
The Mixer and Transcript Reading Corpora: Resources for Multilingual, Crosschannel Speaker Recognition Research*

Christopher Cieri¹, Walt Andrews², Joseph P. Campbell³, George Doddington⁴, Jack Godfrey², Shudong Huang¹, Mark Liberman¹, Alvin Martin⁴, Hirotaka Nakasone⁵, Mark Przybocki⁴, Kevin Walker¹

1. Linguistic Data Consortium, 3600 Market Street, Philadelphia, PA 19104

2. U. S. Department of Defense, MD, USA

3. MIT Lincoln Laboratory, Lexington, MA, USA

4. National Institute of Standards and Technology, Gaithersburg, MD, USA

5. Federal Bureau of Investigation, Quantico, VA, USA

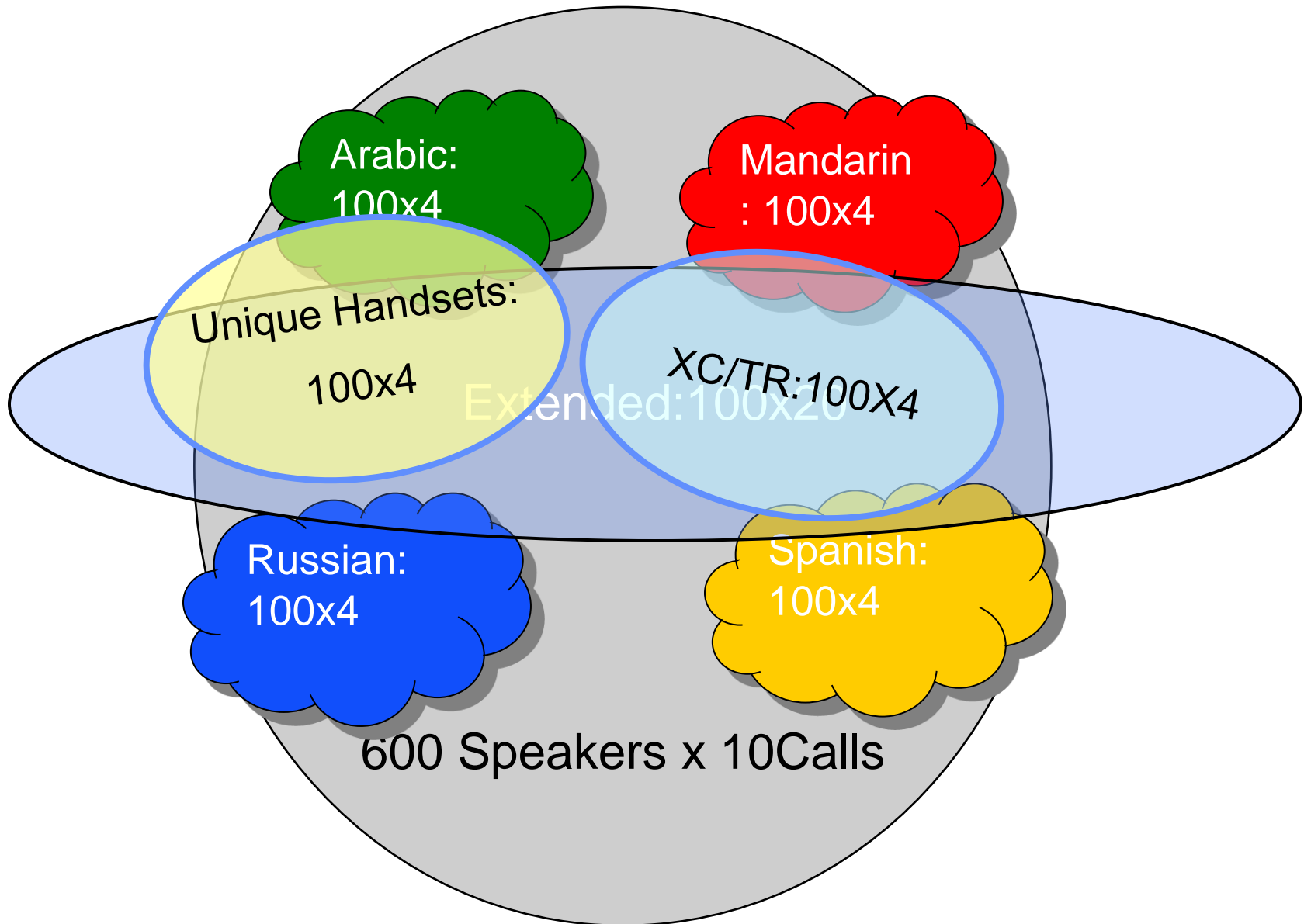
**ccieri@ldc.upenn.edu, waltandrews@gmail.com, j.campbell@ieee.org,
george.doddington@nist.gov, godfrey@afterlife.ncsc.mil, shudong@ldc.upenn.edu,
myl@ldc.upenn.edu, alvin.martin@nist.gov, hnakasone@fbiacademy.edu,
mark.przybocki@nist.gov, walkerk@ldc.upenn.edu**

***This work was supported by funding from the Federal Bureau of Investigation, the Department of Defense and the Intelligence Technology Innovation Center under Air Force contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.**

