Speaker Identification (SID) in Low-Rate Coded Speech

Andrew Catellier and Stephen Voran
{acatellier,svoran}@its.bldrdoc.gov
Institute for Telecommunication Sciences
Telecommunications Theory Division

Effect of Transmission on Multimedia Quality of Service
17-19 June 2008 – Prague, Czech Republic

SID: Motivation

- New equipment for first responders
- Anecdotal complaints about system performance
  - Speaker emotional state, and speaker identity obscured
- Legitimate concerns!

SID: Motivation

- To help improve the platform, these problems must be measured
- Two experiments were designed to measure the problem
- We'll be talking about SID today

SID: Experiment Design

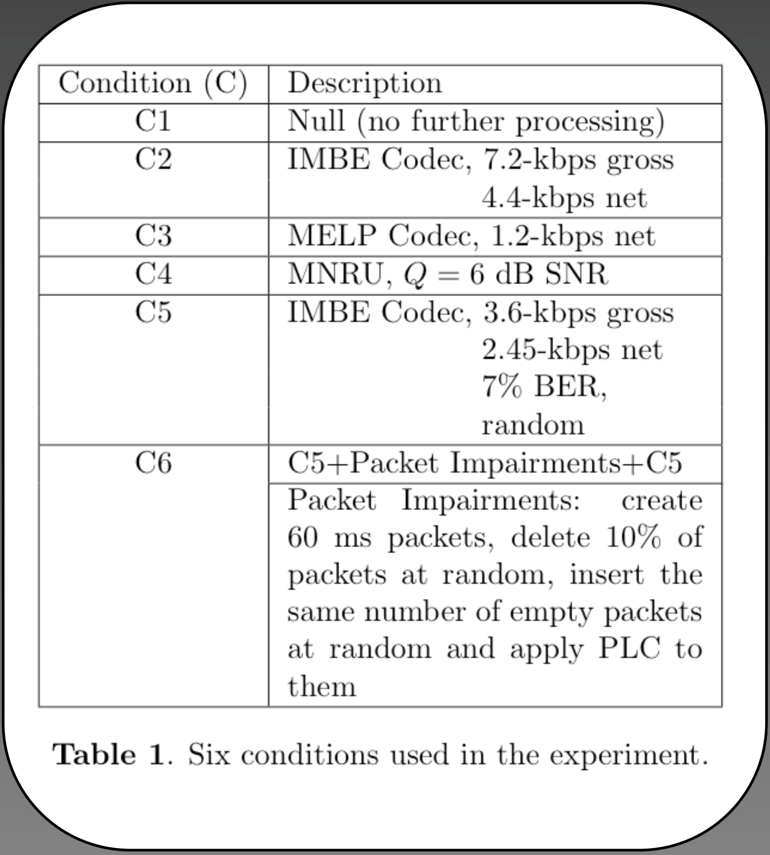
- Specifications:
  - Unfamiliar talkers
  - Clips with and without prosodic information (short and long clips)
  - Six simulated communication systems
  - Manageable experiment length

SID: Experiment Design

- Realization:
  - – Tactical Speaker Identification Database
    - Used three males and three females
  - – Three clip lengths: sentence, four digits, two digits
  - – MELP, IMBE 7.2, 3.6 kbps (with and without impairments), MNRU
  - – 360 total clips, experiment length around one hour

SID: Experiment Design

- C1 clips produced by resampling at 8 kHz, filtering (160-3640 Hz bandpass), and then normalized to -26 dB below clipping
- Low-rate vocoders in C2, C3, C5 and C6 are similar to those used in Public Safety communication systems
- C5 and C6 have additional transmission impairments



Condition (C)	Description
C1	Null (no further processing)
C2	IMBE Codec, 7.2-kbps gross 4.4-kbps net
C3	MELP Codec, 1.2-kbps net
C4	MNRU, $Q = 6$ dB SNR
C5	IMBE Codec, 3.6-kbps gross 2.45-kbps net 7% BER, random
C6	C5+Packet Impairments+C5
	Packet Impairments: create 60 ms packets, delete 10% of packets at random, insert the same number of empty packets at random and apply PLC to them

Table 1. Six conditions used in the experiment.

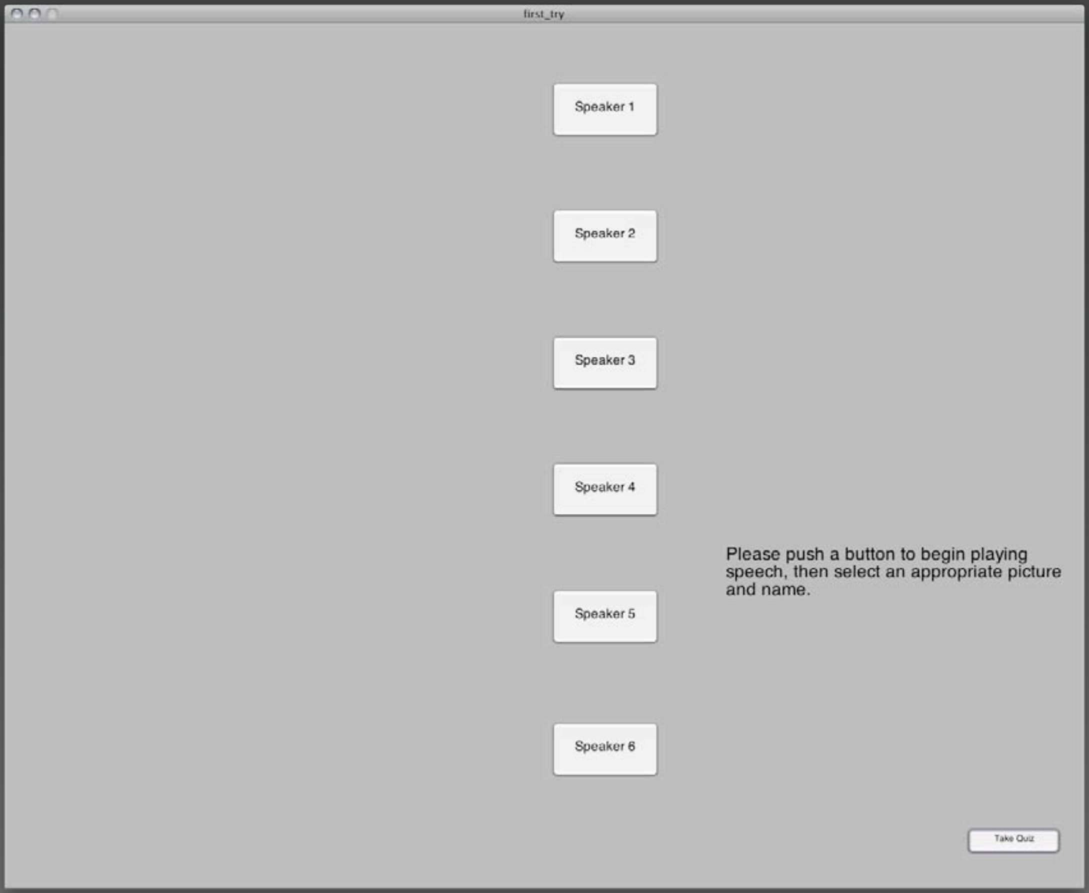
SID: Experiment Design

- Problems:
  - Training listeners to accurately recognize any given speaker
  - False confidence in training
  - Mid-test mistraining

SID: Experiment Design

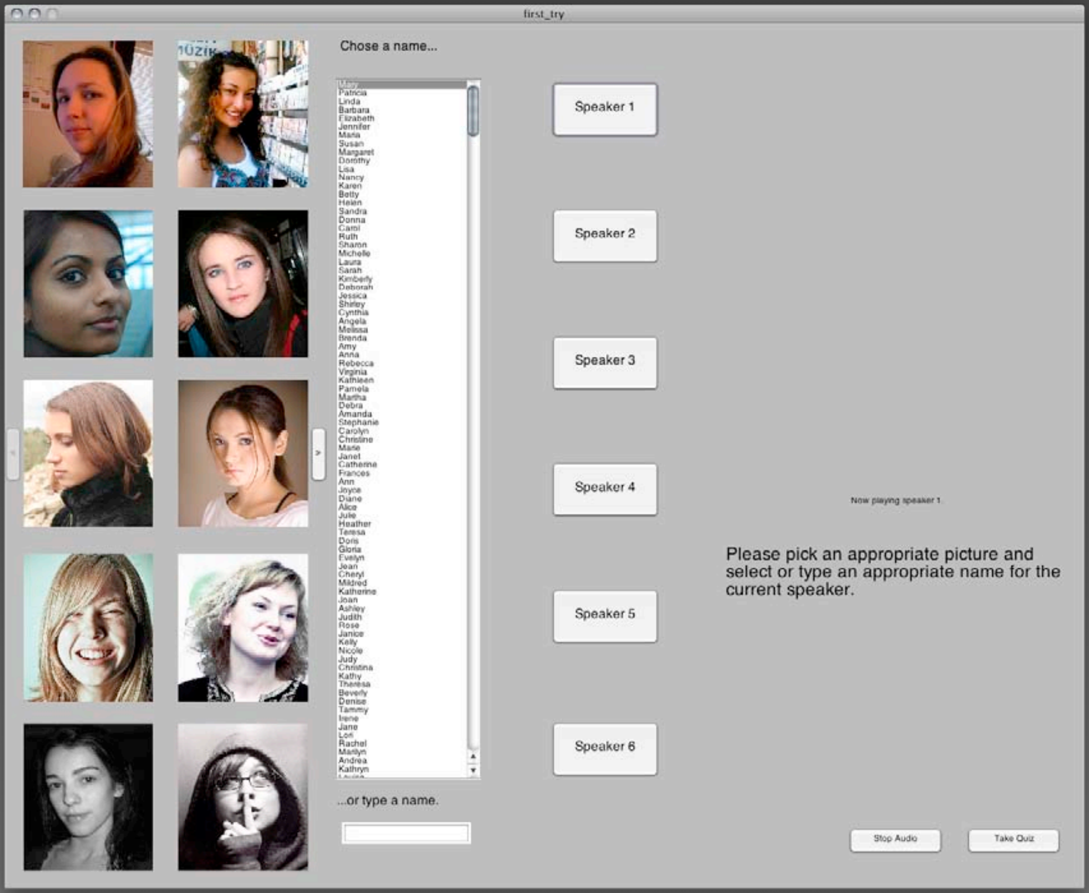
- Training:
  - Used set of clips where speakers were giving directions (semi-spontaneous)
  - Allowed listeners to assign appropriate memory aids: a name and a face

SID: Experiment Design



Please push a button to begin playing speech, then select an appropriate picture and name.

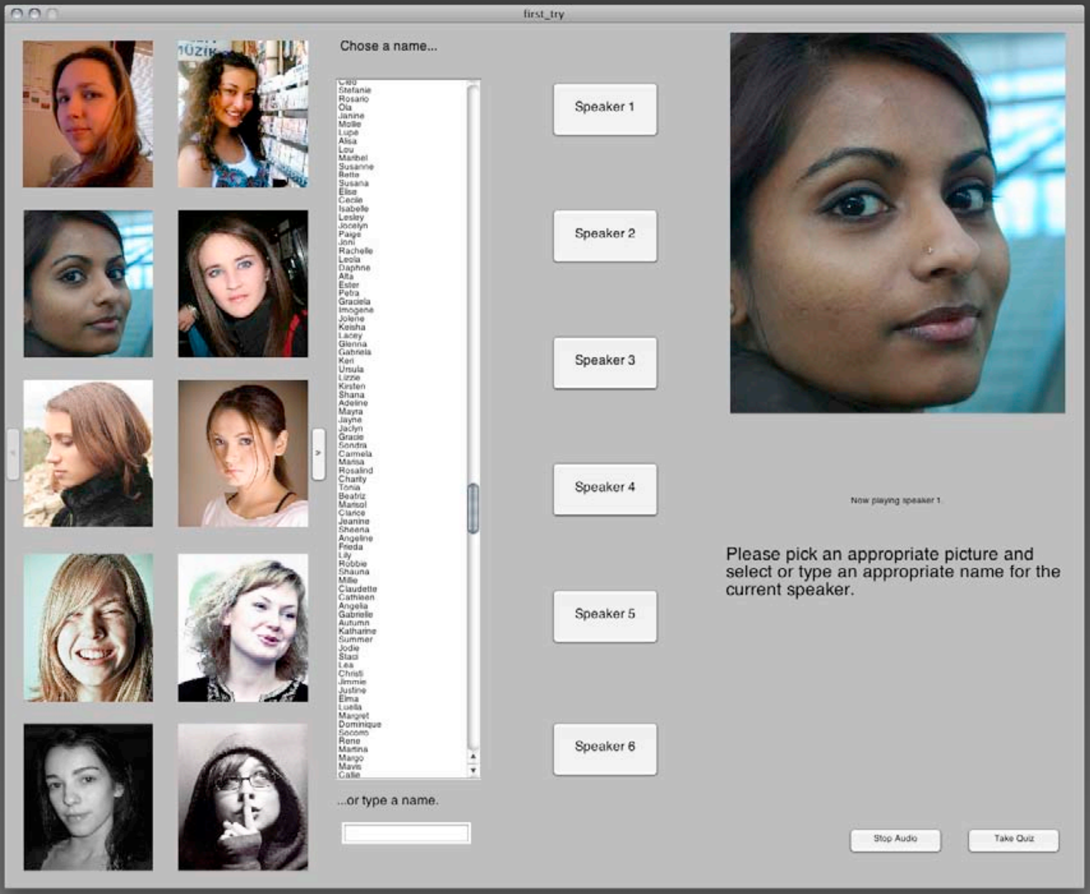
SID: Experiment Design



Now playing speaker 1.

