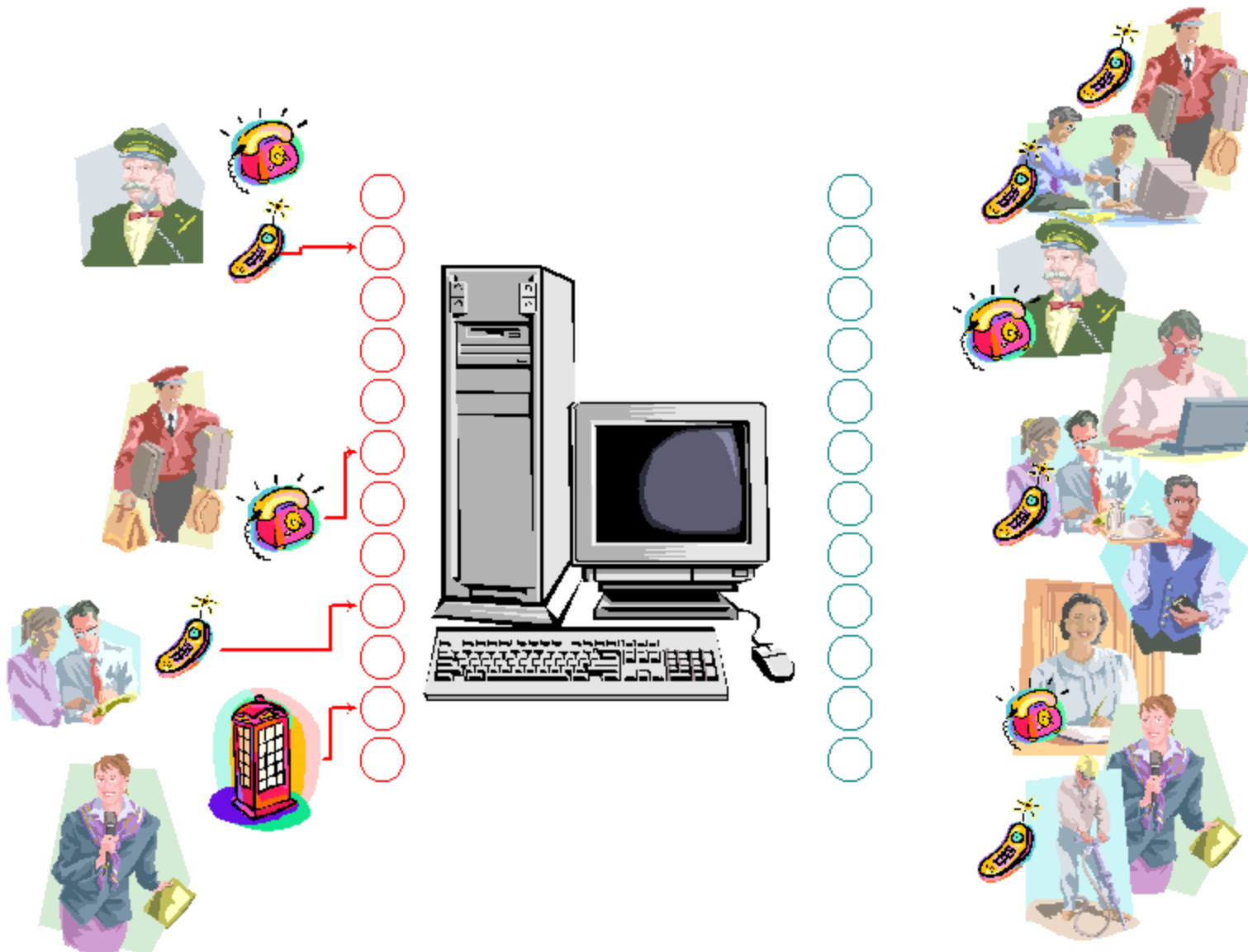
**University of Pennsylvania  
Linguistic Data Consortium**

- ◆ Mixer supports R&D of speaker recognition systems robust to variation in:
  - language: Arabic, Mandarin, Russian, Spanish
  - channel: telephone + 8 to 14 microphones
  - conversational situation: telephone conversation, interviews, reading words, phrases, sentences, transcripts, written texts
- ◆ Mixer 4
  - channel variation
- ◆ Mixer 5
  - channel
  - conversational situation