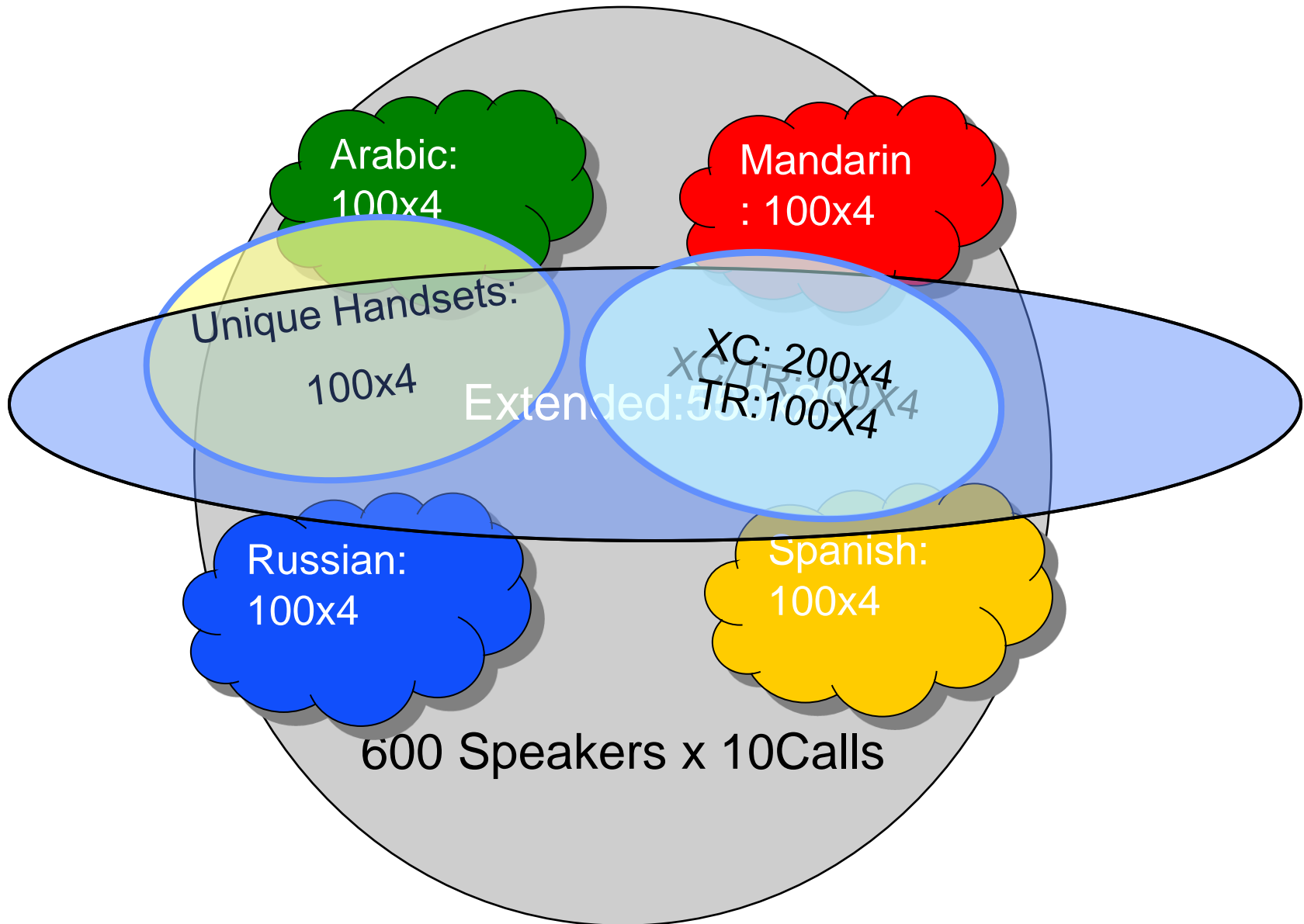
- **Requirements for speaker recognition**
 - text-independent
 - channel-independent
- **New requirements**
 - multiple languages
 - bilingual speakers with train/test mismatch in language
 - extended notion of varying channel
- **Program**
 - multi-language, bilingual, cross-channel collection with
 - data disseminated to multiple research sites and
 - used in metrics based, common task evaluation leading to
 - system performance evaluation and improvement
 - identification of remaining challenges

- **Switchboard style collection**
 - each speaker makes multiple calls
 - brief: six-minutes in duration
 - speaking to other participants
 - using assigned topics
 - collected as 4-wire data
- **Extensions**
 - use variant of Fisher Protocol
 - » adapted to today's telephone use (voice mail)
 - multiple languages collected simultaneously
 - bilingual speakers
 - intensively cross-channel