Please pick an appropriate picture and select or type an appropriate name for the current speaker.

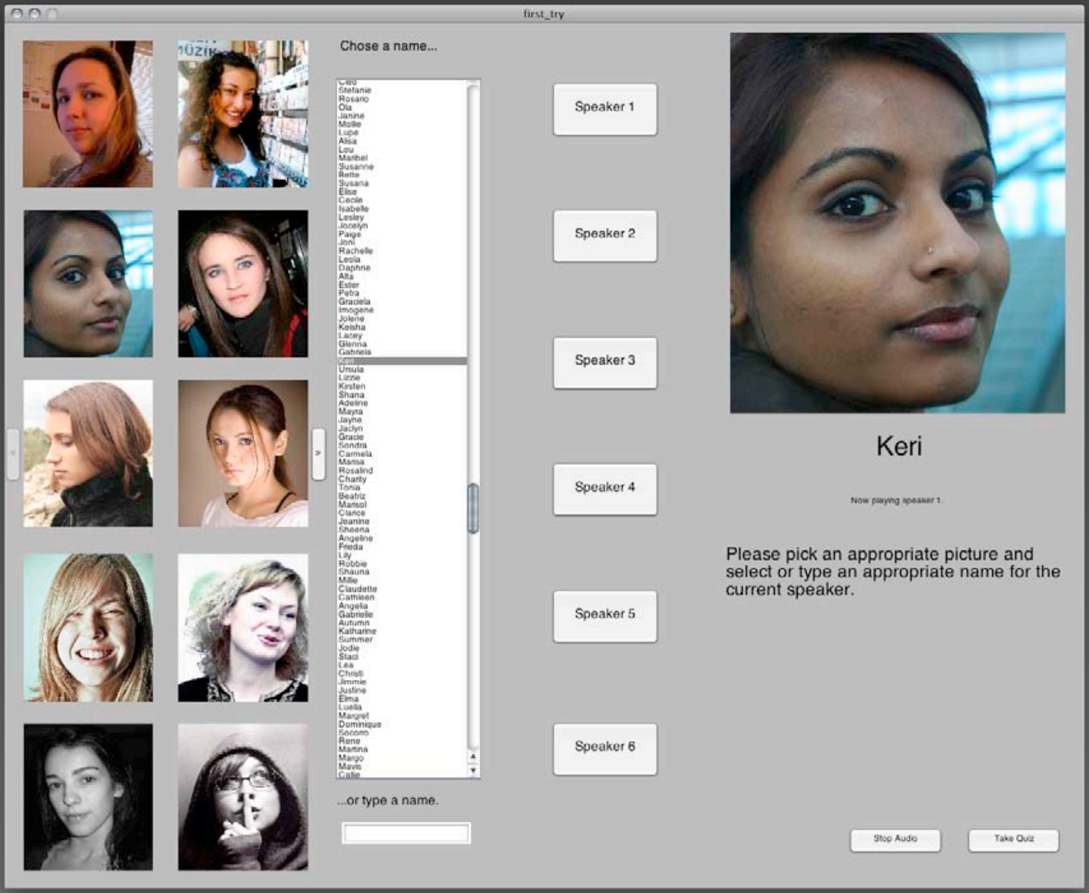
SID: Experiment Design



Now playing speaker 1.

Please pick an appropriate picture and select or type an appropriate name for the current speaker.

SID: Experiment Design

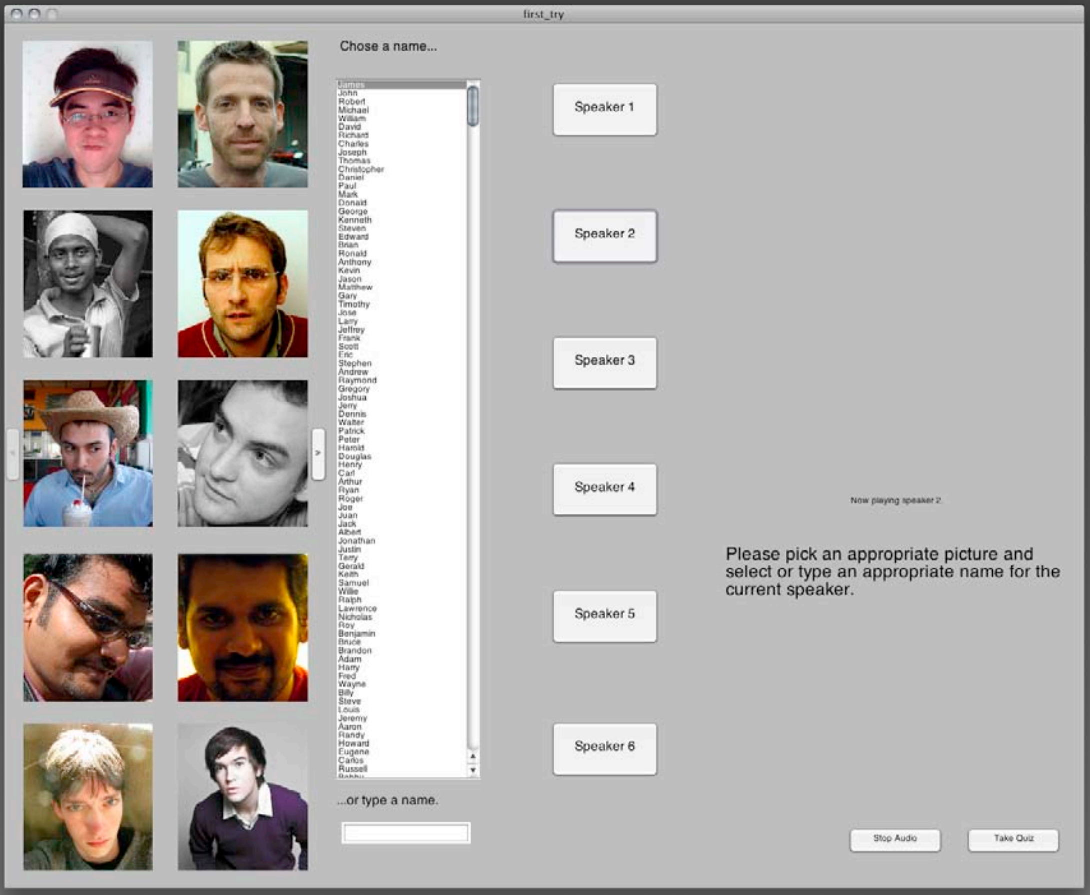


Keri

Now playing speaker 1.

Please pick an appropriate picture and select or type an appropriate name for the current speaker.

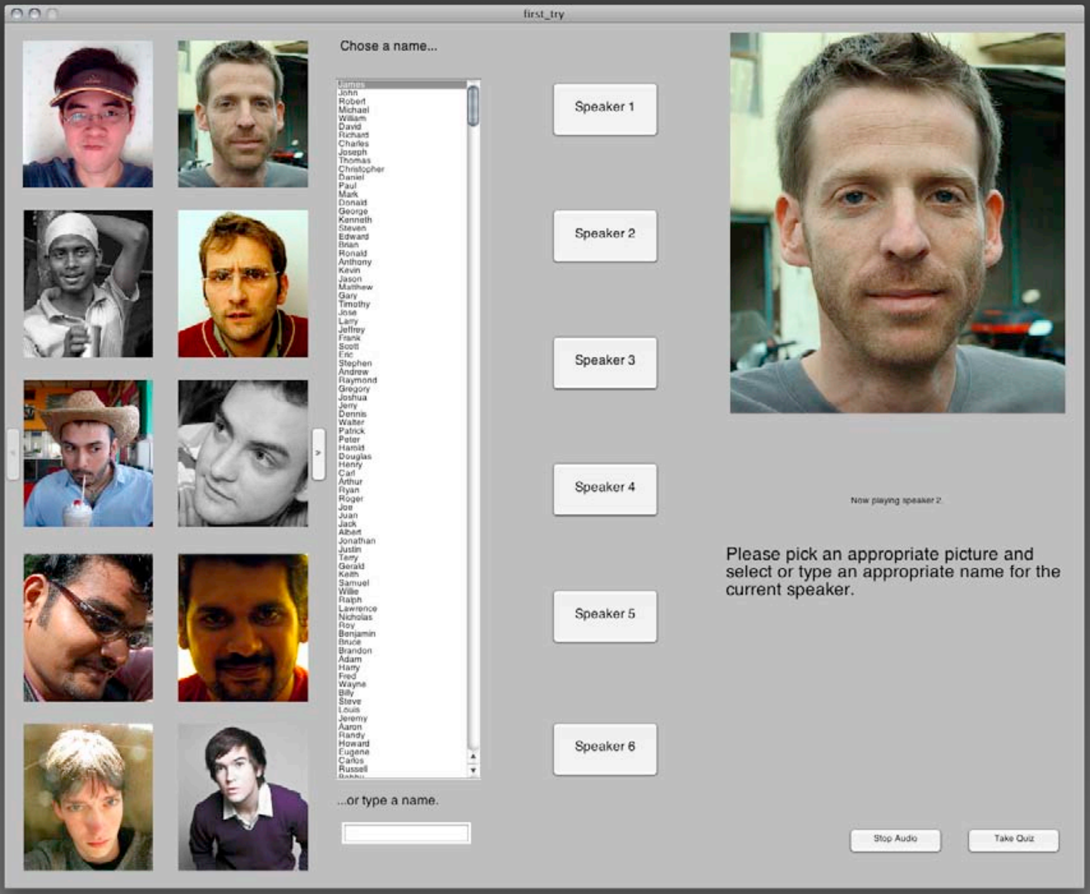
SID: Experiment Design



Now playing speaker 2

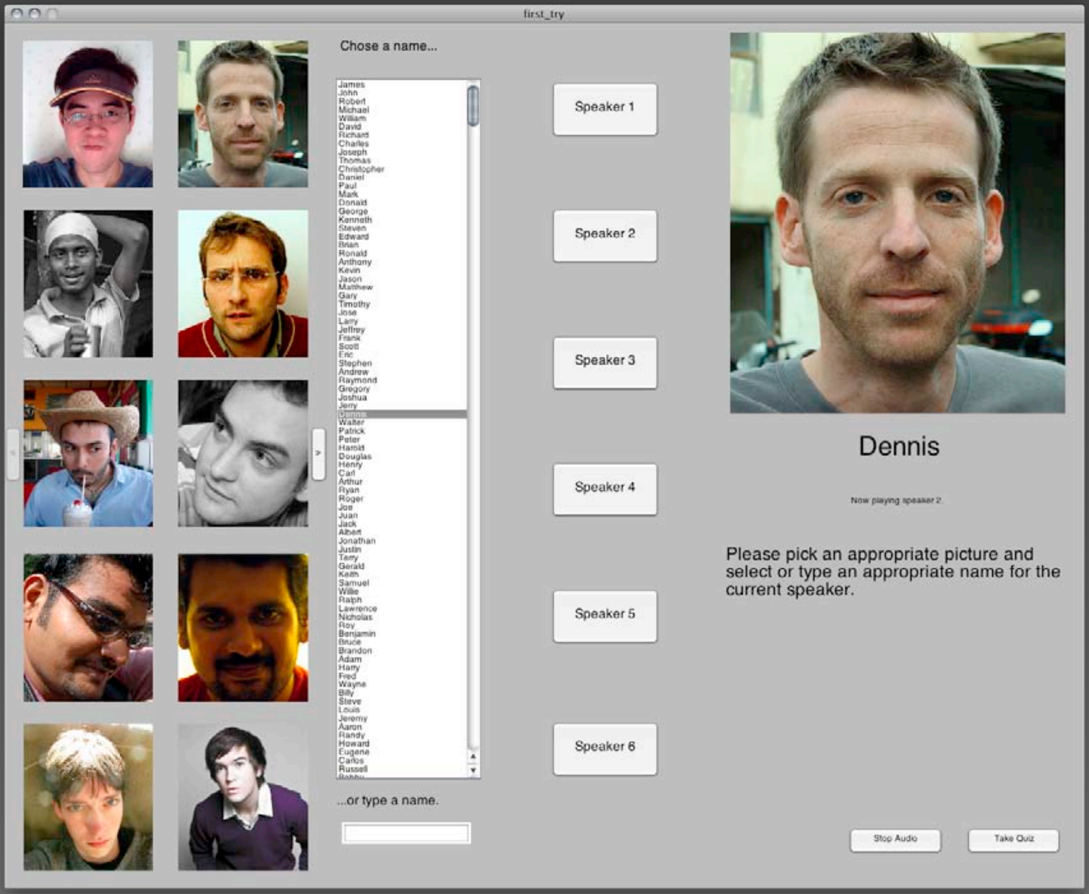
Please pick an appropriate picture and select or type an appropriate name for the current speaker.