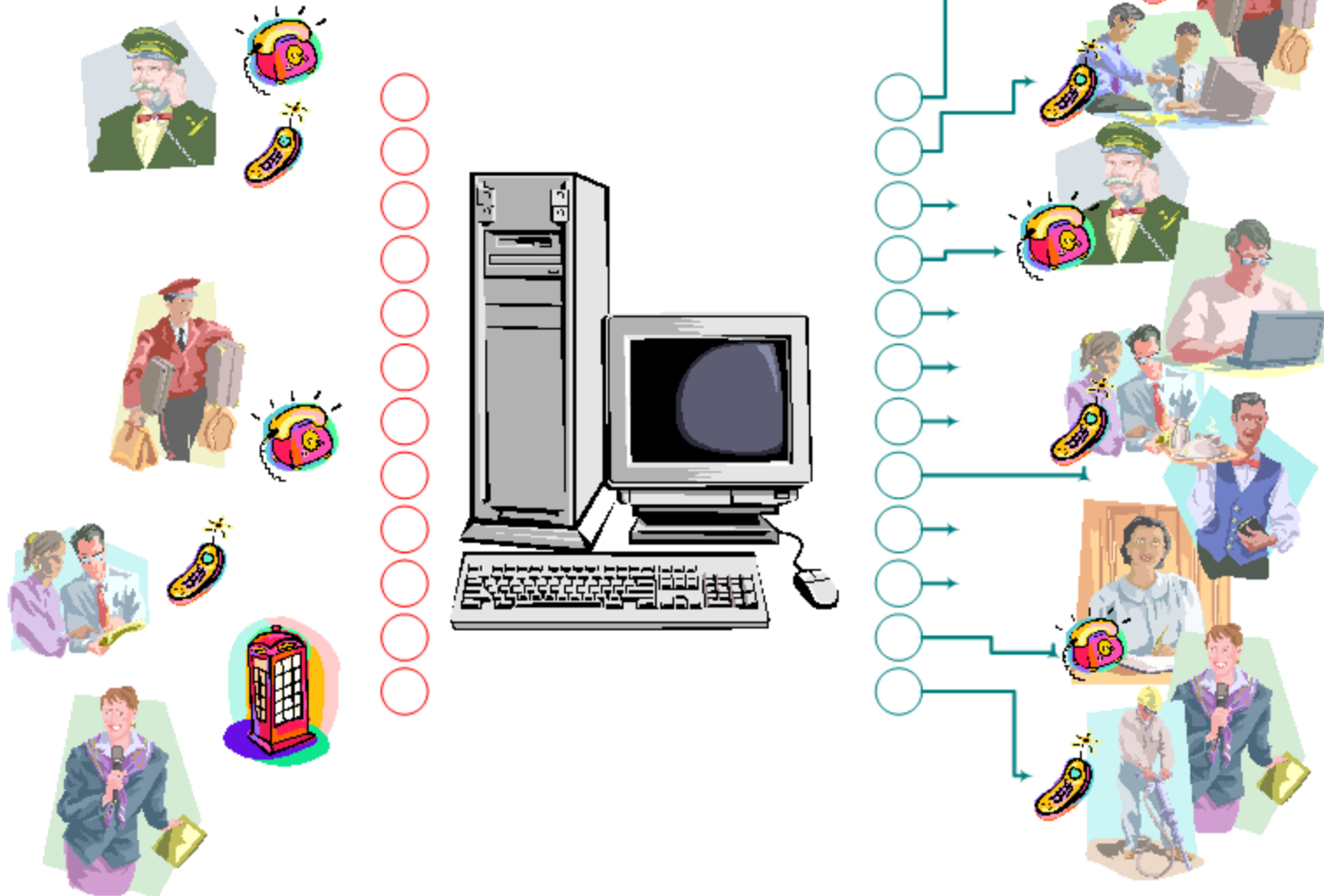
# Comparison of Phases

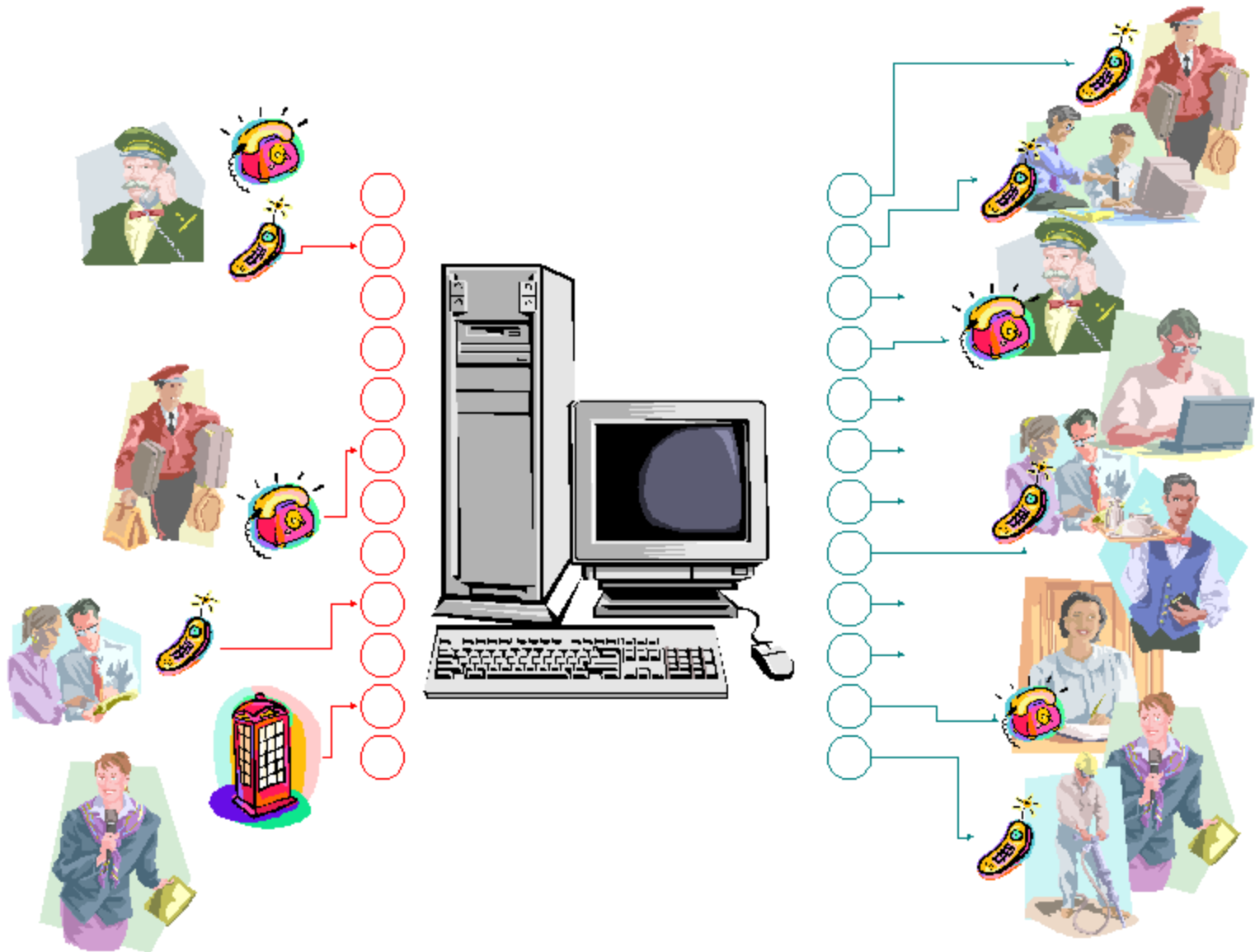
	SB	M1	M2	M3	M4	M5
Core Calls (8+)	✓	✓		✓	✓	✓
Variable Environments	✓					
Unique Handset (4+)	✓	✓	✓	✓		✓
Extended Data (20+)		✓	✓	✓	✓	
Multilingual (4+)		✓		✓		
Cross Channel (2 or 4)		✓	✓		✓	
Transcript Reading (2+)		✓				✓
Interviews (6)						✓