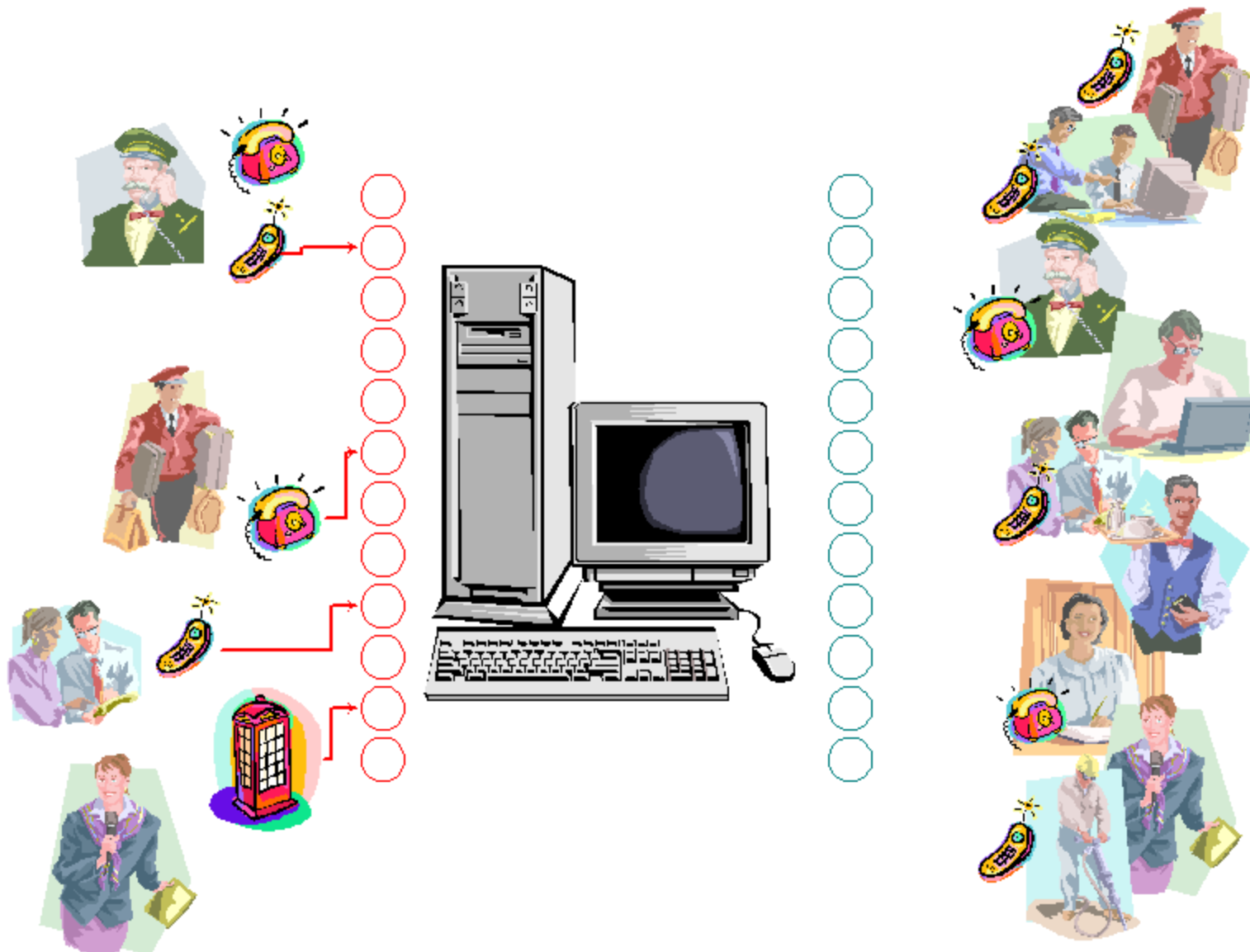
Phase I



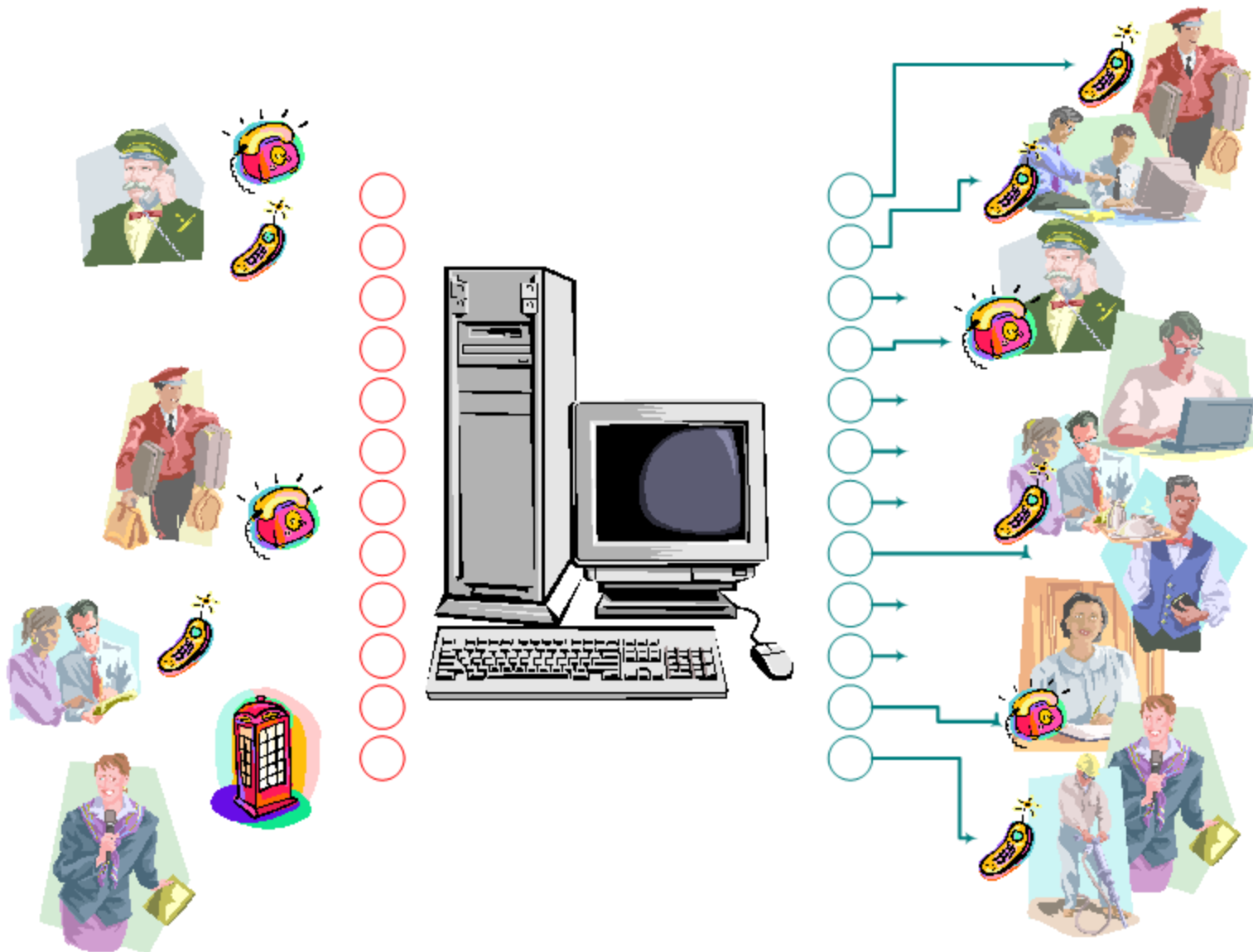
Phase II



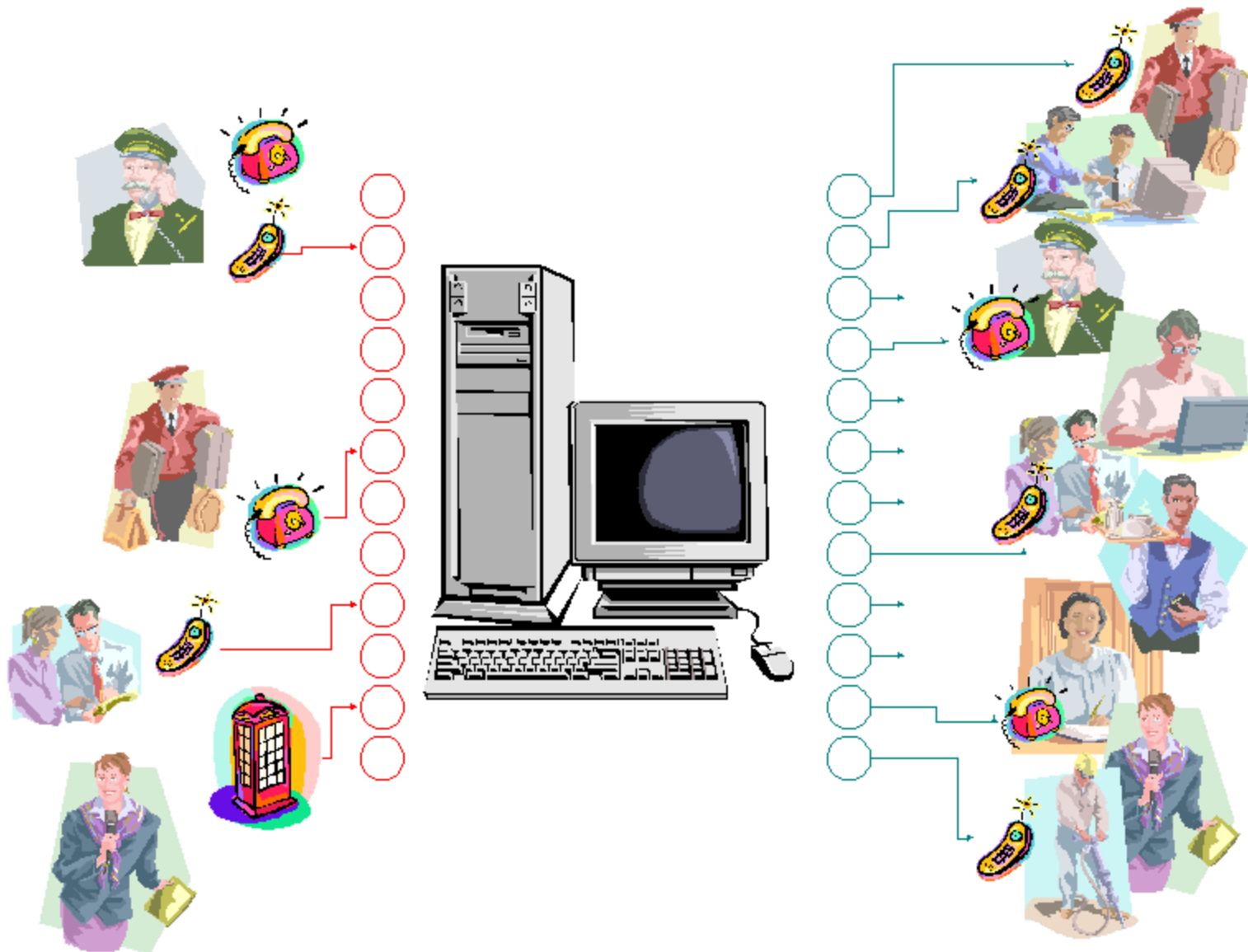
Protocol



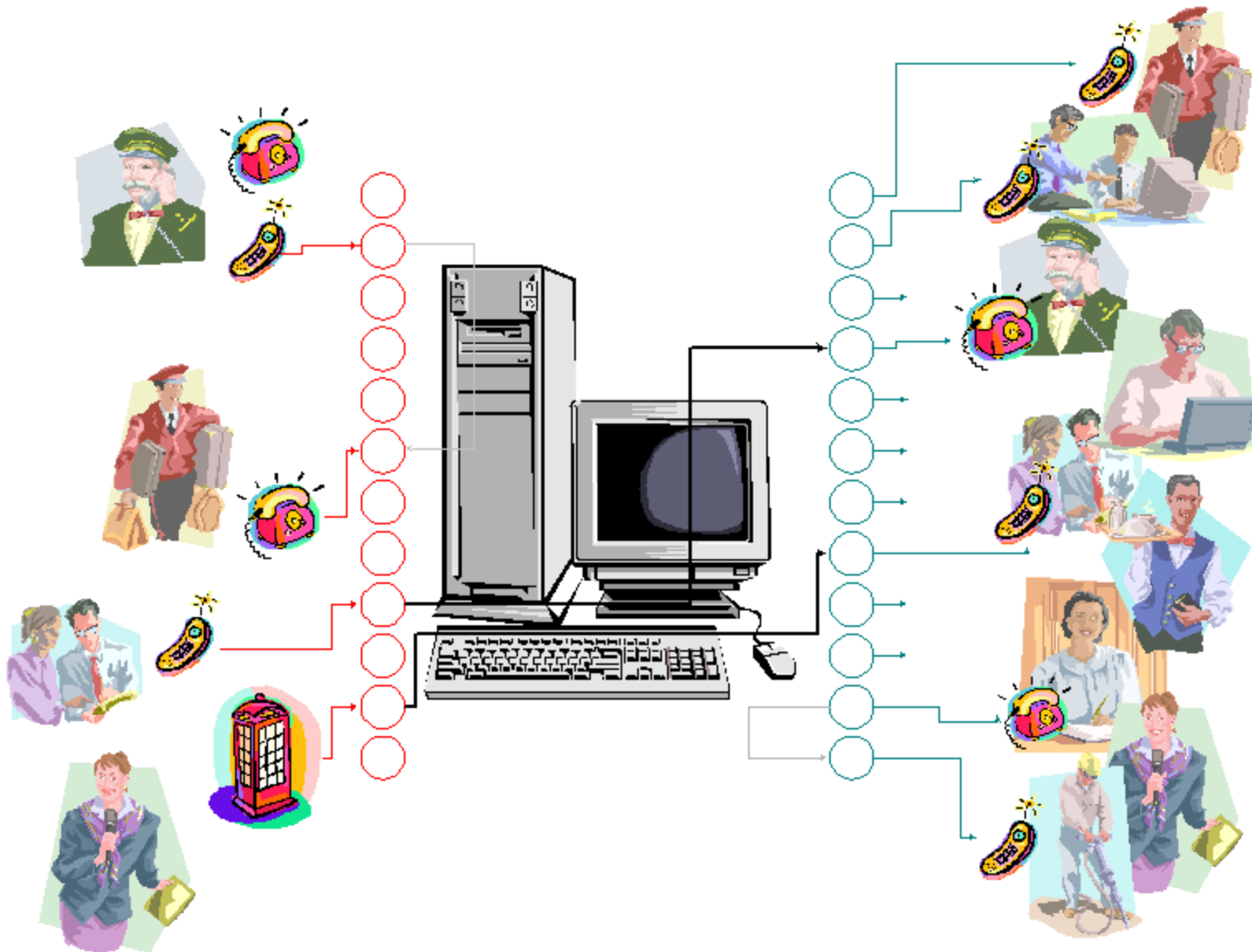
Protocol



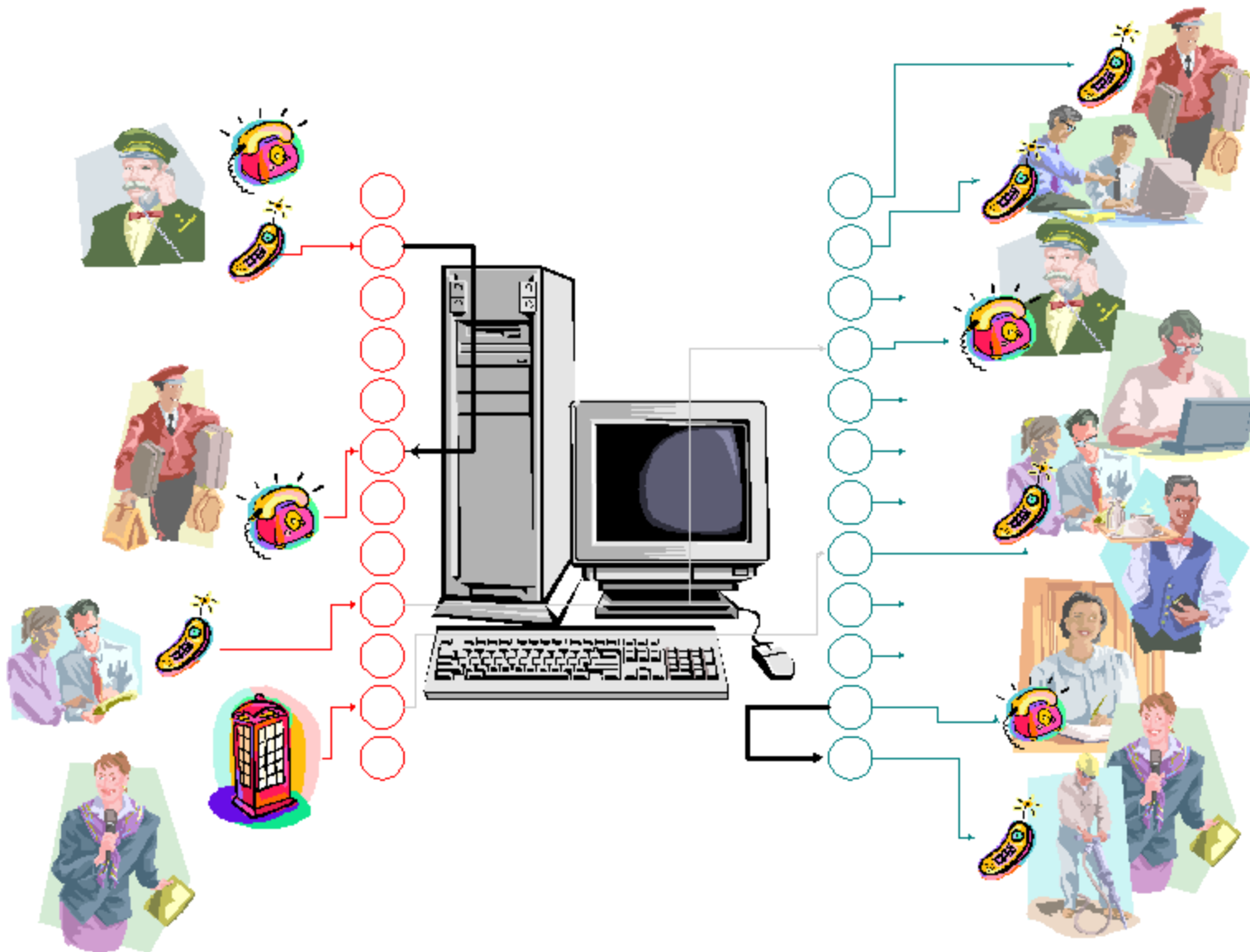
Protocol



Protocol



Protocol



Implementation

- **Facts of life**
 - 50% of recruits participate
 - 70% of participants accomplish 80% of study goals
- **So**
 - Recruit twice as many subjects as needed
 - Set goals 20-25% higher than needed
 - » where study needs 4 calls, ask subjects for 5
 - » where study needs 10 calls, ask subjects for 12
 - » where study needs 20 calls, ask subjects for 25
- **Recruiting**
 - little required due to energy generated by **Fisher** recruiting
 - subjects sign up via Internet or by calling 800 number
 - From Fisher allow multilingual, graybeard and x-channel subjects; block others