SID: Experiment Design



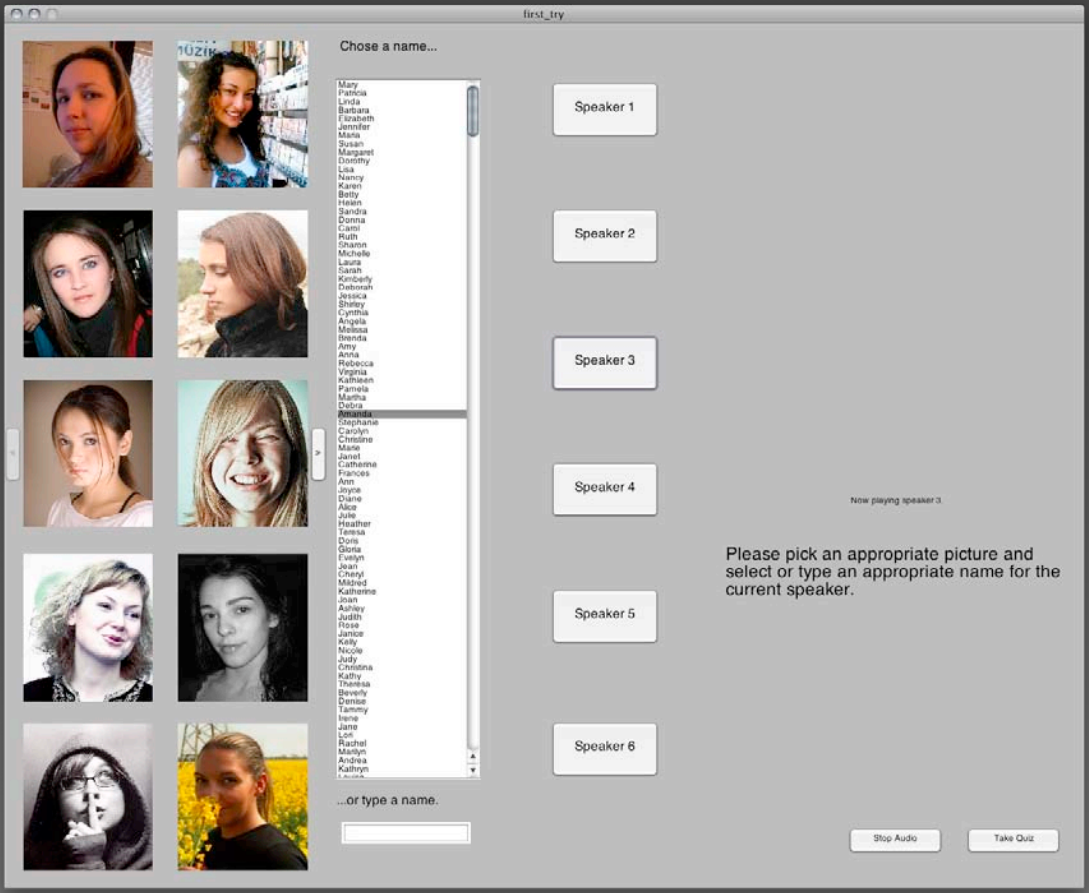
Please pick an appropriate picture and select or type an appropriate name for the current speaker.

SID: Experiment Design



Please pick an appropriate picture and select or type an appropriate name for the current speaker.

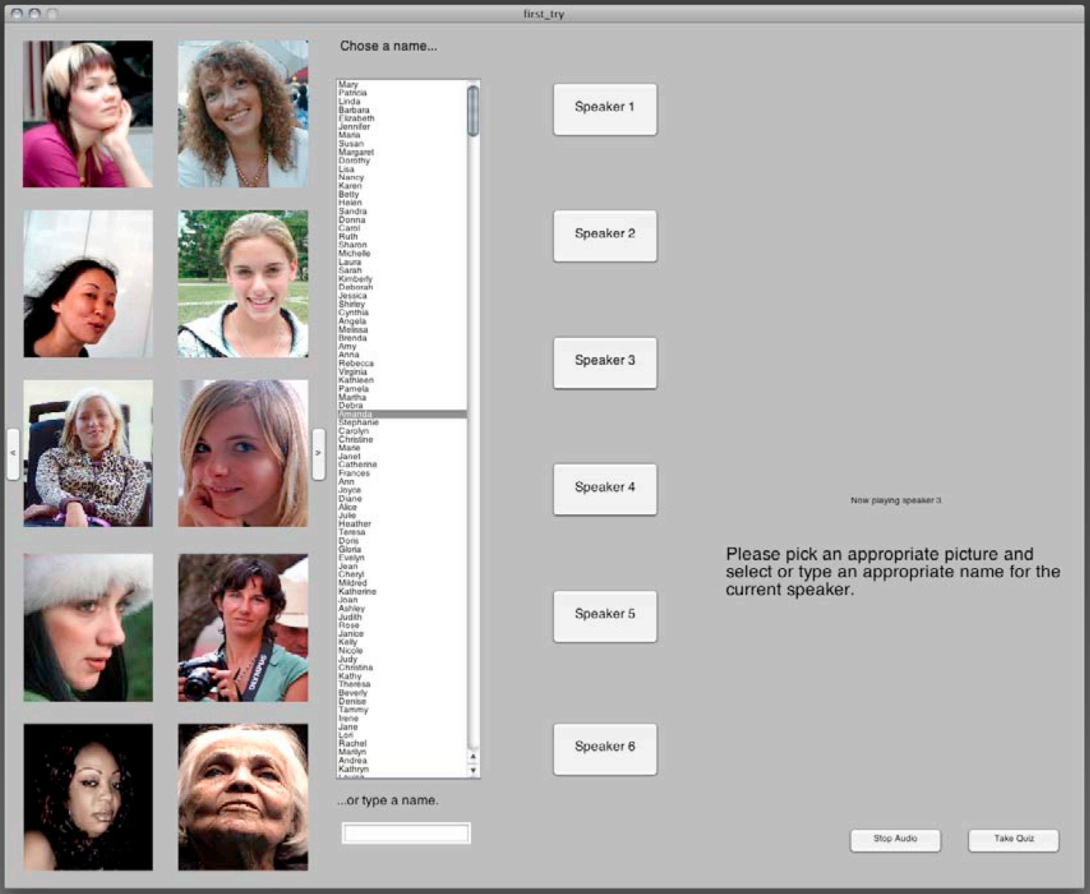
SID: Experiment Design



Now playing speaker 3

Please pick an appropriate picture and select or type an appropriate name for the current speaker.

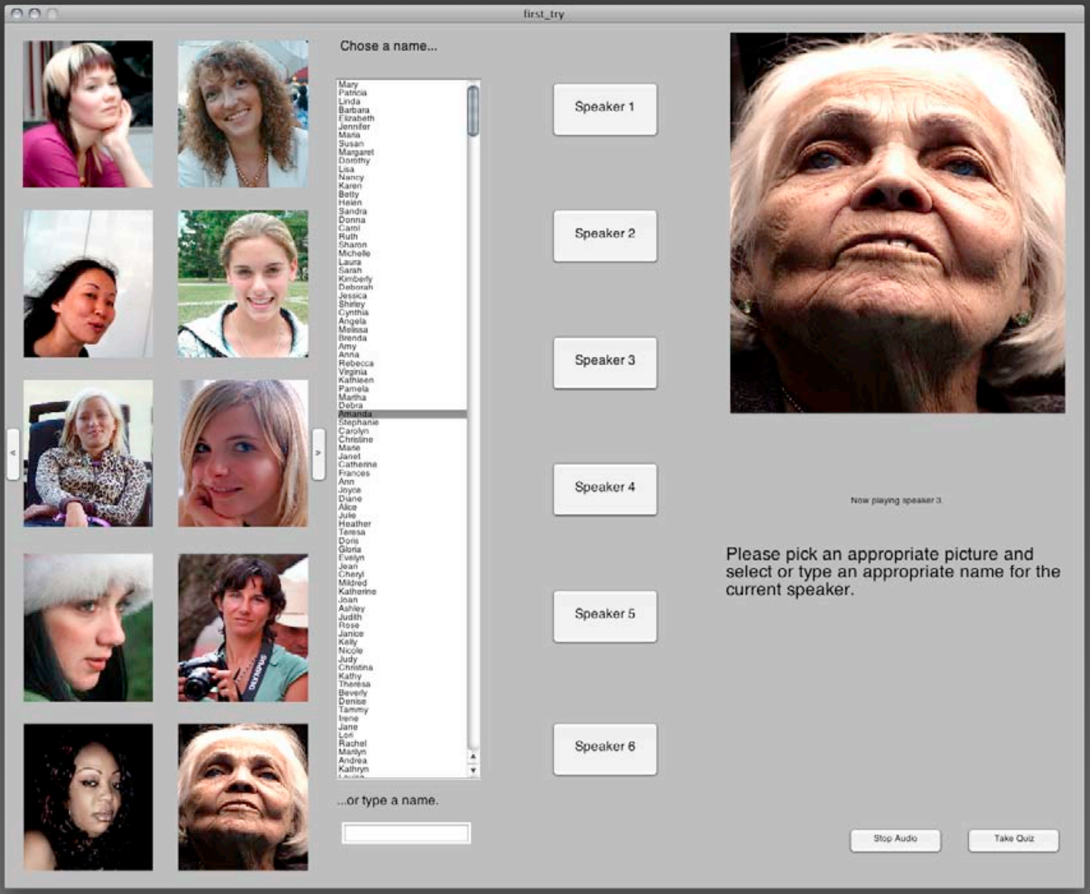
SID: Experiment Design



Now playing speaker 3

Please pick an appropriate picture and select or type an appropriate name for the current speaker.

SID: Experiment Design



Chose a name...

Mary
Patricia
Linda
Barbara
Elizabeth
Jennifer
Maria
Susan
Margaret
Dorothy
Lisa
Nancy
Karen
Betty
Helen
Sandra
Donna
Carol
Ruth
Sharon
Michelle
Laura
Sarah
Kimberly
Deborah
Jessica
Shirley
Cynthia
Angela
Melissa
Brenda
Amy
Anna
Rebecca
Virginia
Kathleen
Pamela
Martha
Debra
Amanda
Stephanie
Carolyn
Christine
Marie
Janet
Catherine
Frances
Ann
Joyce
Diane
Alice
Julie
Heather
Teresa
Doris
Gloria
Evelyn
Jean
Cheryl
Mildred
Katherine
Joan
Ashley
Judith
Rose
Janice
Kelly
Nicole
Judy
Christina
Kathy
Theresa
Beverly
Denise
Tammy
Irene
Jane
Lori
Rachel
Marilyn
Andrea
Kathryn

...or type a name.