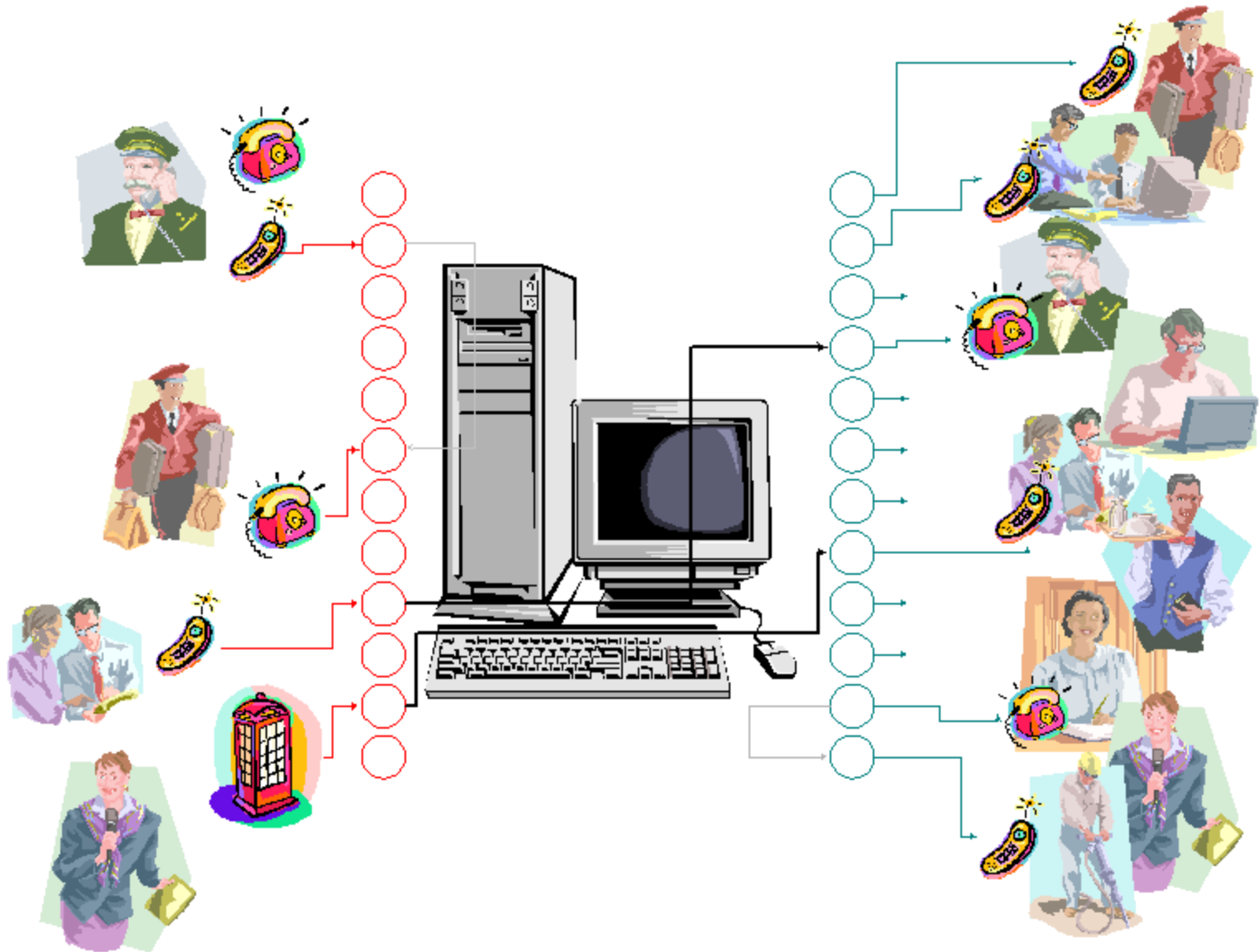
- ◆ Mixer platform designed to address changing telephony
  - Issues Encountered
    - increased cell phone use
    - inexpensive domestic and international calling rates
    - rise in use of call forwarding and call-screening
  - Solutions
    - reduce hours of the study
    - exploit all lines available 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



