

The Mixer Corpus of Multilingual, Multichannel Speaker Recognition Data

Christopher Cieri¹, Joseph P. Campbell², Hirotaka Nakasone³, David Miller¹, Kevin Walker¹

¹University of Pennsylvania, Linguistic Data Consortium, Philadelphia, PA, USA

²MIT Lincoln Laboratory, Lexington, MA, USA

³Federal Bureau of Investigation, Quantico, VA, USA

ccieri@ldc.upenn.edu, j.campbell@ieee.org, hnakasone@fbiacademy.edu, damiller@ldc.upenn.edu walker@ldc.upenn.edu

Forensic Automatic Speaker Recognition

• Minimum Requirements

- text-independent
- channel-independent

• New requirements

- capable of handling multiple languages including bilingual speakers

• Plan

- multilingual, multi-channel collection
- dissemination to research sites
- system performance improvement
- system performance evaluation

Corpus Design

		Calls			
	Speakers	4	10	20	
Phase I	100	Arabic	Unique Handset		
	100	Mandarin			
	100	Russian			
	100	Spanish			
	100	XC & TR			
	100	Extended Data			
Phase II	100	Xchannel			
	100	Extended Data			
	100	Extended Data			
	100	Extended Data			
	100	Extended Data			
	50	Extended Data			