Topics

- **5. Music: “Music is the universal language of mankind” - Henry Longfellow. Is music a way to bring people together from different backgrounds? Can music help to achieve peace or help make the world a better place, or is music the source of social divisions?**
- **7. Music: What do you think about downloading music and movies from the internet? Do you have sympathy for the music industry or believe that unauthorized downloading is a violation of copyright law?**
- **34. Craftsmanship: ... For example, as a consumer would you be more interested in purchasing a hand-made individualized item, or something mass-produced for the best cost to function ratio?**
- **36. U.S. obesity: The obesity prevalence rates are at an all-time high right now. ... Do you think that it is a priority that we need to thin down as a country? What do you think are some of the causes of our obesity?**
- **47. Boxers or Briefs: Do you prefer boxer shorts or briefs? What are some of the advantages and disadvantages to either?**
- **65. Food: Is there such a thing as "American" food? Do you like the fact that that there a greater variety of foods are available in the U.S. today then ten years ago?**










Protocols

- **Mixer uses the “Fishboard” protocol**
 - Robot operator live from Noon-3AM EST
 - Unlike Switchboard and CallHome/CallFriend, robot drives study
 - Robot calls subjects at times they list as available during sign-up
 - Subjects may also call robot
 - Pairs any two speakers even if they have already spoken
- **Mixer enhancements**
 - Robot gives priority to speakers of same native language.
 - Some days were devoted to non-English calls.
 - Compensation = core fee + special features + completion bonuses
 - **Multichannel System,**
 - » laptop with two firewire hard drives
 - » multichannel interface, recording application, 8 sensors
 - Each channel sampled at 48Khz with 16bit samples
 - Call collected by robot operator simultaneously
 - Deployed cross channel recording system at four sites
 - » LDC, ICSI, MSU/ISIP => Rutgers

Cross Channel

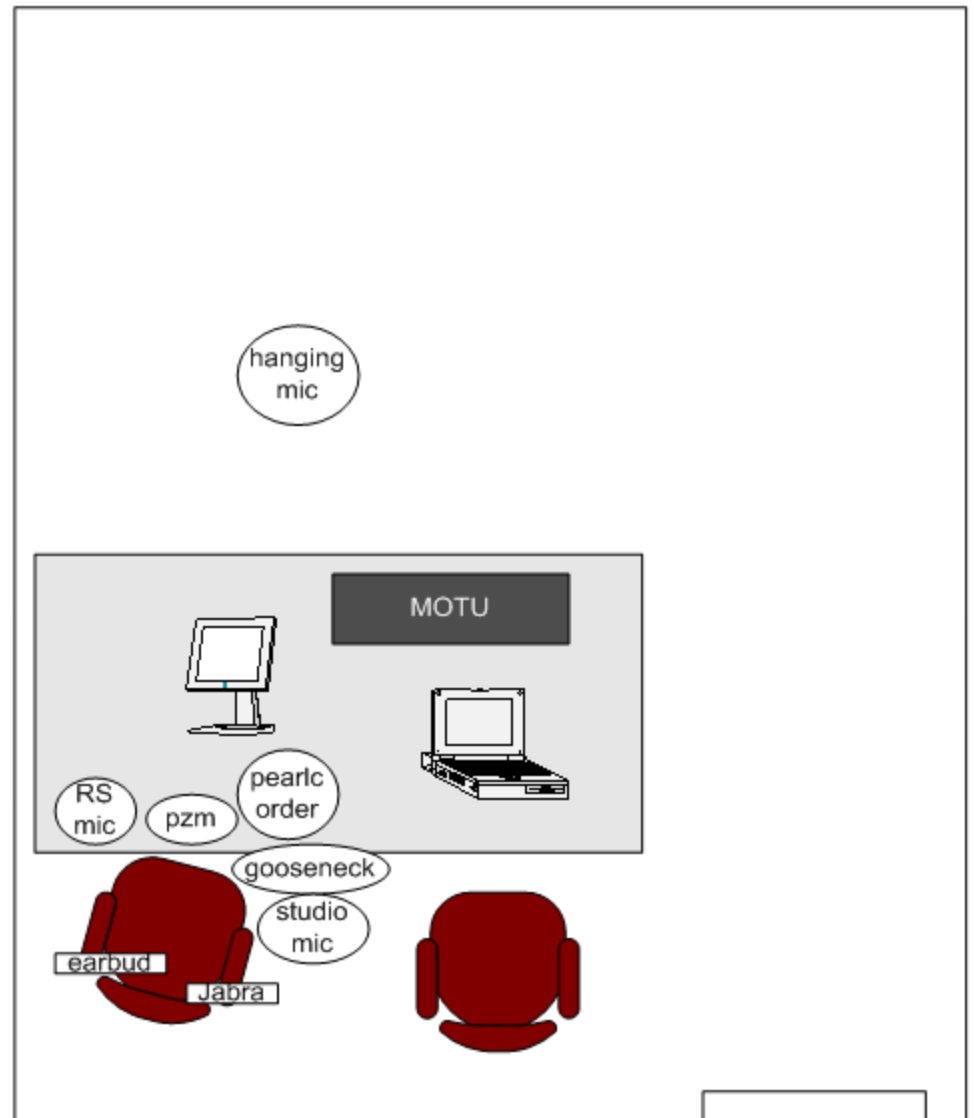
- 8 microphone connected to cross-channel platform (XCP)
- preamplifier used for all 8 channels, provided up to 40dB of gain
- gain set to record the strongest signal for each channel without clipping
- recordings at 48khz, 16bit
- Microphones fixed at set locations over course of collection
 - 2 worn, 1 hanging 7 feet away, 5 on desk 1–2 feet from speaker
- mic placement, usage balance real world use, best practices.
 - mic provide no feedback to speaker level
 - physical constraints
 - 2 cell mic'c use external biased power
- session moderator confirmed that
 - speaker is on axis with the desk/hanging microphones
 - cell mics were worn properly
 - earboom worn over the ear
 - earbud microphone was clipped to the collar