# Protocol

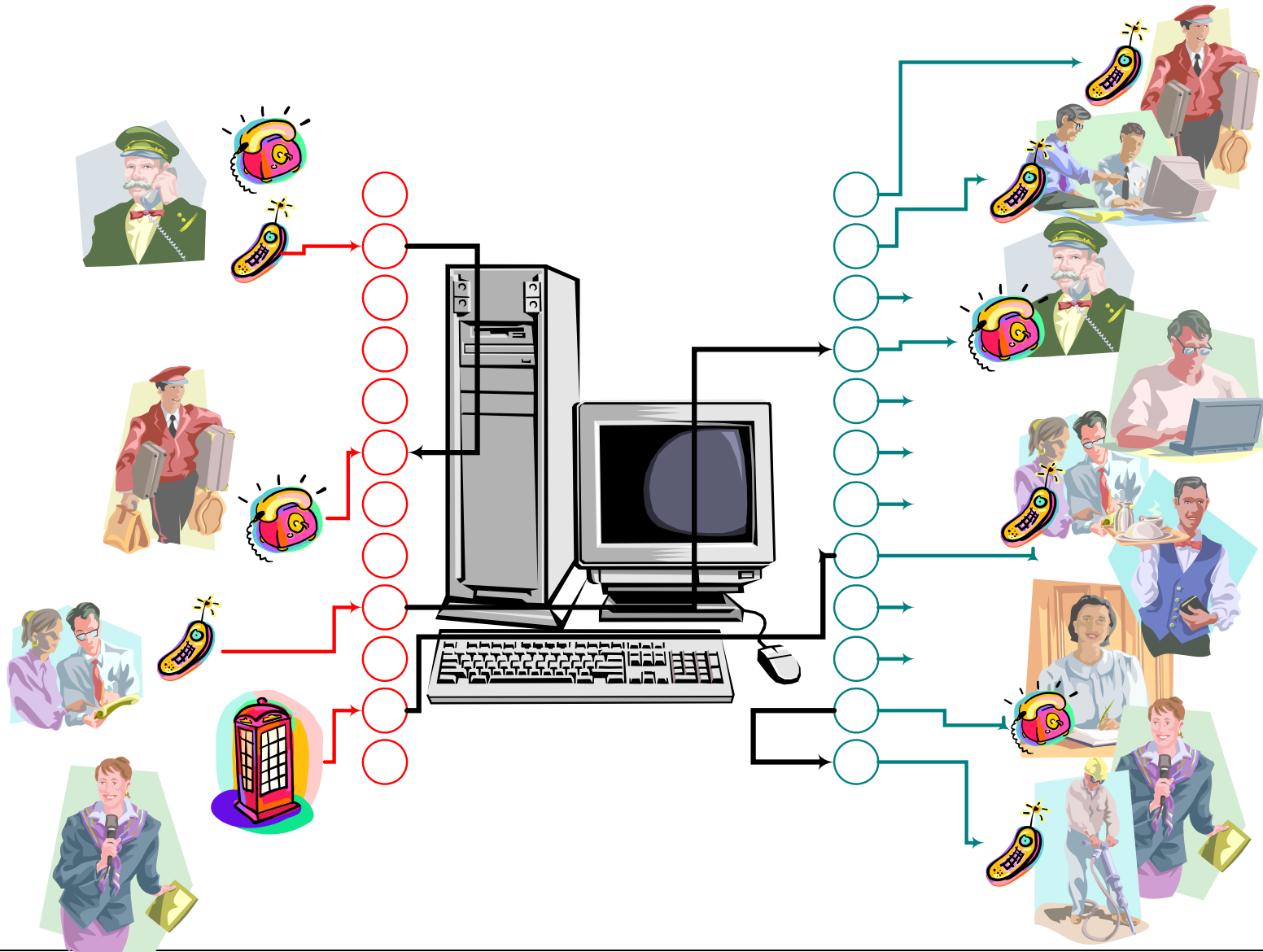






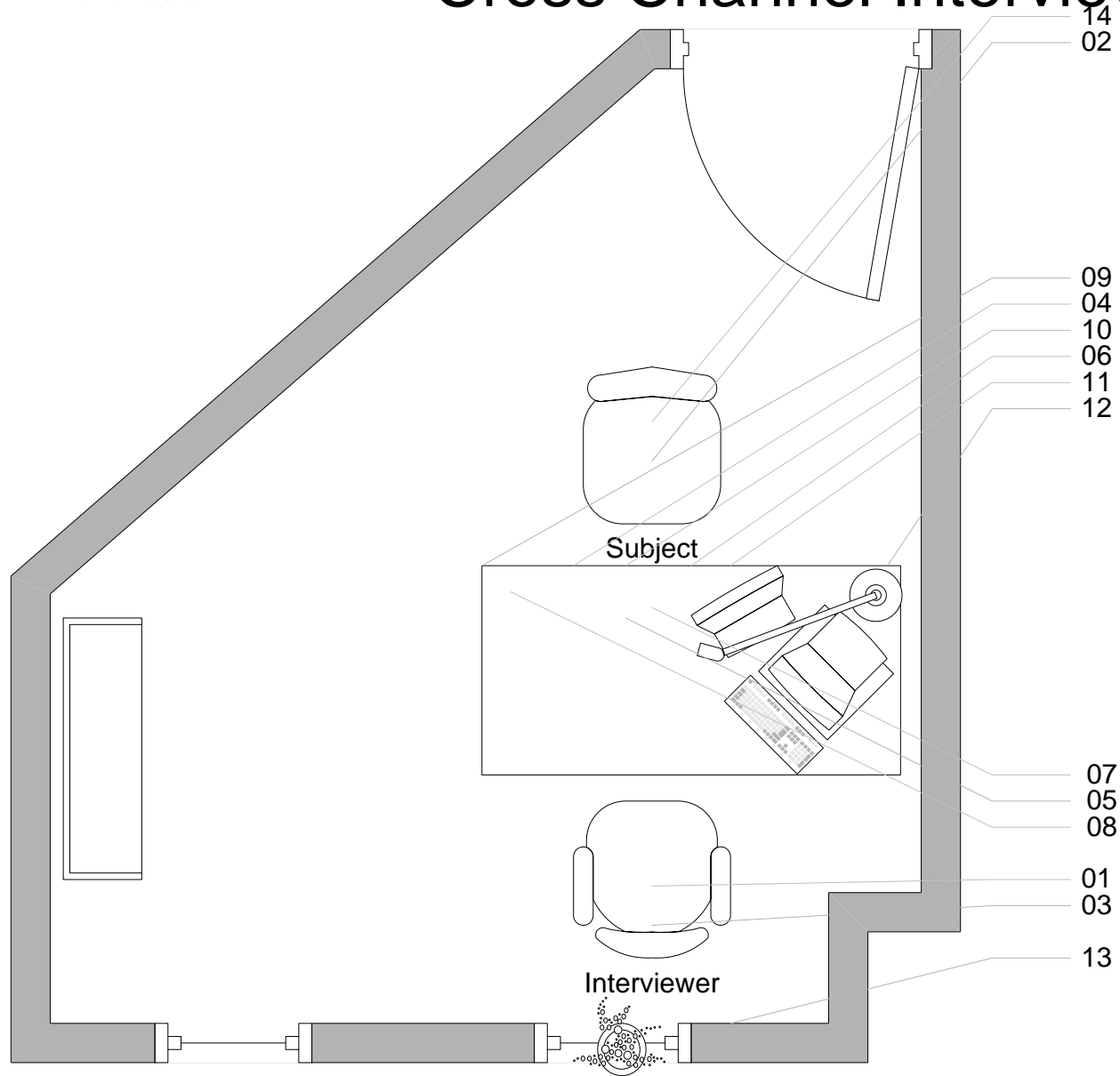


# Diagram of Platform Protocol



- ◆ Mixer 4 & 5 conducted simultaneously
- ◆ Studies began when participant pool  $\geq 200$
- ◆ 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