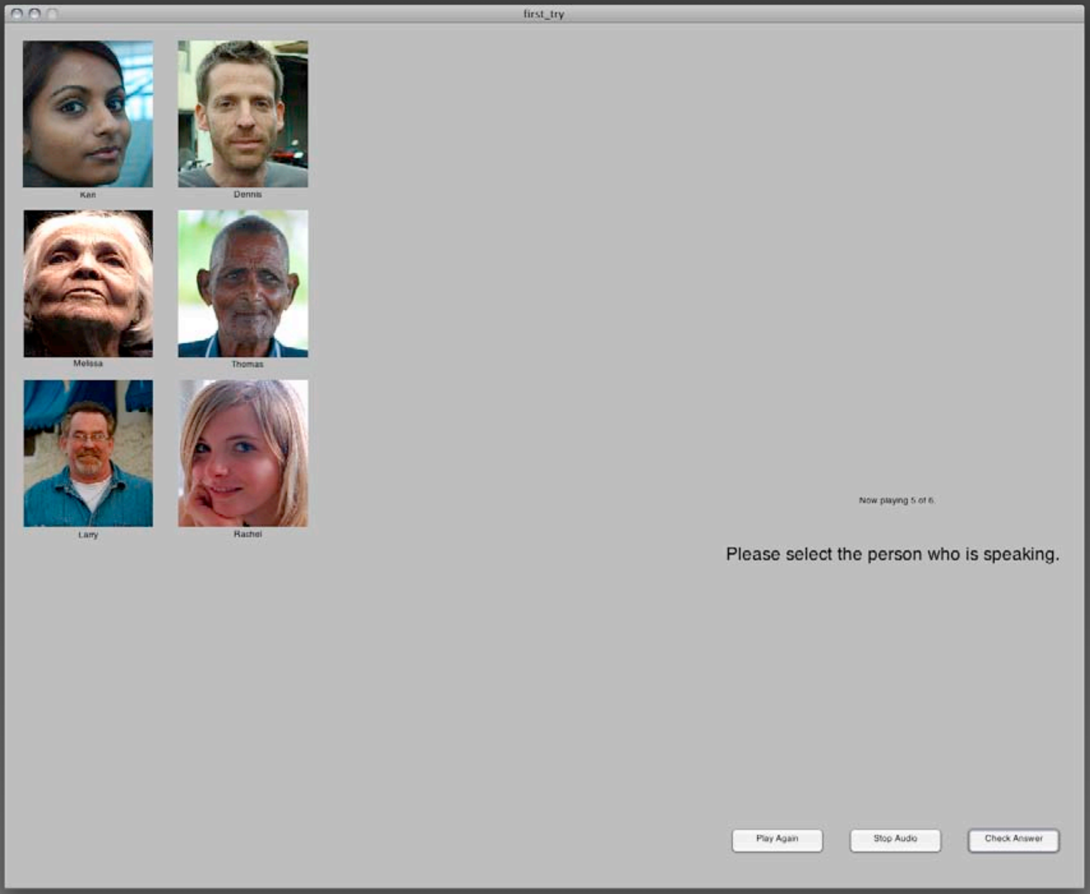
Now playing speaker 3

Please pick an appropriate picture and select or type an appropriate name for the current speaker.

SID: Experiment Design

- Quiz:
  - Undistorted speech
  - Provided feedback to listener about state of training
  - Familiarized listener with test process

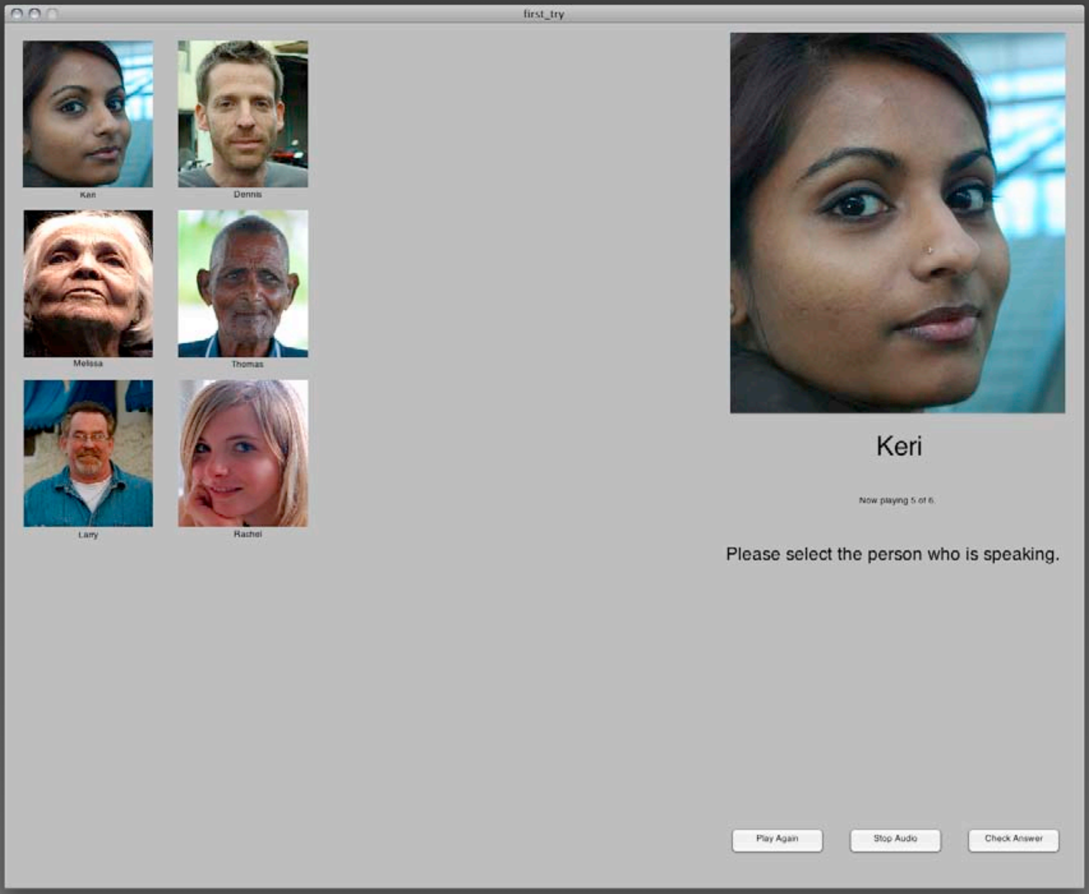
SID: Experiment Design



Now playing 5 of 6.

Please select the person who is speaking.

SID: Experiment Design

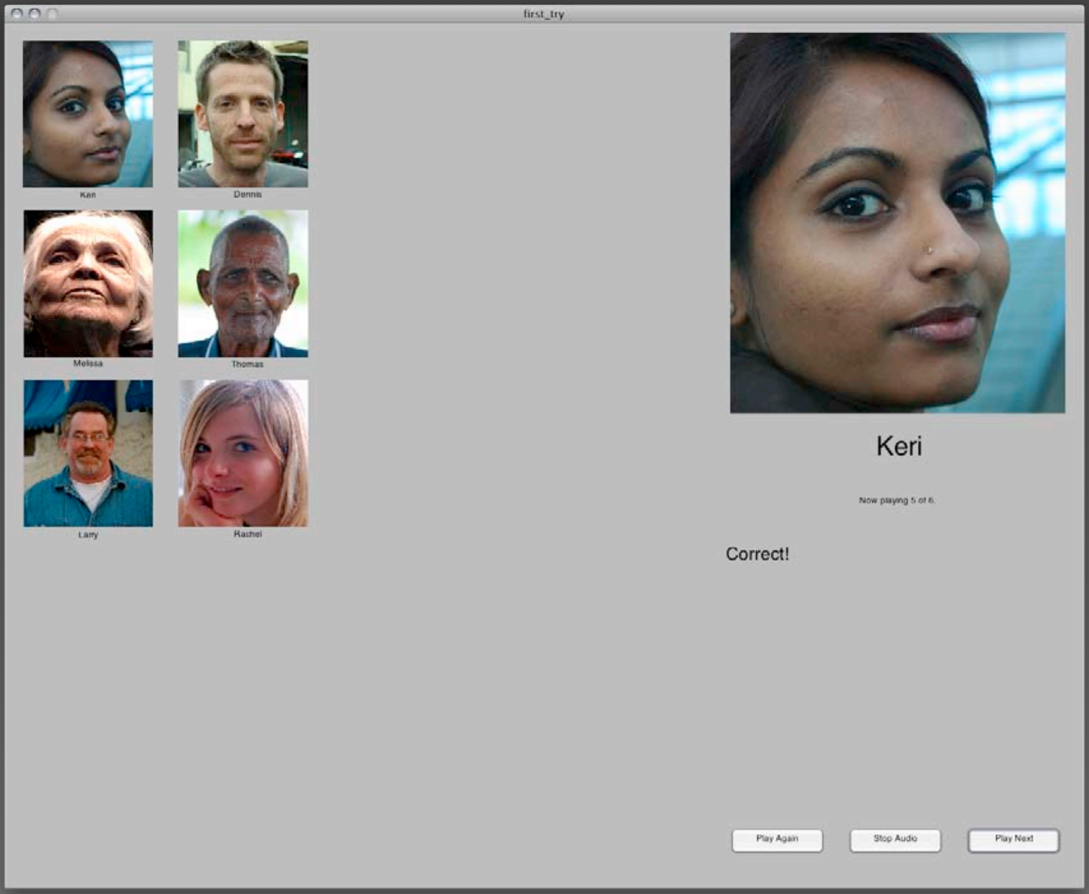


Keri

Now playing 5 of 6.

Please select the person who is speaking.

SID: Experiment Design

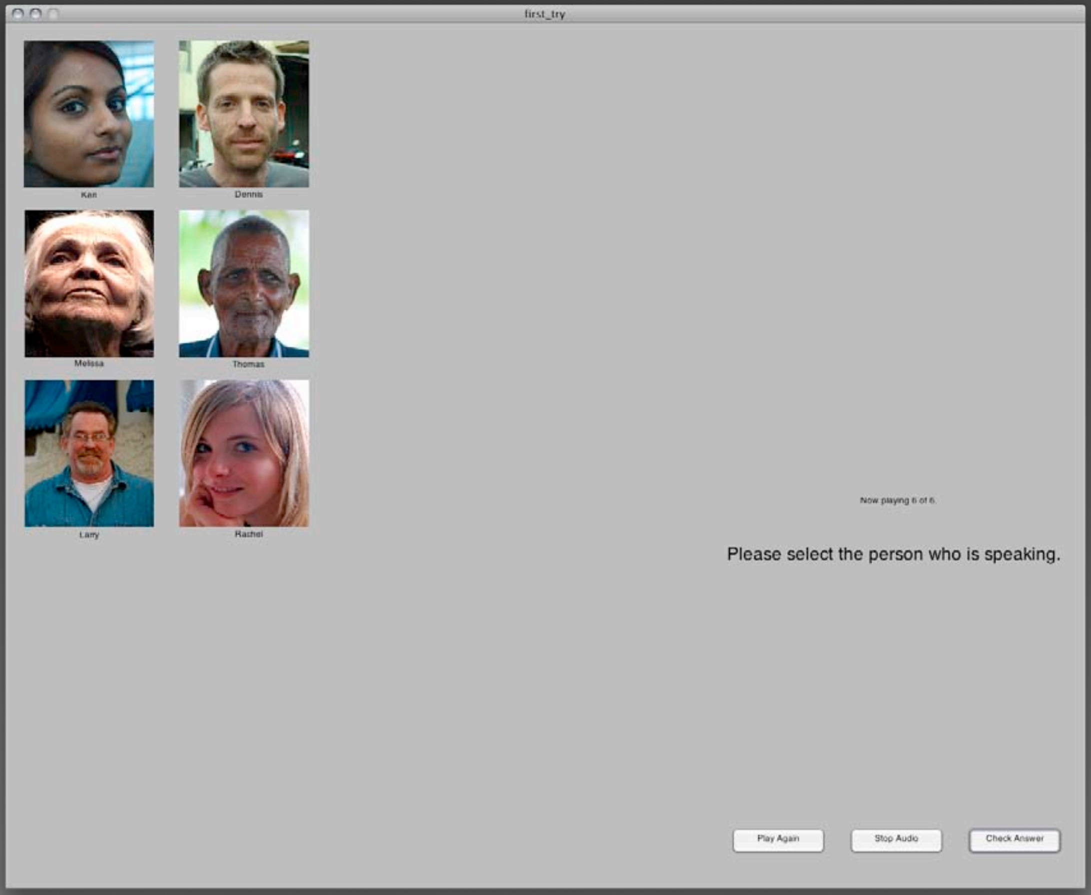


Keri

Now playing 5 of 6.

Correct!