Microphone Details

	Channel, Type	Sensor	Balance, Shielding	Impedance	Power, Comments, Cost
	0. Wireline Telephone				Connects to robot operator, not to XCP
	1. Studio (AT3035)	Cardioid Condenser	Balanced, Shielded	Low impedance	Phantom power <XCP, \$200
	2. Podium (Shure MX418S)	Supercardioid	Balanced, Shielded	Low impedance	Phantom power <XCP, \$185
	3. Hanging (AT Pro 45)	Cardioid Condenser	Balanced, Shielded	Low impedance	Phantom power <XCP, \$91
	4. Dictaphone (Olympus Pearlcor S725)		Line level shielded	Line level (unbalanced)	Bias power <dictaphone wall wart, \$66
	5. Cell Earboom (Jabra EarWrap #43-1914)		Unbalanced manuf. cable, unshielded	High impedance	Required external bias power <3 rd party supply (9 volt battery), \$30
	6. Cell Earbud (Motorola SYN8390)		Unbalanced manuf. cable, unshielded	High impedance	Required external bias power <3 rd party supply (9 volt battery), \$12
	7. Conference (Crown Sound Grabber II)	pressure-zone mic	Unbalanced manuf. cable, shielded	High impedance	Bias power <internal battery (AAA 1.5 volt battery), \$70
	8. Computer (Radio Shack #33-3031)		Unbalanced manuf. cable, shielded	High impedance	Bias power <internal battery (two AA 1.5 volt batteries), \$27

Recording Configuration

- Shure Gooseneck and Audio Technica Studio Mic on microphone stand (1 foot from speaker)
- Radio Shack mic, Crown PZM, and Olympus Dictaphone on Desk (1.5 – 2 feet from speaker)
- Audio Technica HangingMic placed behind desk (7 feet from speaker)
- Cell phone mics (Jabra & Earbud) worn by speaker.
- Fluorescent bulbs were replaced With incandescent lighting.
- An instruction manual with detailed Information about best practices for recording sessions was prepared and stayed in the room throughout the collection. This manual was reviewed by all collection coordinators.



Transcript Reading

- **120 dense, 30 second segments from previously recorded Mixer cross-channel, 1 for each TR subject, selected and transcribed**
- **Selection process maximizes speech from target**
 - Based on auto-segmentation (human in some cases)
 - Minimum 30 seconds from target speaker
 - Maximum density of speech from target subject
 - Segments with low type/token ratio examined by humans
- **Each subject reads their own and other's transcripts in random order**
- **Back-channel transcription visible to the subject**
- **Two or more sessions, each beginning with subject reading own transcript.**
- **Recorded on multi-channel platform and telephone collection platform using same software**
 - multichannel recording software modified to allow external control

Transcript Reading

MCTR_Socket

Calibrate Begin Pause End

Um, do you ever download uh movies, or just music?
No, I've never downloaded movies, actually.
I don't even have, I don't think I have a ~DVD, um, recordable.
Because you can download them um and then burn them onto um ~CDs.
There's a, a ~VCD format.
Oh really?
Mhm.
I didn't even know that. Do you download them?
Mhm.
I don't, but I have a lot of friends who do, so, and it's g- it's cool because you can get the movies right away, which I know is really bad.
And they come out good?
Yeah, they come out really good. Like, the, one of my friends downloaded um
The Matrix, like, The Matrix three before it
had come out in the theaters, or like the day that it came out in the theater.
Oh wow.
And so we could watch it at home like that
That's awesome.
night. Yeah.
And, I think it's just kind of a strategy they use to make people distrust the media that they get, and to, um, sort of start leaning more towards the right. And, um, there's some, um, I guess one specific example I was thinking of was, um, Fox News, which, um, I think everyone would pretty much agree is very right leaning, and um But, you know they act like like, in fact, I think their slogan is something like fair and balanced, um... But, when you really look at their coverage it's very skewed, so, I think, um, that's kind of irresponsible of them and I don't

Prompt Number: 55 Percent Done: 3% Prompt: Yeah, they come Elapsed Time: 528

Prepare Record Stop

Start SRE2005.ppt C:\WINDOWS... F:\MCTR 2005... MCTR_Soc... Mixer Cross C...

- Recording begins with subject consent
- Each session begins with subject reading own snippet.
- Backchannel is indented, gray.
- Change in background color marks speaker change.
- No coaching, no imitation. Auditor does correct mis-readings.
- Prompting software records time at which each prompt is highlighted.
- “Pause” pauses prompting but not recording.