# Cross Channel Interview Room



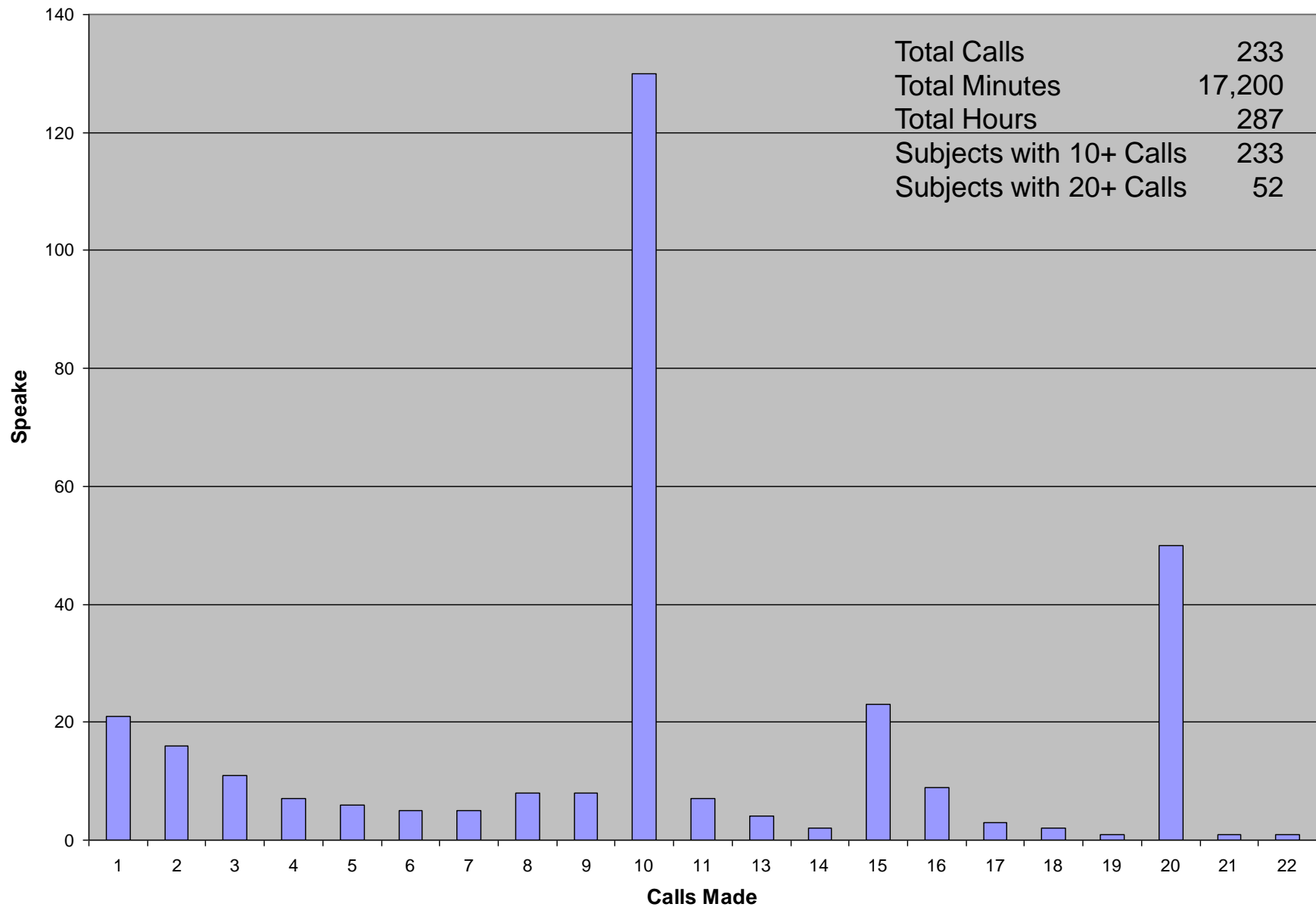


# Multi-Channel Set-Up

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	

- ◆ 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)