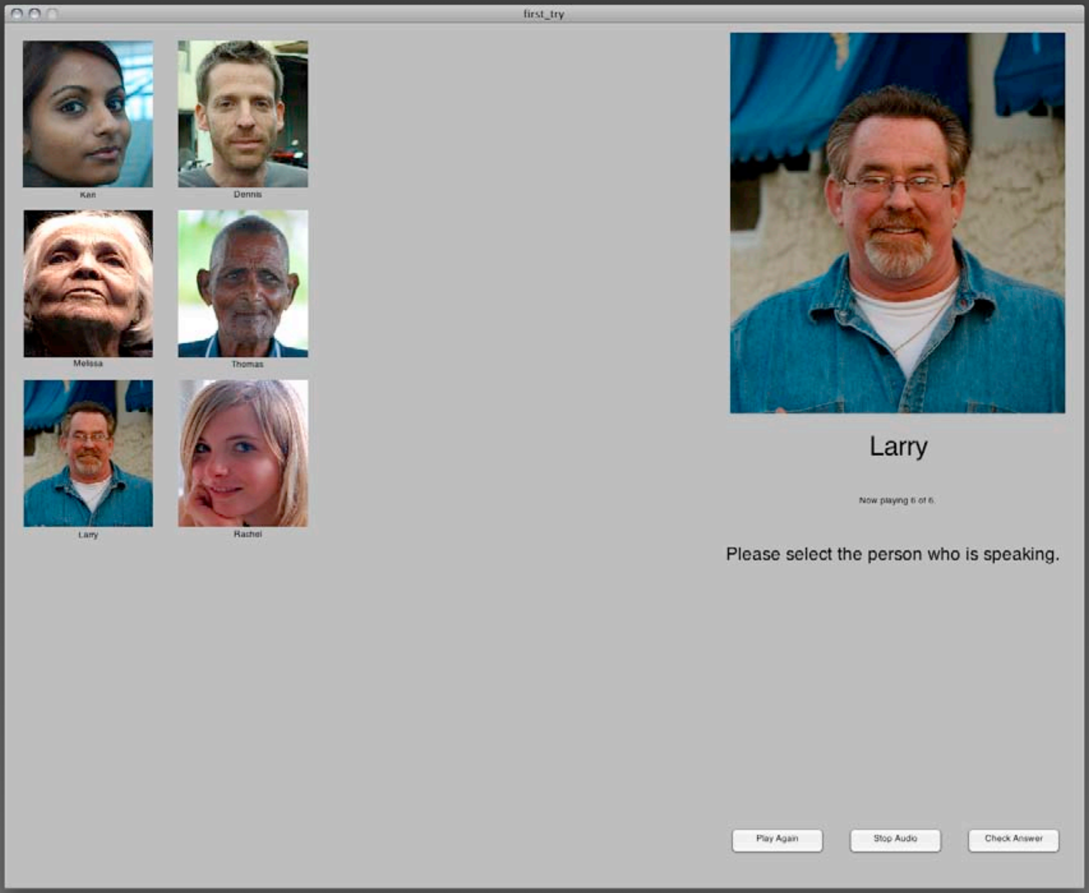
SID: Experiment Design



Now playing 6 of 6.

Please select the person who is speaking.

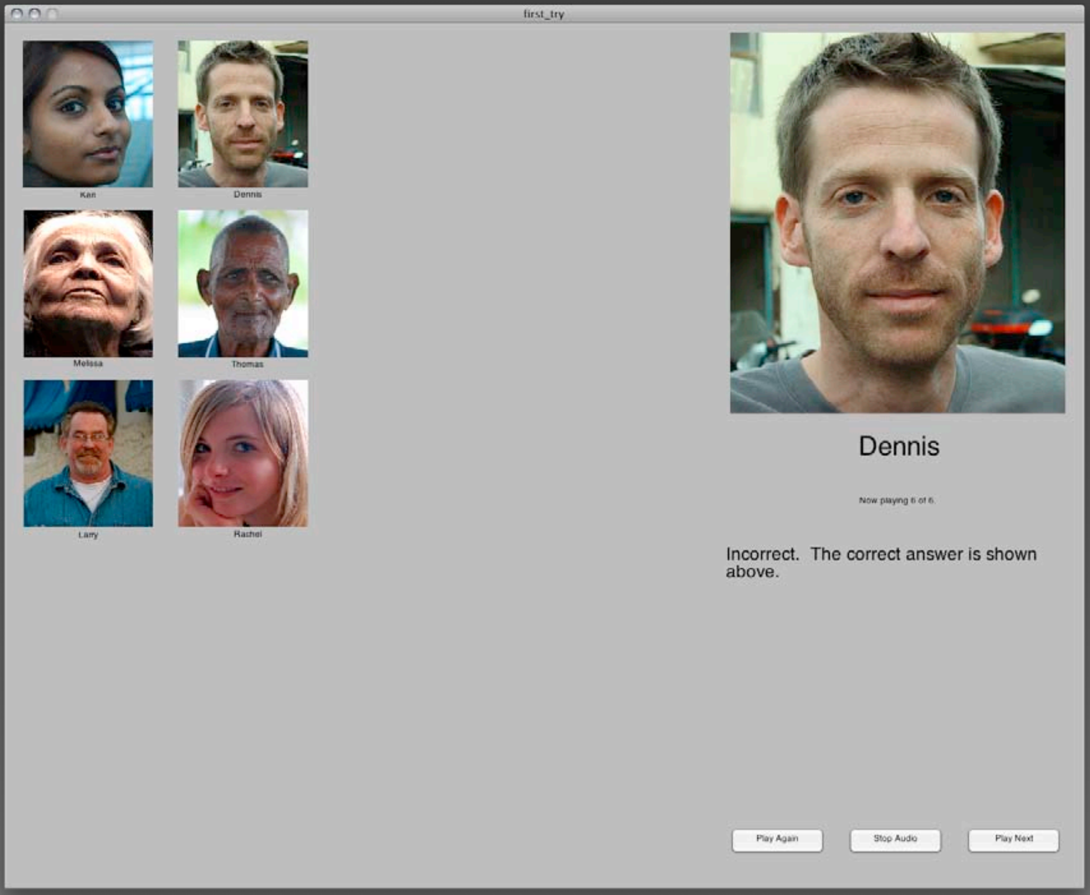
SID: Experiment Design



Now playing 6 of 6.

Please select the person who is speaking.

SID: Experiment Design



Dennis

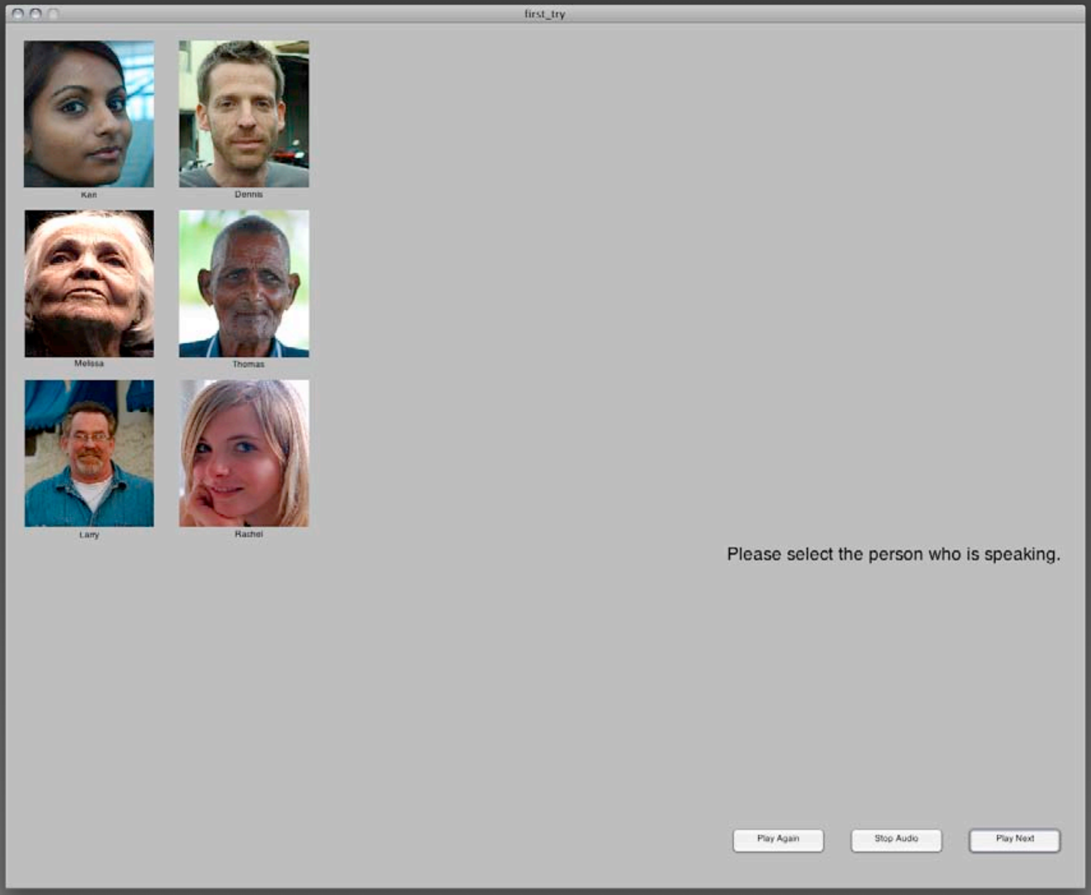
Now playing 6 of 6.

Incorrect. The correct answer is shown above.

SID: Experiment Design

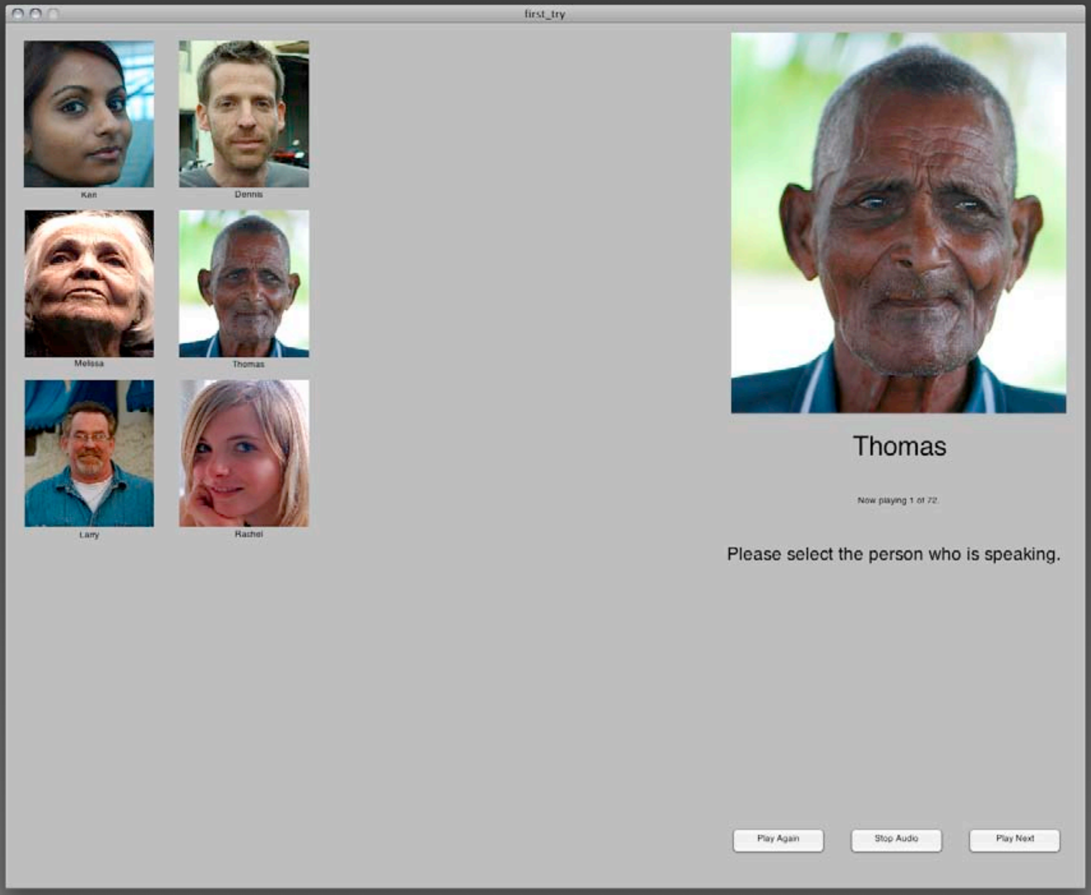
- Test sessions
  - Similar to the quiz session, but with no feedback
  - Three test sessions
  - Each clip length had its own session (sentences were first, then four digit clips, then two digit clips)
  - Same interface for all three clip lengths

SID: Experiment Design



Please select the person who is speaking.

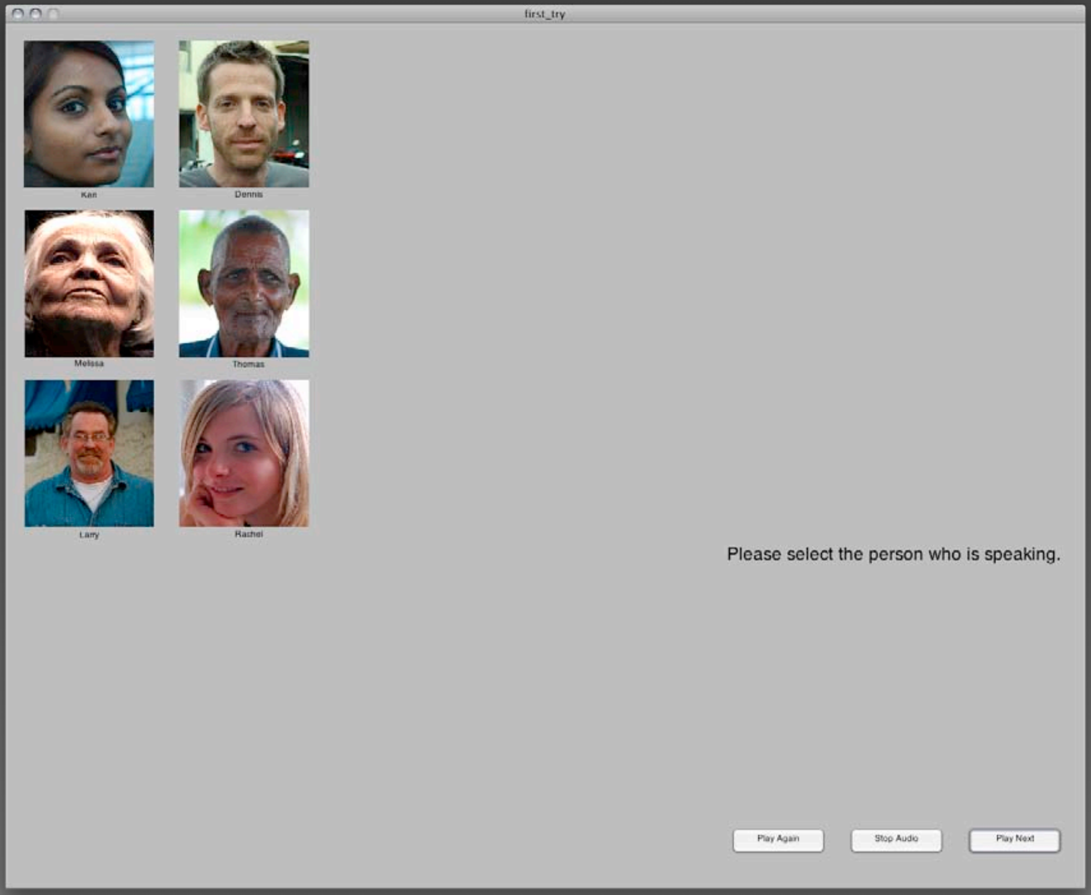
SID: Experiment Design



Now playing 1 of 72.