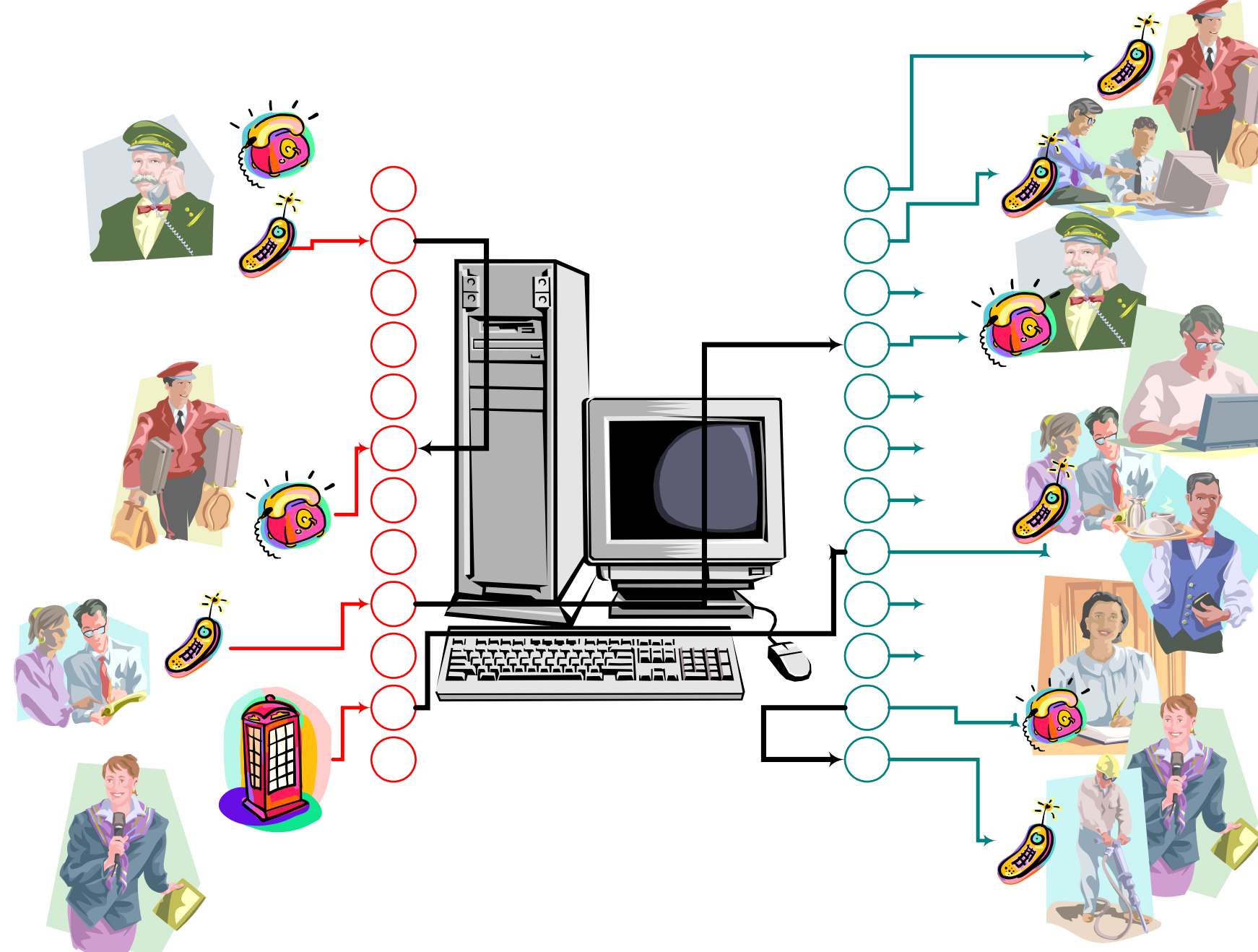
• Generalities

- calls are 6 minutes in duration
- subjects speak to each other
- assigned topics change daily
- robot operator logs ANI
- unique handsets encouraged
- subjects indicate phone and handset type

• Components

- Core: 600 subjects*10 calls
- Extended: 100 subjects continue to 20 calls
- Multilingual
 - 100 subjects * 4+ of their calls in Arabic
 - 100 subjects * 4+ of their calls in Mandarin
 - 100 subjects * 4+ of their calls in Russian
 - 100 subjects * 4+ of their calls in Spanish
- Multi-channel Data: 100 subjects * 4 of their calls via multi-channel device
- Transcript Reading: 100 subjects read extracts of transcripts of each others' and their conversations via multi-channel device

Fisher Platform Behavior



Multichannel System

- Laptop with two firewire hard drives
- Multichannel interface
- Sensors



Wireline telephone

call goes to robot operator platform, not to MOTU)



1. Cell-phone style ear-bud in-line lapel mic
Motorola Earbud Handsfree, SYN8390, \$12



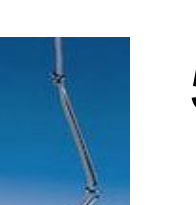
2. Over the ear miniboom mic
Jabra® EarWrap Headset Radio Shack #43-1914 \$30