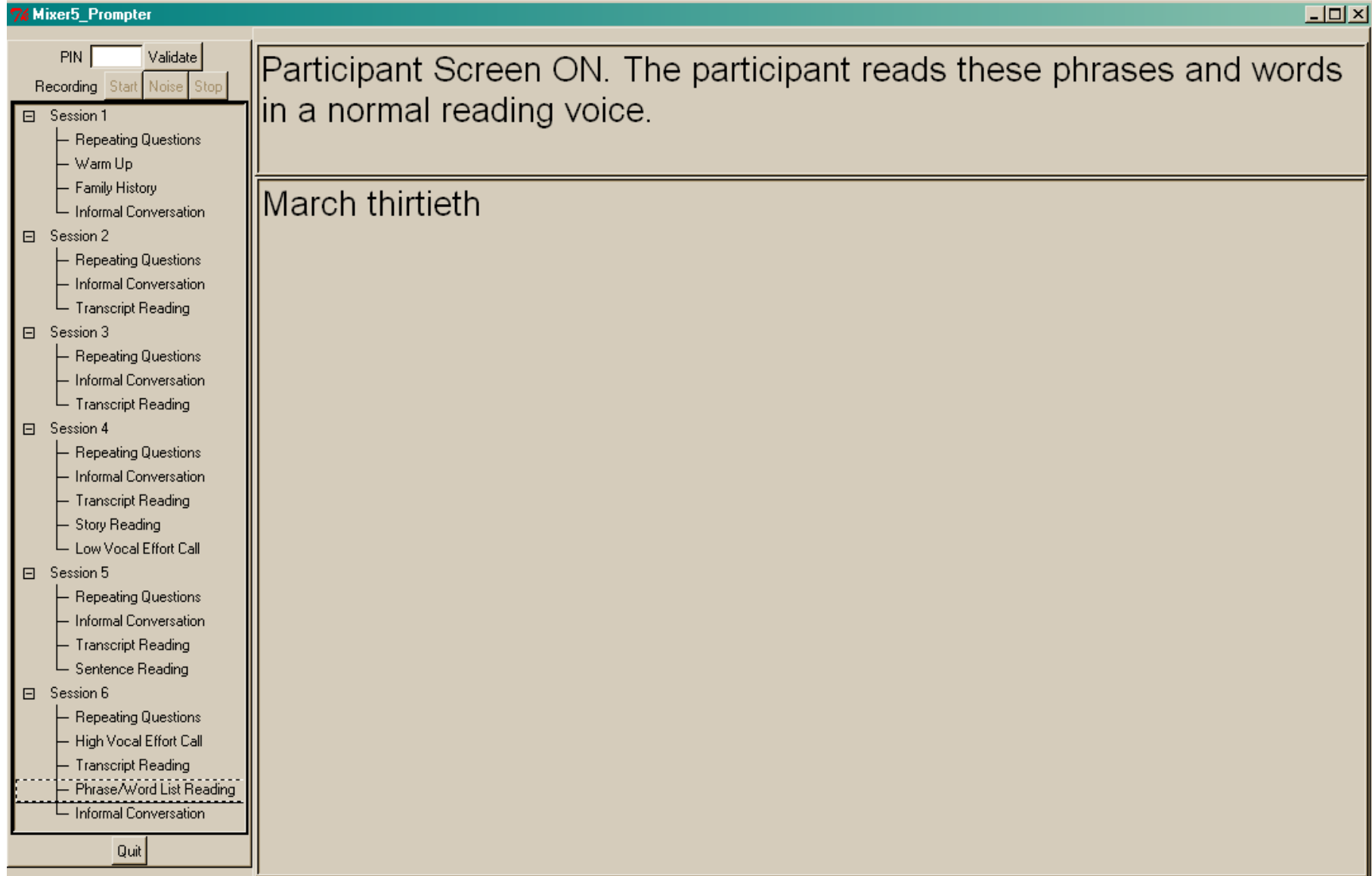
# Mixer 4 Call Yields



- ◆ 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.



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
Phrase/Word List Reading						5	5
Low Vocal/Effort				5			5
High Vocal/Effort						4	4
Total Session	30	30	30	30	30	30	180



**Mixer5\_Prompter**

PIN  Validate

Recording **Start** Noise Stop

- Session 1
  - Repeating Questions
  - Warm Up
  - Family History
  - Informal Conversation
- Session 2
  - Repeating Questions
  - Informal Conversation
  - Transcript Reading
- Session 3
  - Repeating Questions