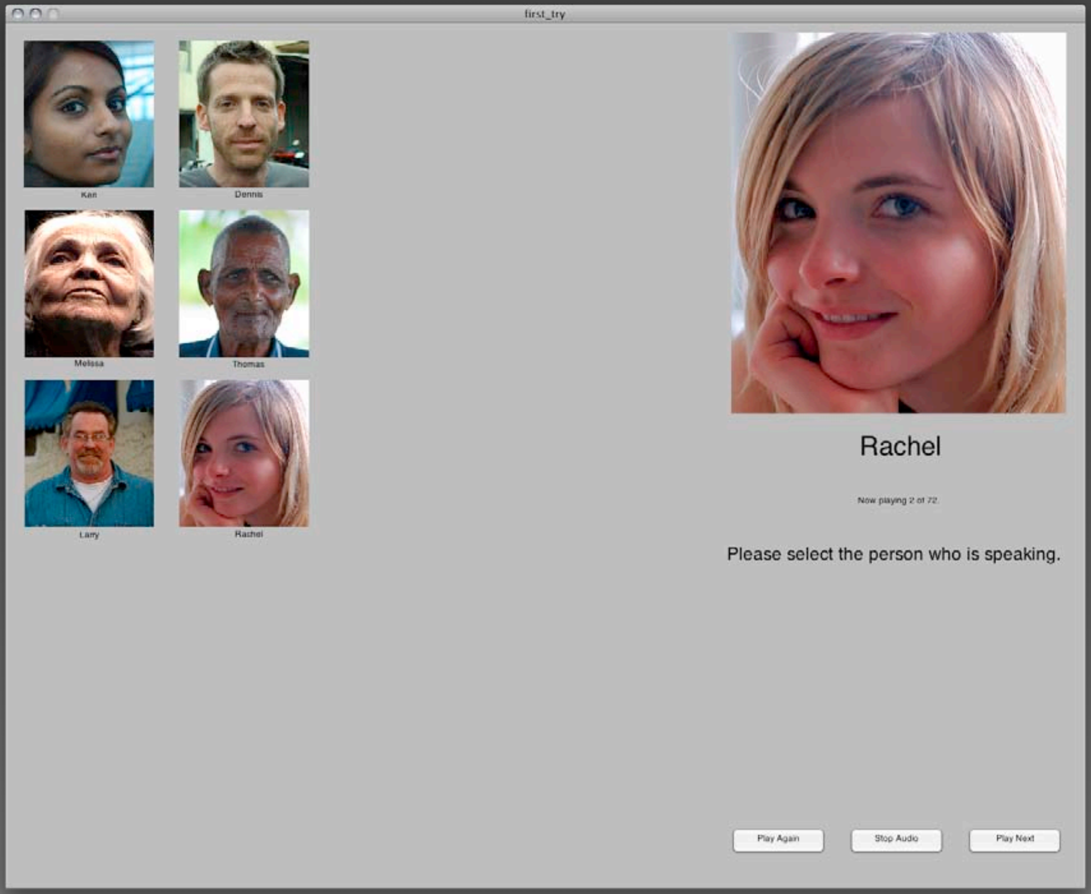
Please select the person who is speaking.

SID: Experiment Design



Please select the person who is speaking.

SID: Experiment Design



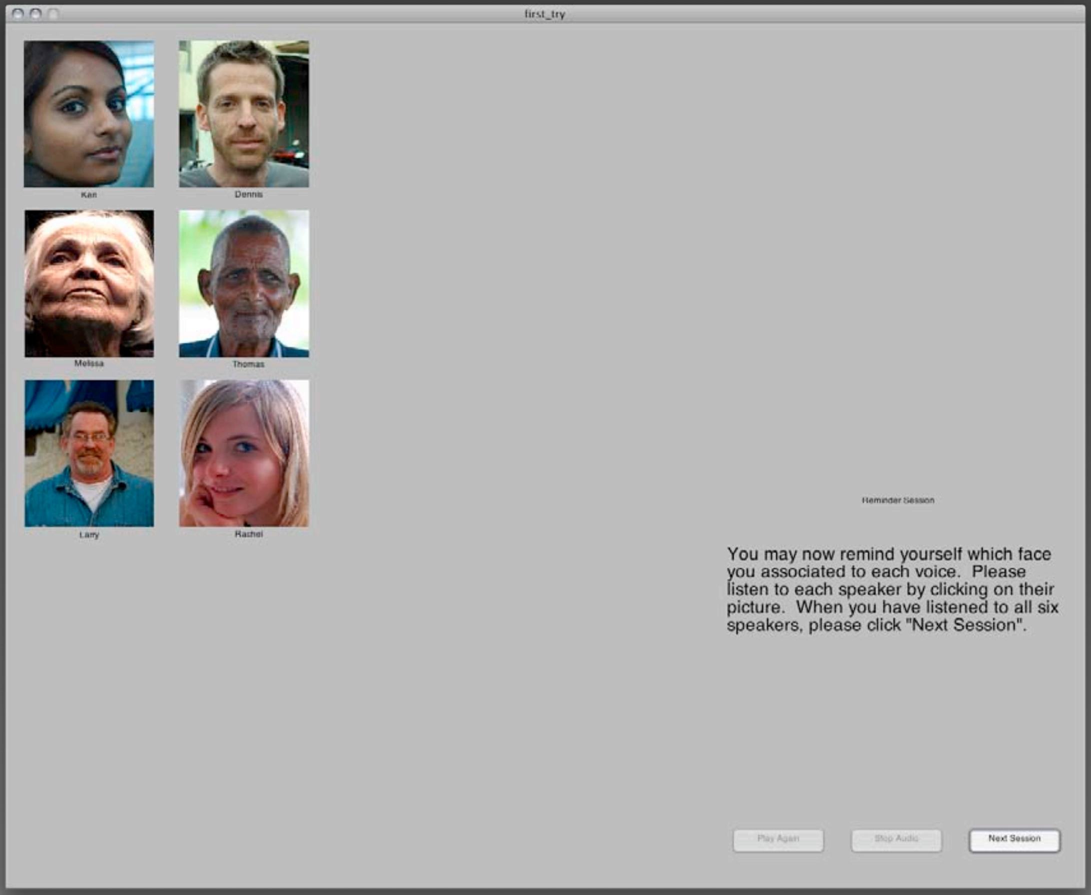
Now playing 2 of 72.

Please select the person who is speaking.

SID: Experiment Design

- Reminder sessions
  - Designed to keep the effects of mid-test mistraining from tainting results of next session
  - Clips representative of those to be heard in the next session were used
  - Listeners had to listen to each speaker at least once

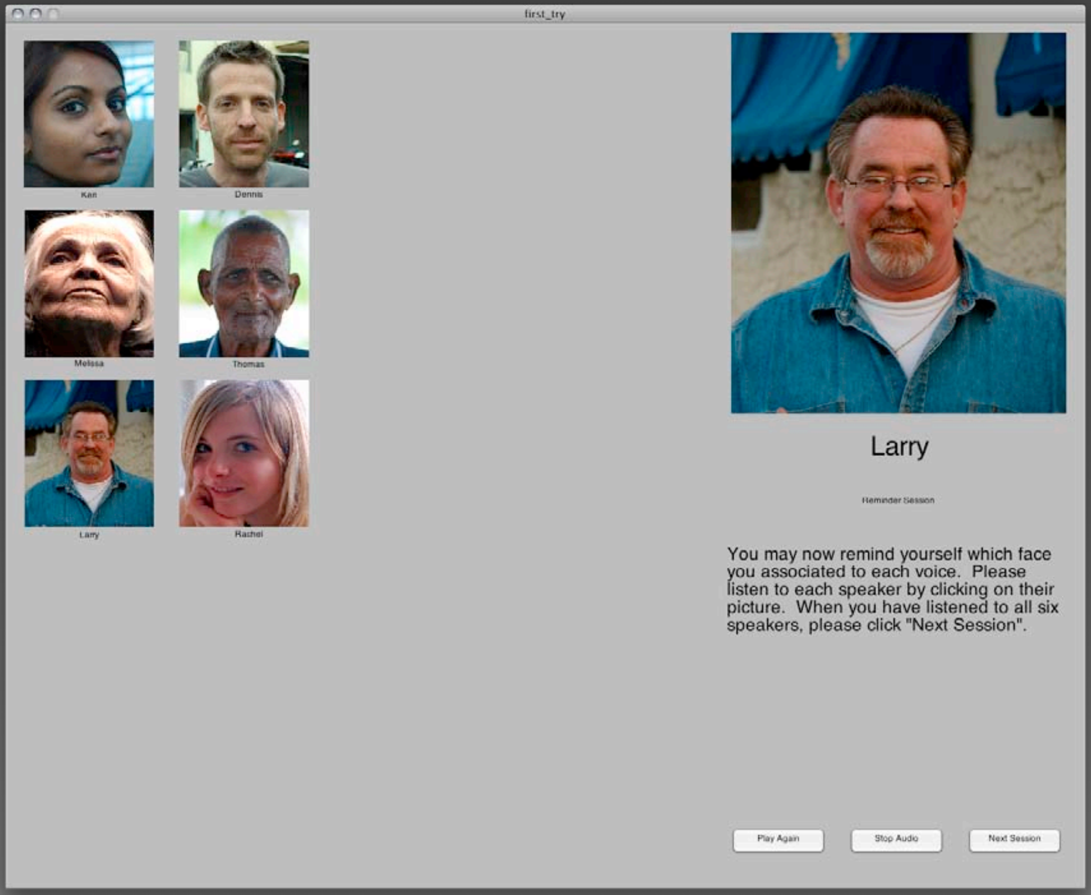
SID: Experiment Design



Reminder Session

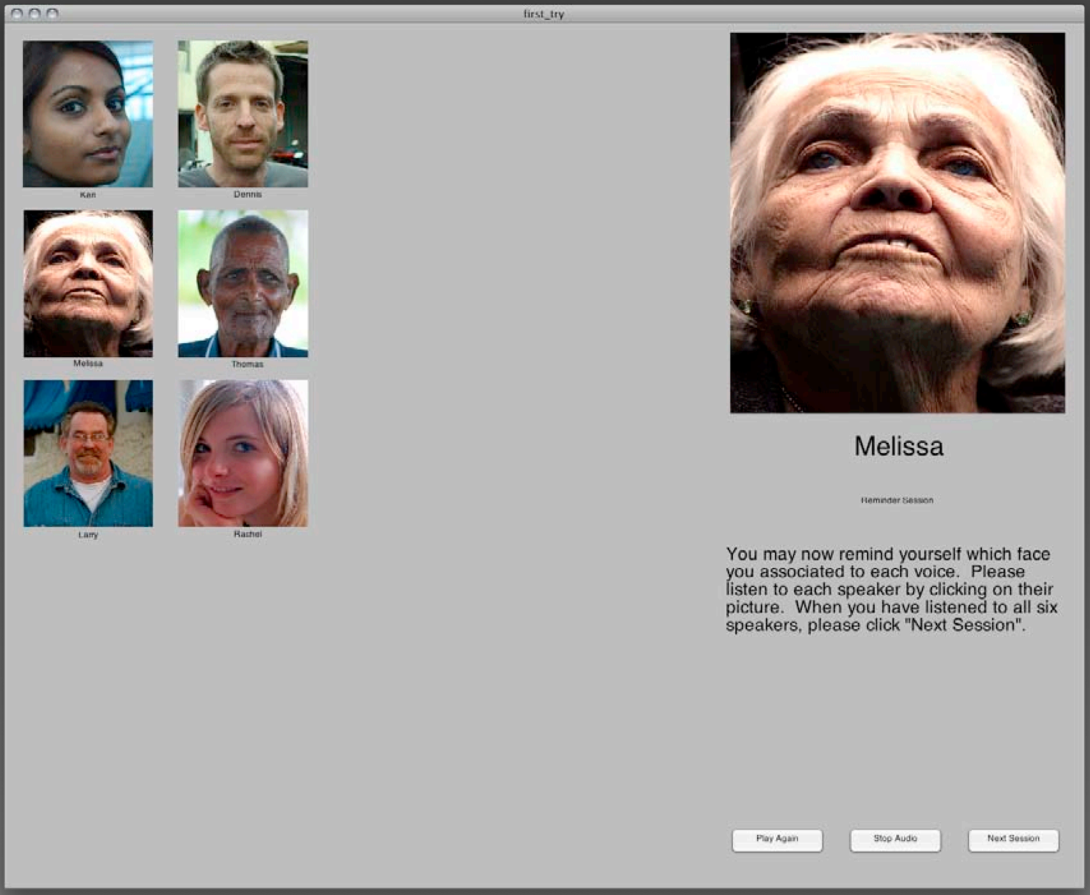
You may now remind yourself which face you associated to each voice. Please listen to each speaker by clicking on their picture. When you have listened to all six speakers, please click "Next Session".