3. Courtroom mic
Shure MX418S Supercardioid Gooseneck Mic \$185, mounts to furniture



4. Conference room mic (table top boundary mic)
Crown SoundGrabber II pressure-zone mic \$70



5. Distant mic (e.g., courtroom mic across the room)
Audio Technica Pro 45 Cardioid Condenser Hanging Mic \$91



6. Studio mic (placed near talker)
Audio Technica AT3035 Cardioid Condenser \$200 (not including stand)



7. PC-style stand mic
Radio Shack Desktop Mic with Noise Canceling #33-3031 \$27



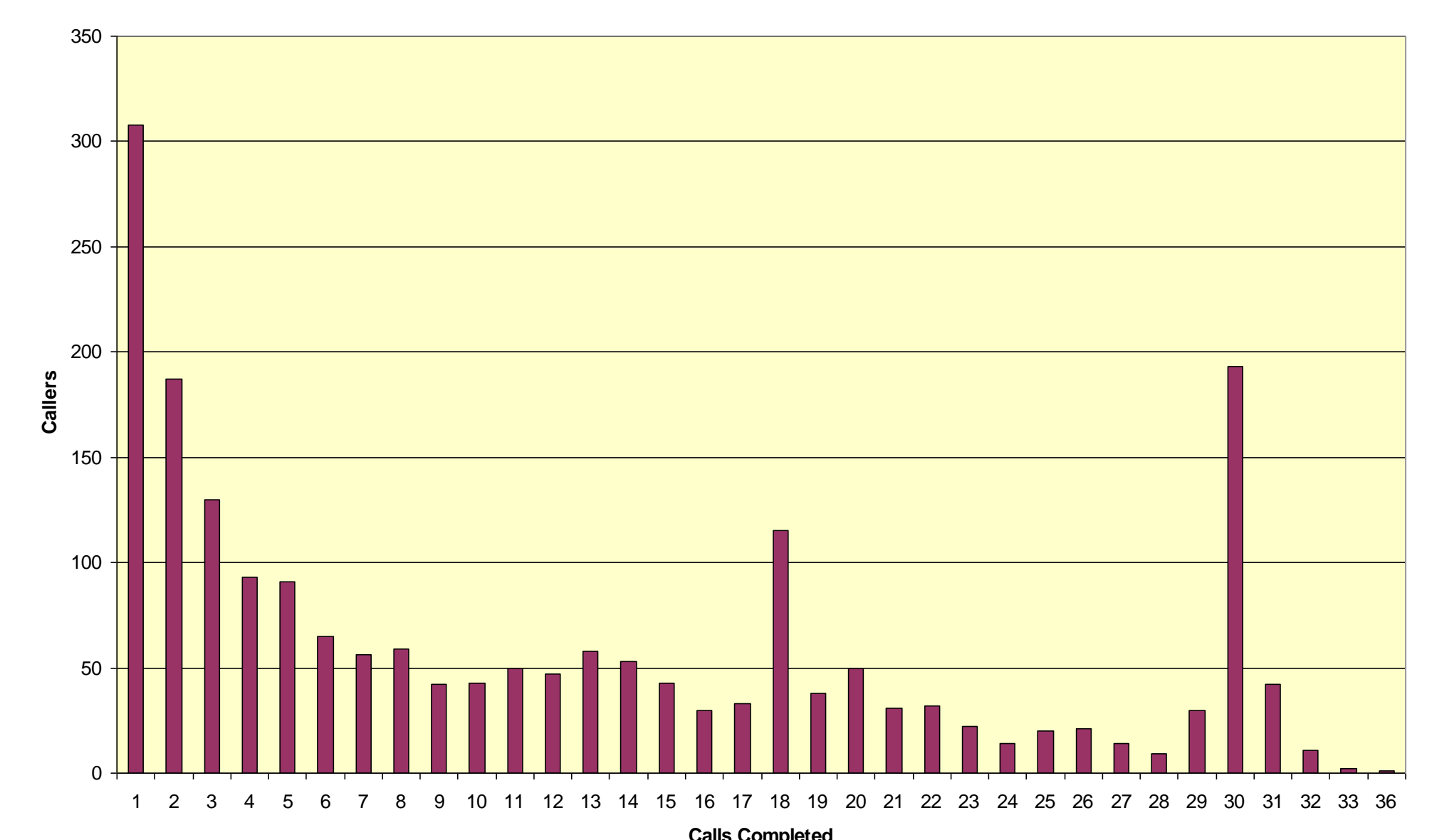
8. Microcassette mic
Olympus PearlCorder S725 (no mic monitoring capability) \$66