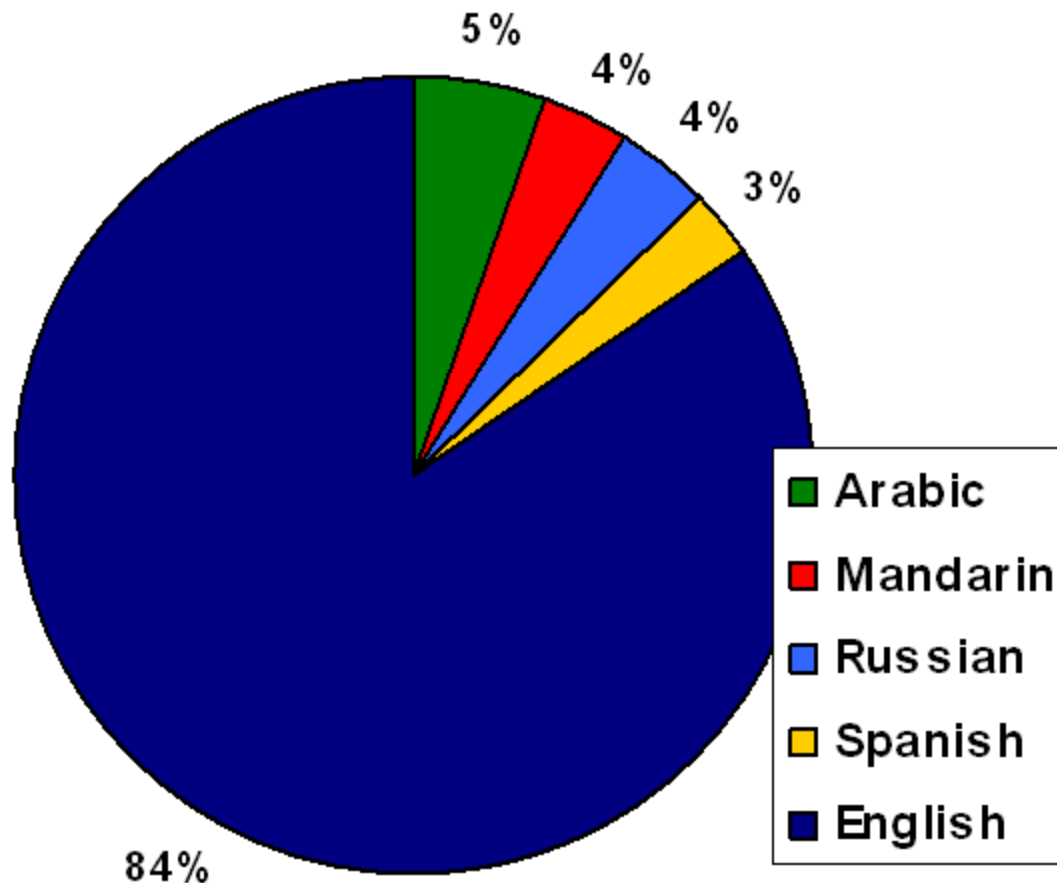
Transcript Reading

PROMPT 87.6060 0 (text)
PROMPT 95.9179 2 (text)
PROMPT 101.3858 3 (text)
REPEAT 104.8508 3 (text) Repeat the current prompt, please.
PROMPT 111.7607 5 (text)
PROMPT 118.5505 6 (text)
PROMPT 120.8438 8 (text)
PROMPT 122.2458 9 (text)
PROMPT 128.3446 10 (text)
PROMPT 131.0985 12 (text)
PROMPT 137.6479 13 (text)
REPEAT 147.9227 13 (text) Repeat the current prompt, please.
REPEAT 157.9371 13 (text)
REPEAT 163.0745 13 tax cut. Basically, they want, they, they,
they espouse the opinion that um, the less tax, the better
because then the companies can invest in Repeat the
current prompt, please.
PROMPT 173.0188 14 (text)

Yields to Date

	Targeted	Achieved
Base (x10 Calls)	650	1124
Arabic (x4 Calls)	100	127
Mandarin (x4 Calls)	100	113
Russian (x4 Calls)	100	106
Spanish (x4 Calls)	100	100
Extended (x20 Calls)	550	563
Super-Extended (x30 Calls)	0	134
Cross Channel (x4 Calls)	200	201
Transcript Reading	100	100