SID: Experiment Design



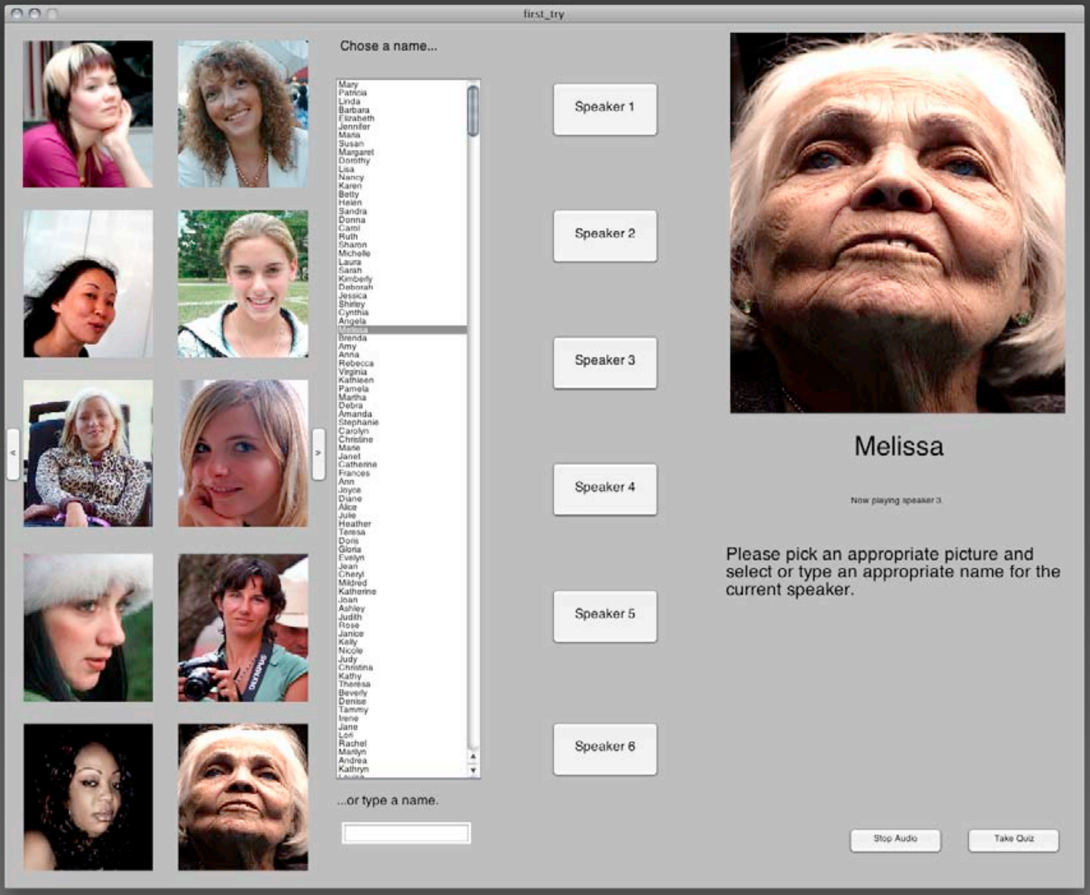
You may now remind yourself which face you associated to each voice. Please listen to each speaker by clicking on their picture. When you have listened to all six speakers, please click "Next Session".

SID: Experiment Design



You may now remind yourself which face you associated to each voice. Please listen to each speaker by clicking on their picture. When you have listened to all six speakers, please click "Next Session".

SID: Experiment Design



Now playing speaker 3

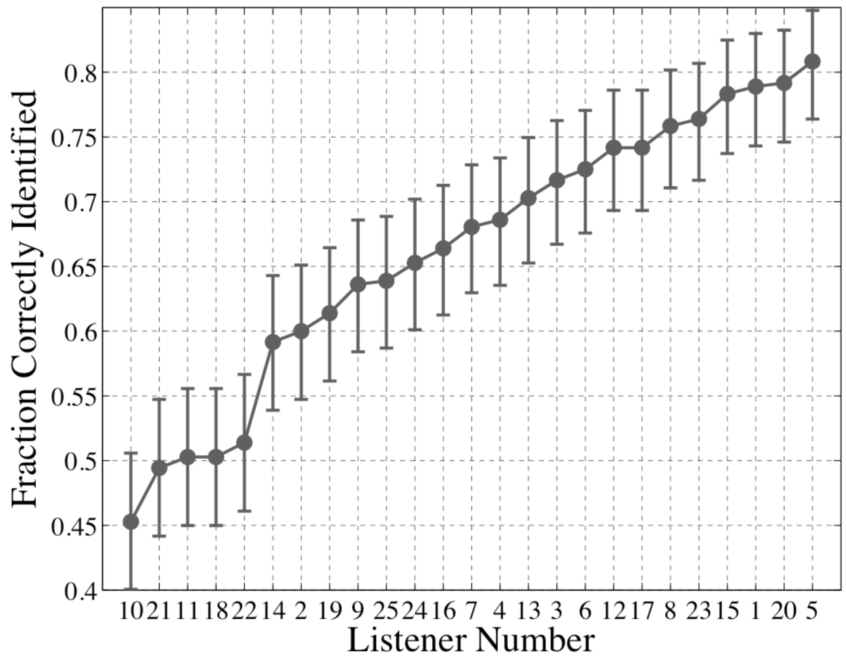
Please pick an appropriate picture and select or type an appropriate name for the current speaker.

SID: Administrivia

- 25 listeners
  - 15 male
  - 10 female
  - Ages 37-64, Mean: 49
  - Scientists, mathematicians, IT professionals, desk workers
  - Native languages: English (22), Spanish (1), German (1), Russian (1)

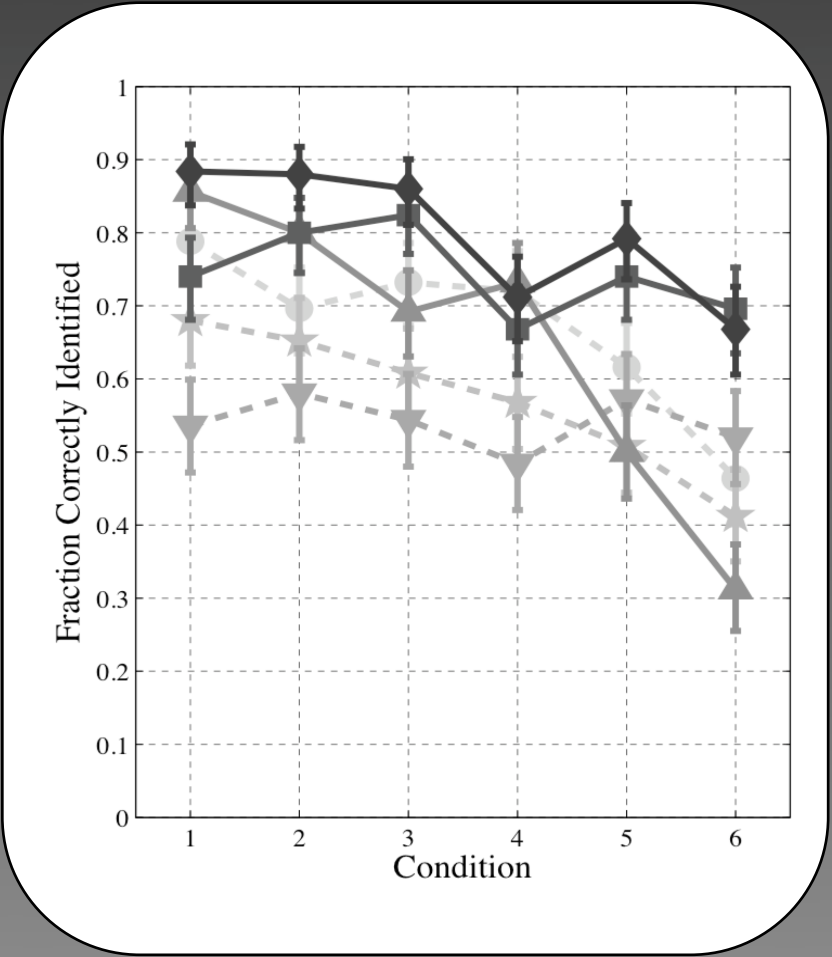
SID: Results

- Per Listener Results
  - Mean fraction of correct identifications: .662
  - 20 listeners fall between fractions .59 and .81
  - Two hearing aid users (14,16), one subject deaf in one ear (20)
  - Experiment administrator achieved a fraction correct of .98 (not included in analysis)



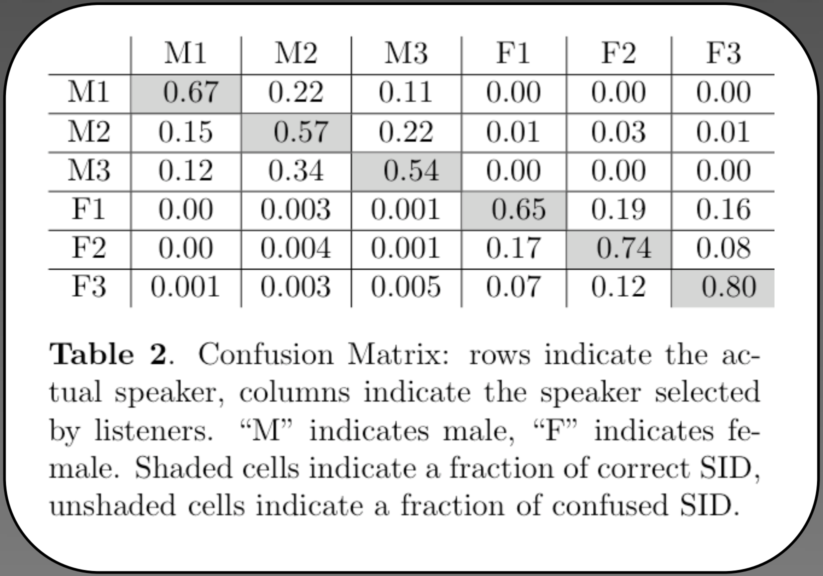
SID: Results

- Per Speaker Results
  - Dotted lines = males
  - Solid lines = females
  - One female very recognizable (also has Ecuadorian accent)
  - Males more often confused



SID: Results

- Confusion Matrix
  - Male-female confusion is very low
  - Males 2 and 3 most often confused
  - Females 2 and 3 most easily recognized