  - Informal Conversation
  - Transcript Reading
- Session 4
  - Repeating Questions
  - Informal Conversation
  - Transcript Reading
  - Story Reading
  - Low Vocal Effort Call
- Session 5
  - Repeating Questions
  - Informal Conversation
  - Transcript Reading
  - Sentence Reading
- Session 6
  - Repeating Questions
  - High Vocal Effort Call
  - Transcript Reading
  - Phrase/Word List Reading
  - Informal Conversation

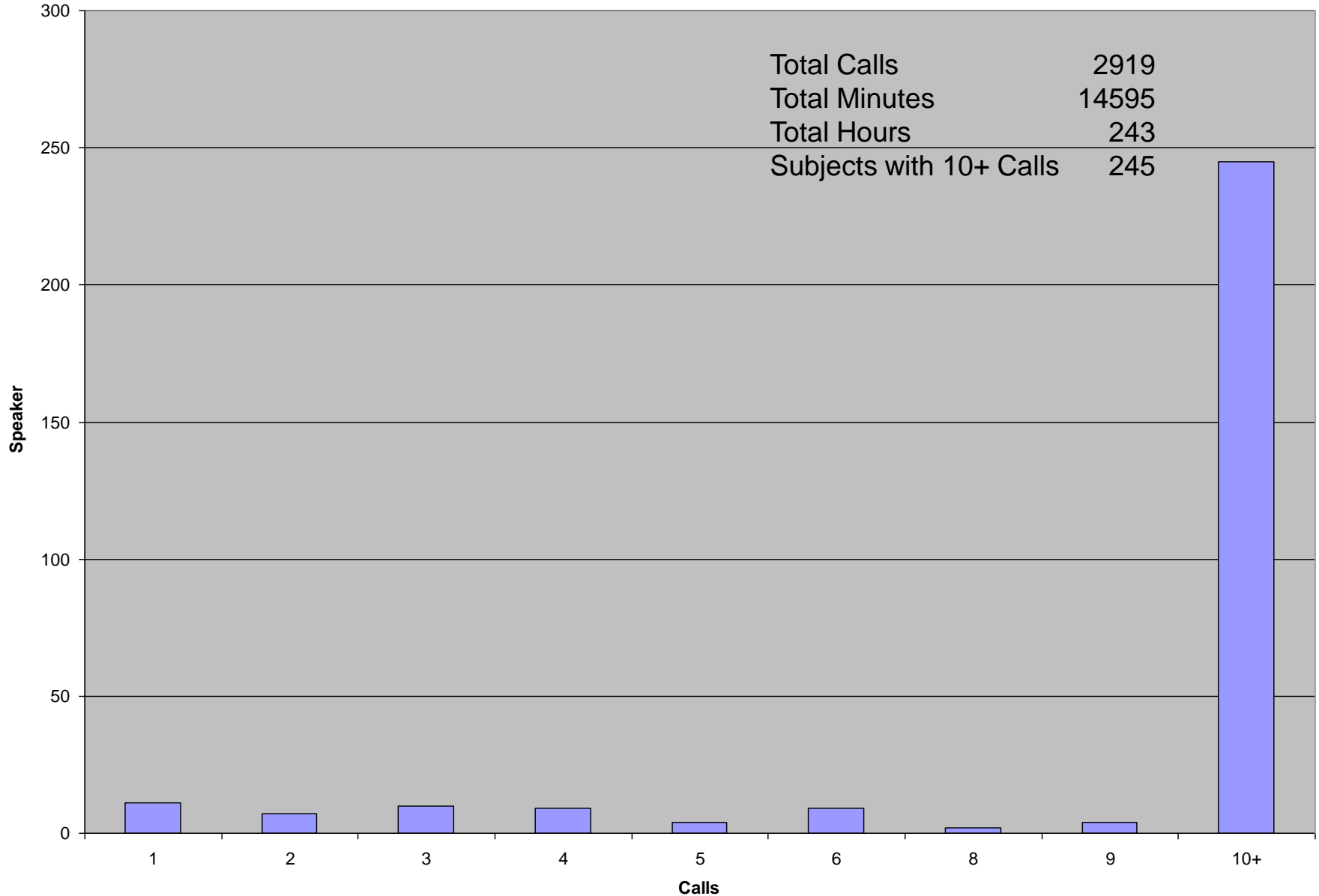
Quit

Participant Screen ON. The participant reads these phrases and words in a normal reading voice.