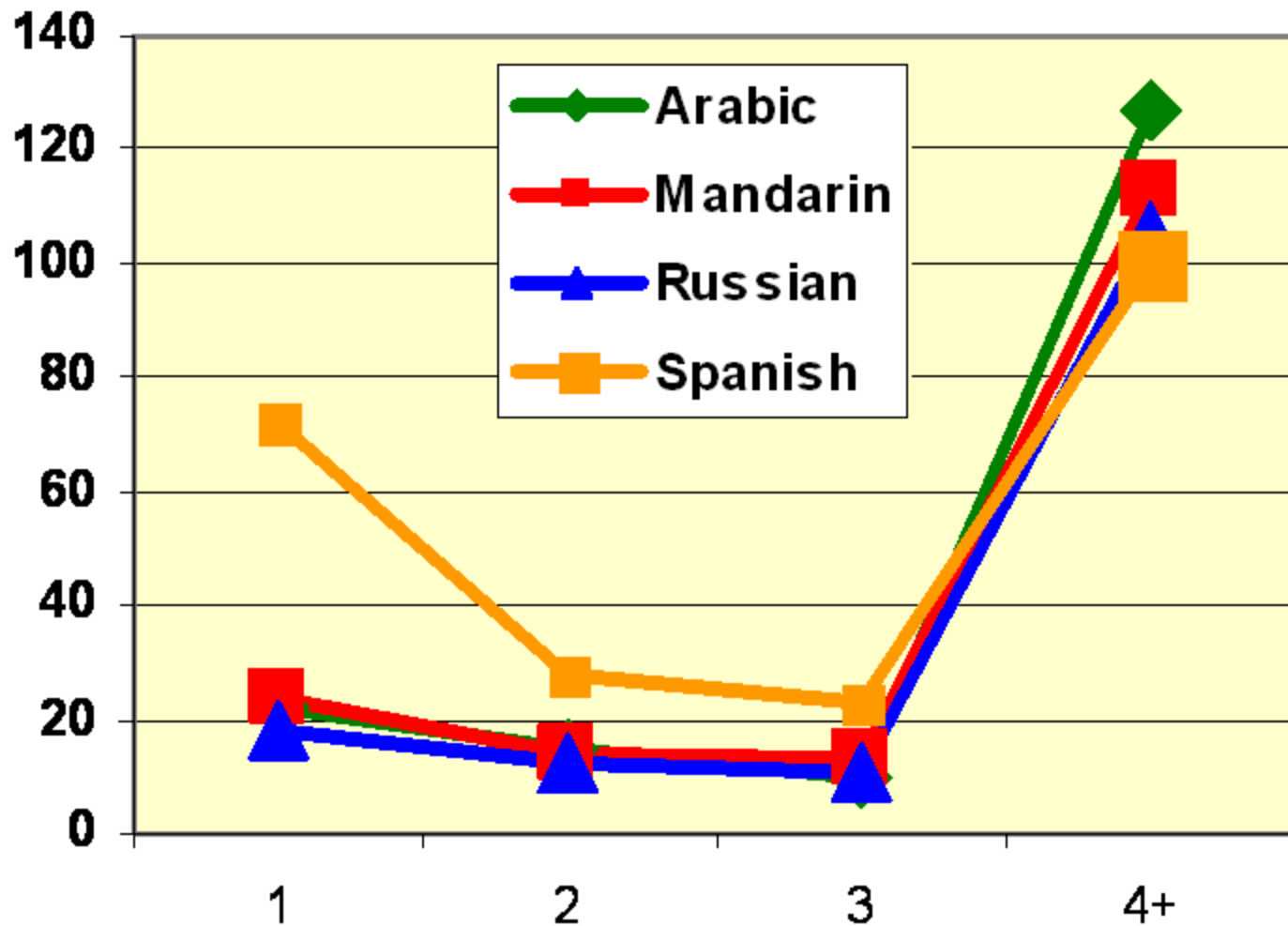
Calls by Language



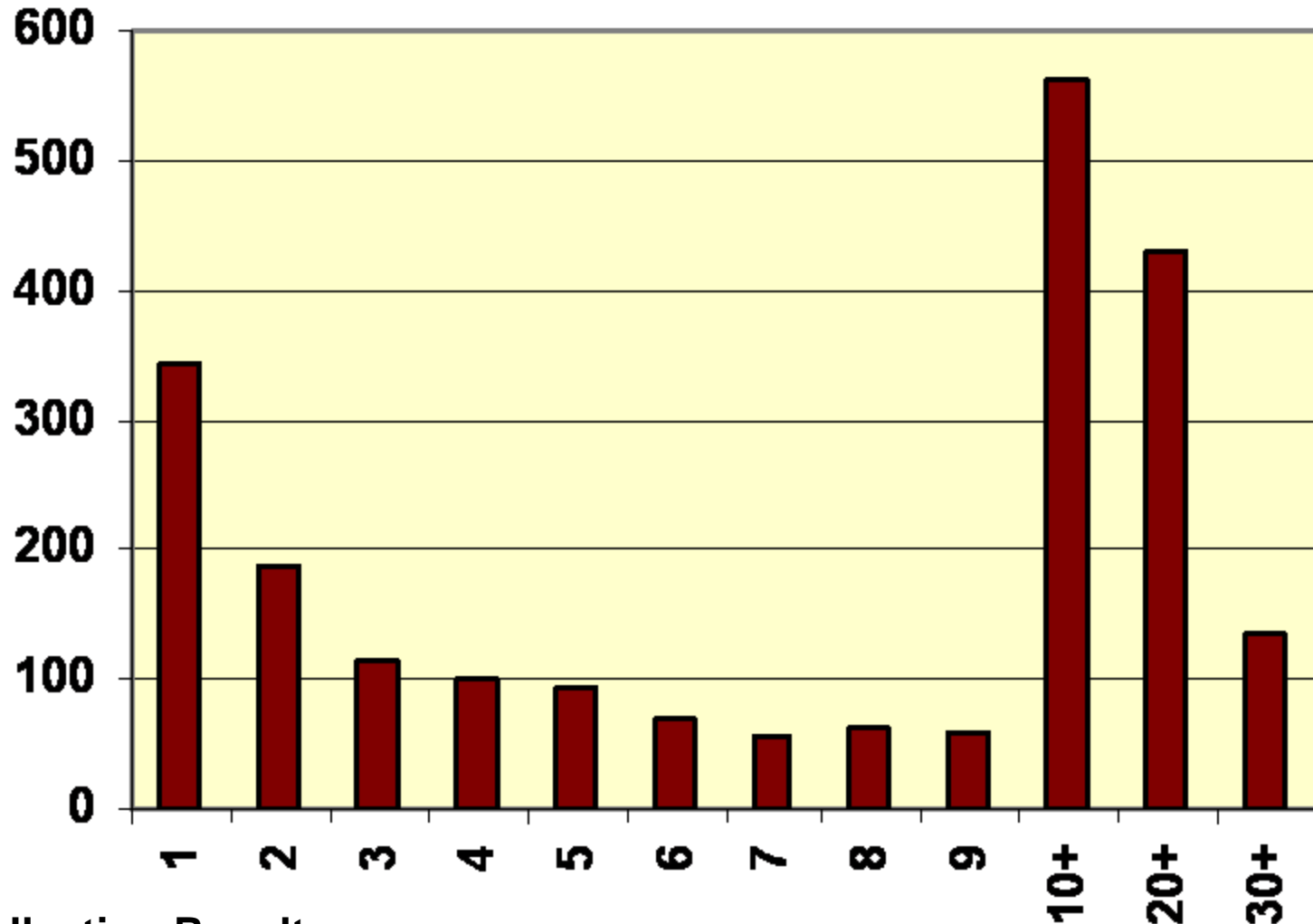
- In 98.8% of calls, subjects chose language as requested; speaking in a shared non-English language where possible and otherwise defaulting to English.

- Arabic: 738
- Mandarin: 520
- Russian: 534
- Spanish: 372
- English: 12207

Subjects by # non-English calls



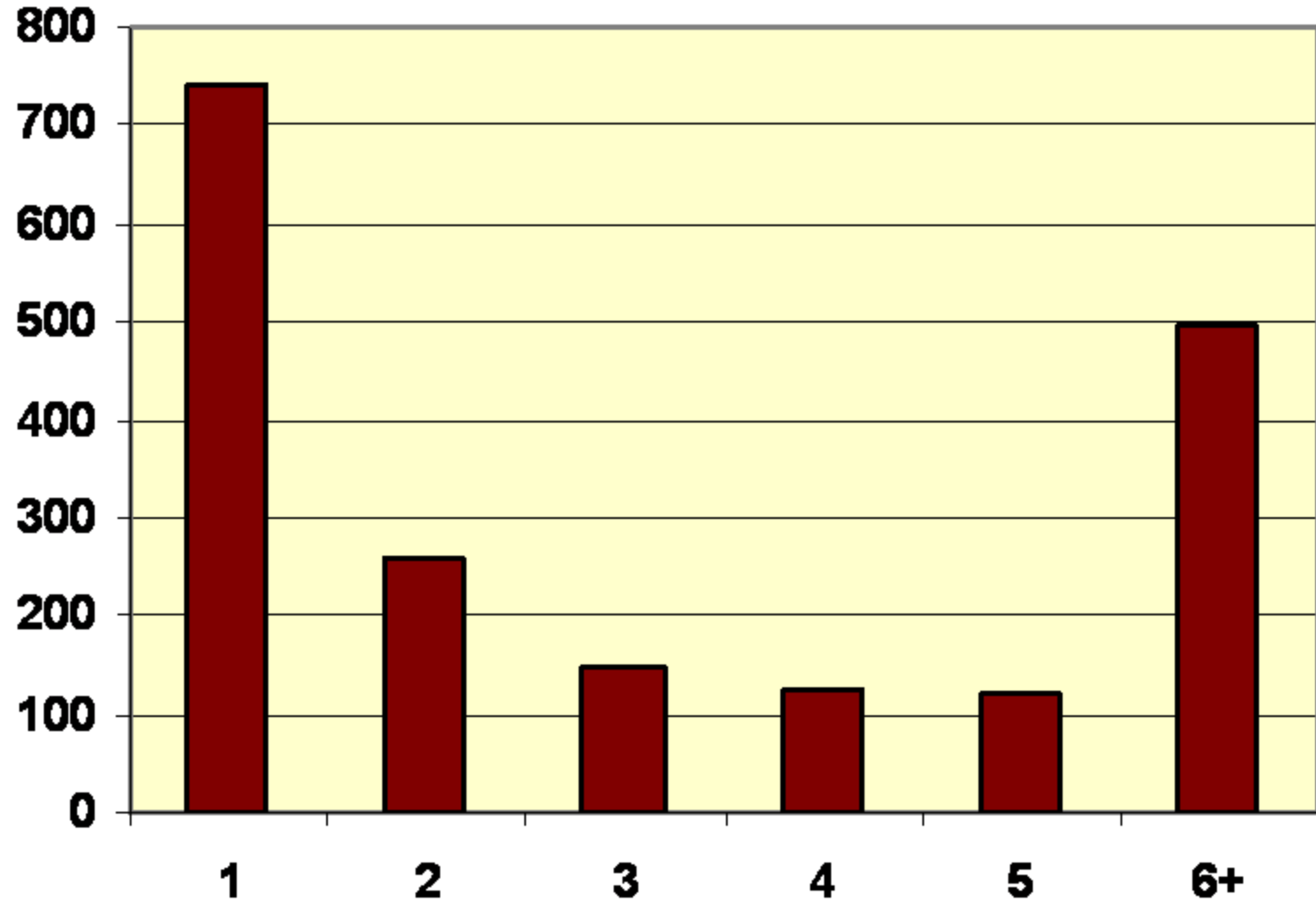
Callers by Calls Made



- **Collection Results**

- 611 subjects completed 20 calls
- ~1150 subjects complete 10 calls
- 2 peaks – 1 call, and 30 calls (thanks to bonus)

Speakers by # Unique Handsets



Future Work

- **Mixer Phases I, II reported here**
- **Phase III**
 - complete, 400 subjects completed 12+ calls
 - Collection coordinate with Language ID community
 - » 22 linguistic varieties represented
 - used in NIST's 2006 Speaker Recognition Evaluation
- **Future Work**
 - Phase IV under discussion; 330 subjects at or near completion
 - interest in more multi-channel, broadband collection
 - Interest in new collection scenarios
 - » not just new topics but
 - » Interviews, other interactive styles
 - » greater within-speaker variation
- **All data to be published; current plan to begin publications in 2006**