- Each channel sampled 48kHz, 16bit
- Calls also collected by robot operator

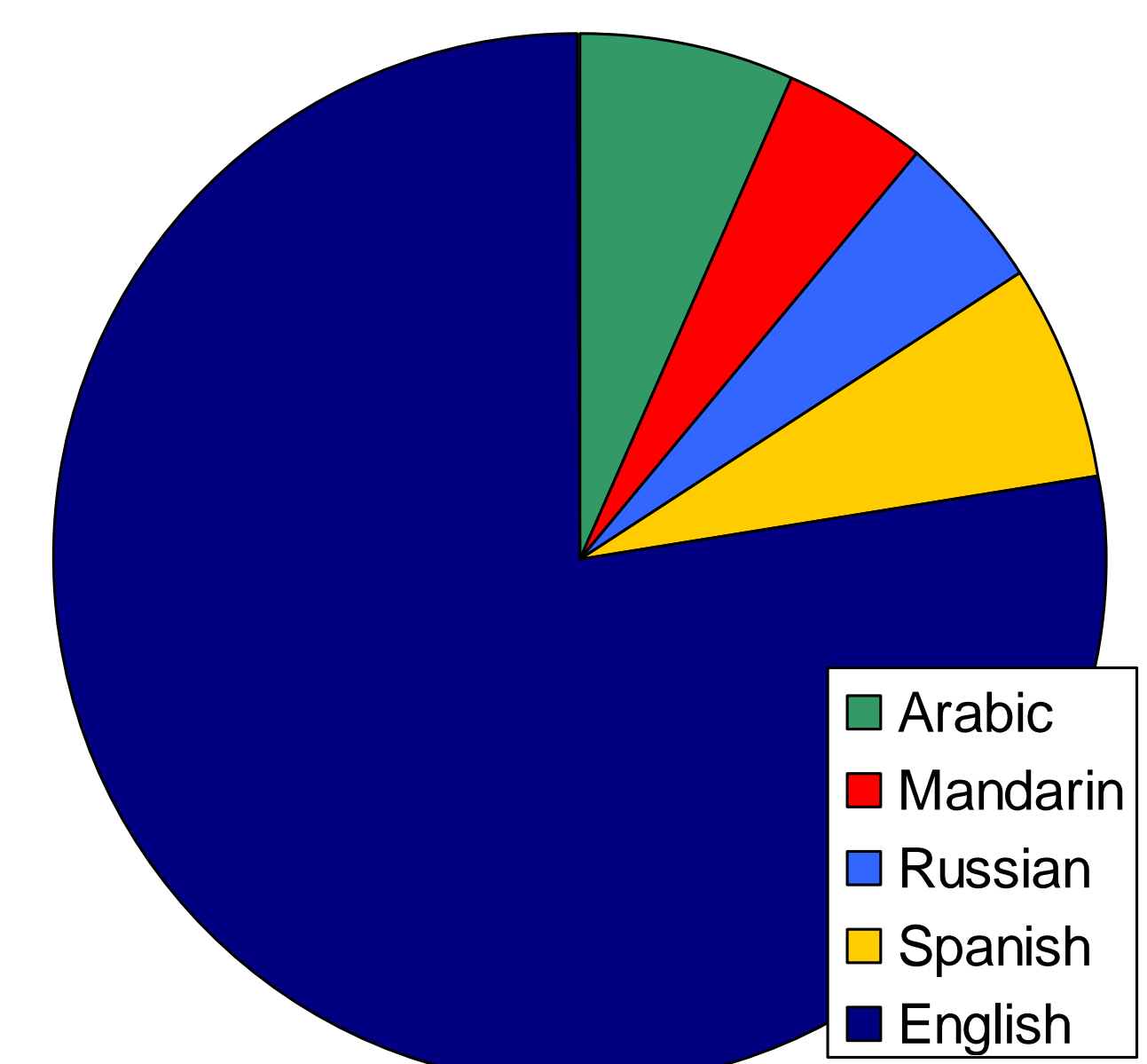
Outcomes To Date

- 4651 subjects recruited
- 12,169 calls (~1200 hours) collected
- >250 new calls each week
- >100 cross channel subjects completed 4 calls
- collection continues

Callers by Calls Made



Calls by Language



Research

• Supports speaker recognition research with emphasis on forensic-style problems:

- telephone conversations
- channel independence
- language independence & bilingualism
- transcript reading

• Mixer is the first large scale, publicly available corpus to address all these dimensions.

• FBI vision is to create a corpus that accurately reflects and focuses research on forensic-style problems.

• Research at MIT-LL aims to produce robust automatic speaker recognition system to support forensic analysis experts.

• Mixer used in 2004 NIST Speaker Recognition Evaluation