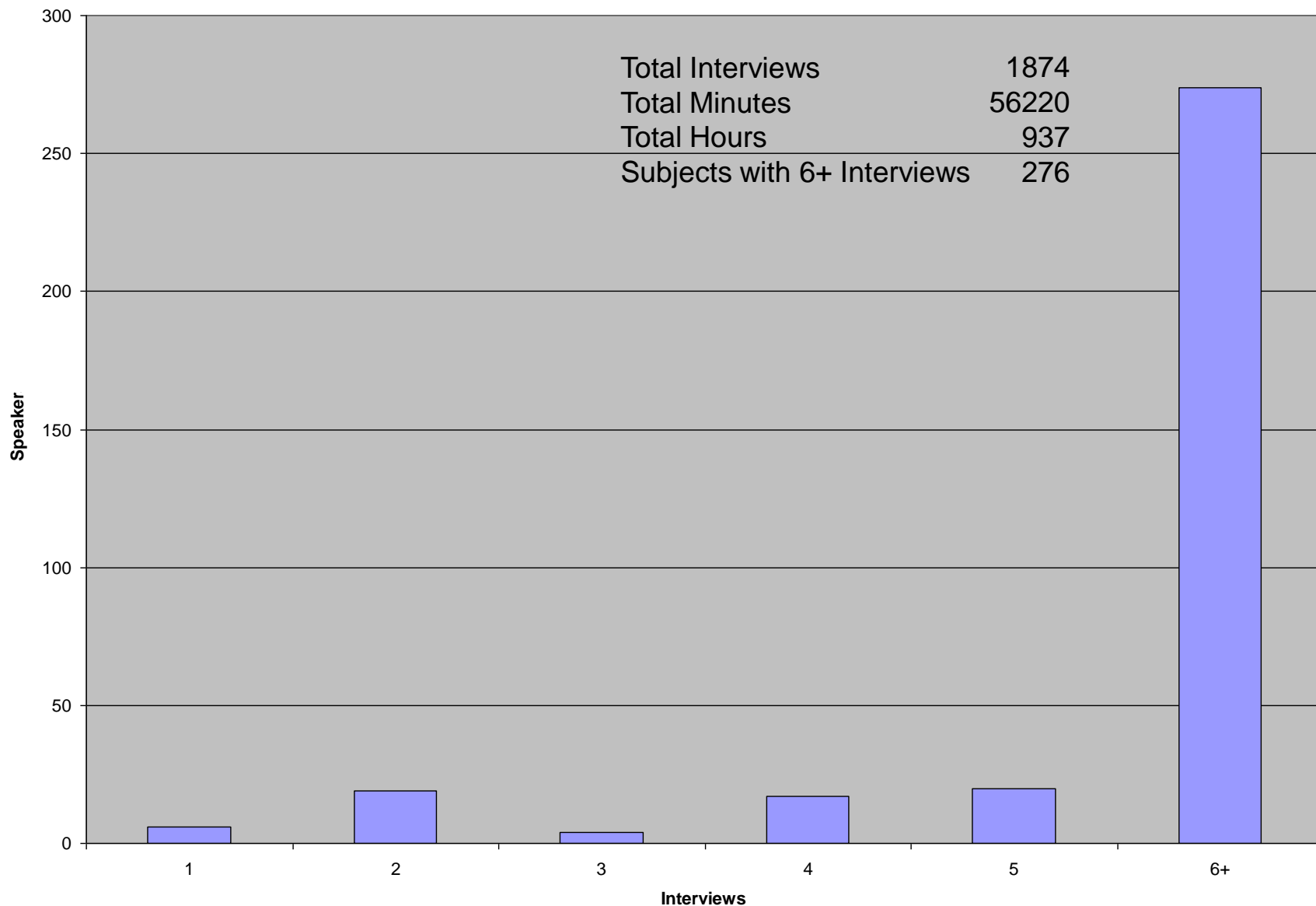
March thirtieth



# Mixer 5 Call Yields



# Mixer 5 Interview Yields



- ◆ Mixer 1 & 2
  - in LDC publication pipeline
- ◆ Mixer 3
  - used in SRE06 & LRE07; remainder reserved for future evaluation
- ◆ Mixer 4
  - collection underway
  - part used in SRE08 remainder reserved for future evaluation
- ◆ Mixer 5
  - interview collection ahead of schedule
  - phone call collection also well underway
  - part used in SRE08; remainder reserved for future evaluation
- ◆ Mixer 6 (Graybeard)
  - subjects from previous CTS collection return to join
- ◆ Potential new studies
  - conduct Mixer 5 style interviews in other languages
  - conduct studies like Mixer 1 & 2 but involving other languages
- ◆ All Mixer data will be published after its use in technology evaluations.