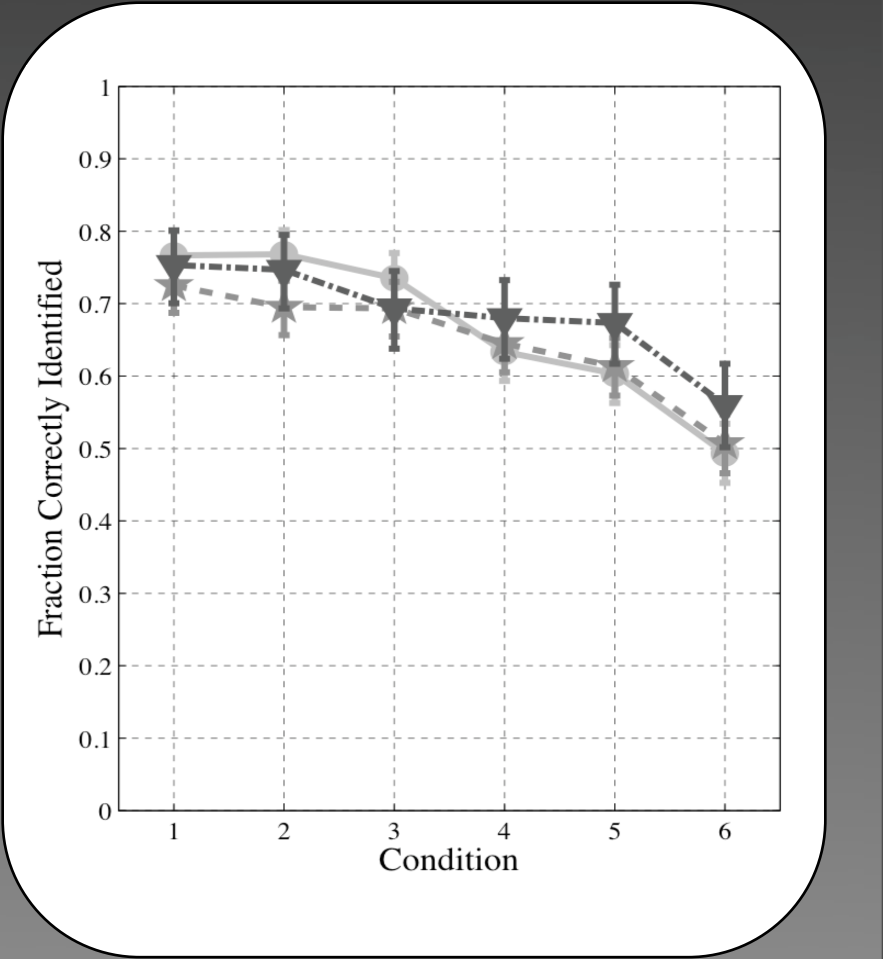


	M1	M2	M3	F1	F2	F3
M1	0.67	0.22	0.11	0.00	0.00	0.00
M2	0.15	0.57	0.22	0.01	0.03	0.01
M3	0.12	0.34	0.54	0.00	0.00	0.00
F1	0.00	0.003	0.001	0.65	0.19	0.16
F2	0.00	0.004	0.001	0.17	0.74	0.08
F3	0.001	0.003	0.005	0.07	0.12	0.80

Table 2. Confusion Matrix: rows indicate the actual speaker, columns indicate the speaker selected by listeners. "M" indicates male, "F" indicates female. Shaded cells indicate a fraction of correct SID, unshaded cells indicate a fraction of confused SID.

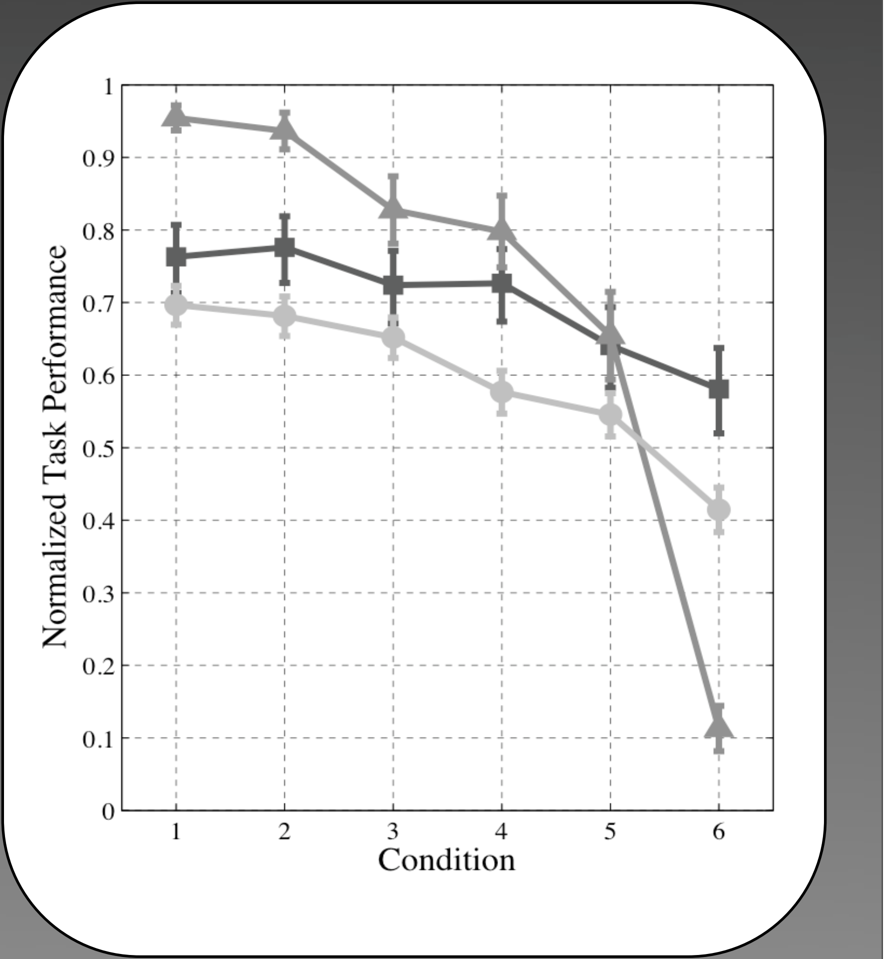
SID: Results

- Per Length Results
  - Interesting outcome: no length is significantly easier!
  - Consistent with prior research, but unintuitive
  - Experimental order (sentence, four digits, two digits) may have had an effect



SID: Results

- SID Vs. Intelligibility and Stress Detection
  - SID is not as robust as dramatized urgency (DU) detection
  - About 3 times more robust than intelligibility
  - Light gray: SID, medium gray: intelligibility, dark gray: DU detection

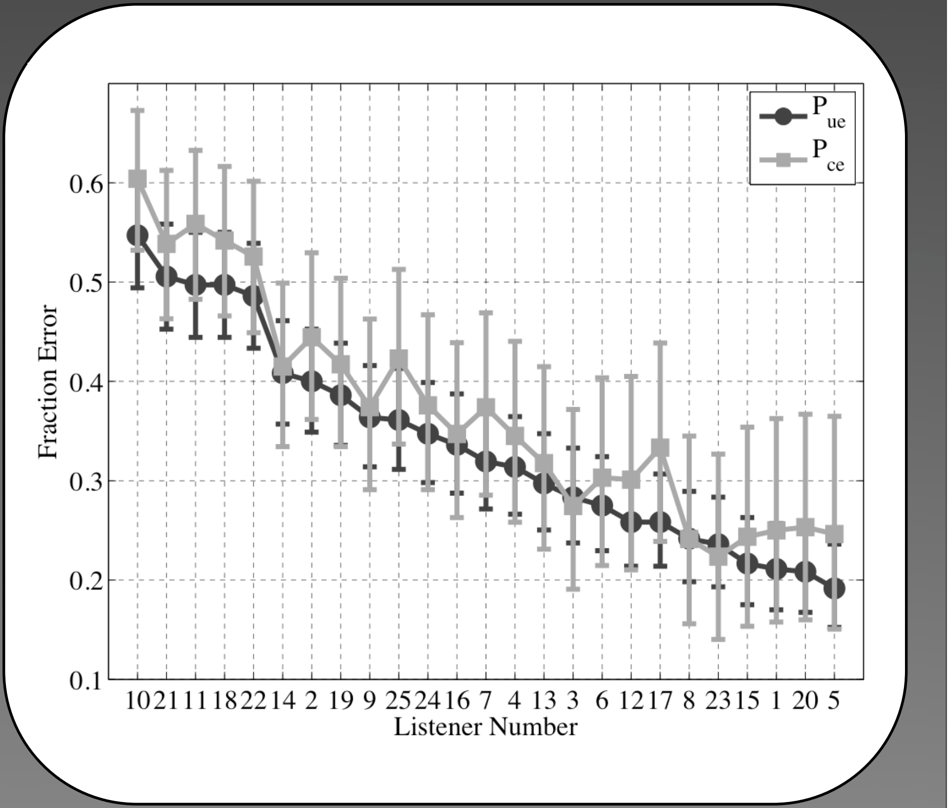


SID: Post-Hoc Work

- We had these questions while we were conducting the test:
  - Is an "event" causing temporary mistraining?
  - How often does a "confusion" result in a more permanent mistraining?
  - How often is a speaker assigned a similar memory aid?
  - How often are clips replayed?

SID: Post-Hoc Work

- Many listeners showed a slight tendency towards “bursty” errors
- Clearly not enough data
- Can’t say anything about permanent mistraining either



SID: Post-Hoc Work

M1

3 listeners

3 listeners

3 listeners

3 listeners

M2

5 listeners

4 listeners

3 listeners

M3

3 listeners

2 listeners

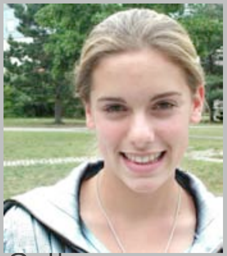
2 listeners

SID: Post-Hoc Work

F1

7 listeners

4 listeners



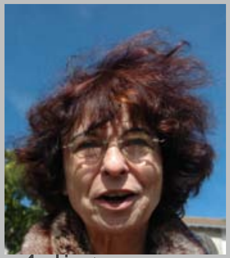
2 listeners

F2

5 listeners

4 listeners

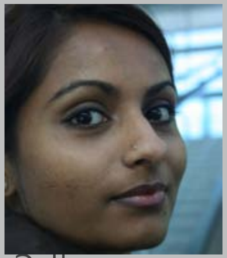
4 listeners



4 listeners

F3

5 listeners



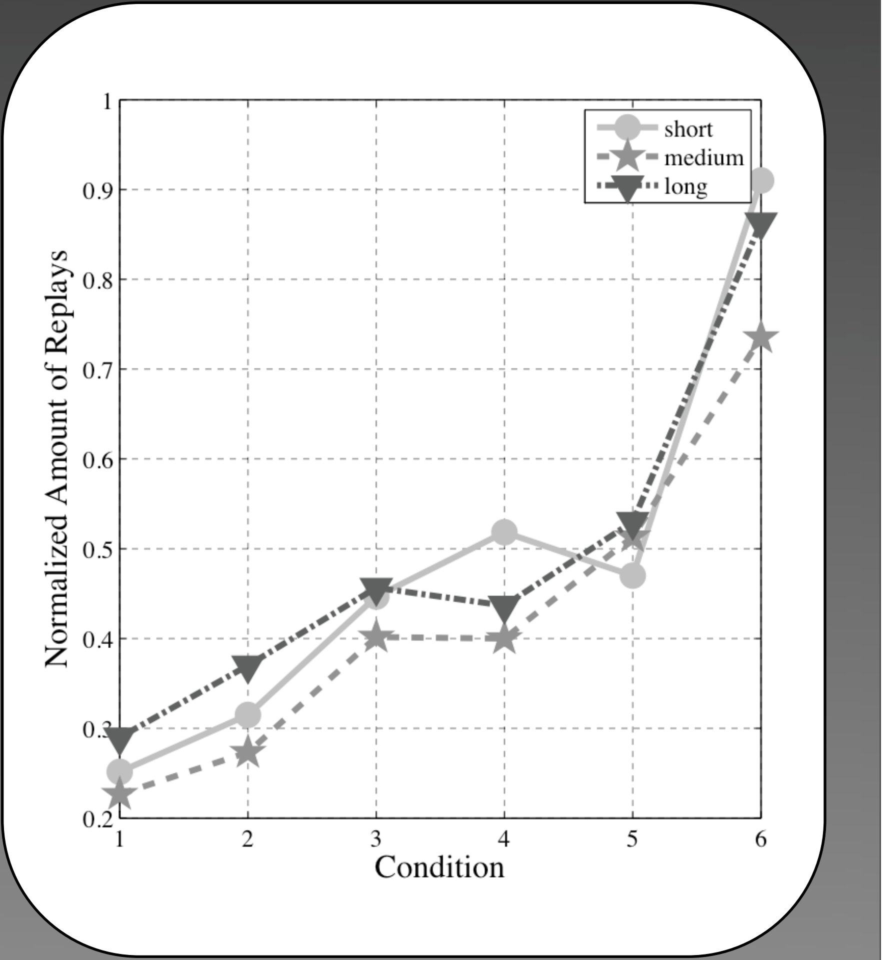
3 listeners

3 listeners

3 listeners

SID: Post-Hoc Work

- C1 replayed 20-30% of the time, on average
- C6 replayed 70-90% of the time, on average
- Number of replays goes up with difficulty
- Amount of prosodic information might have been a source of listener confusion



SID: Open Questions

- Consult with experts in psychology and neurology to design lab tests that more closely model real world situations
- Attempt an experiment with better controlled recordings and familiar speakers

SID: In depth

- Paper covering results published in the conference proceedings of MESAQIN 2008: http://wireless.feld.cvut.cz/mesaqin/contributions.html