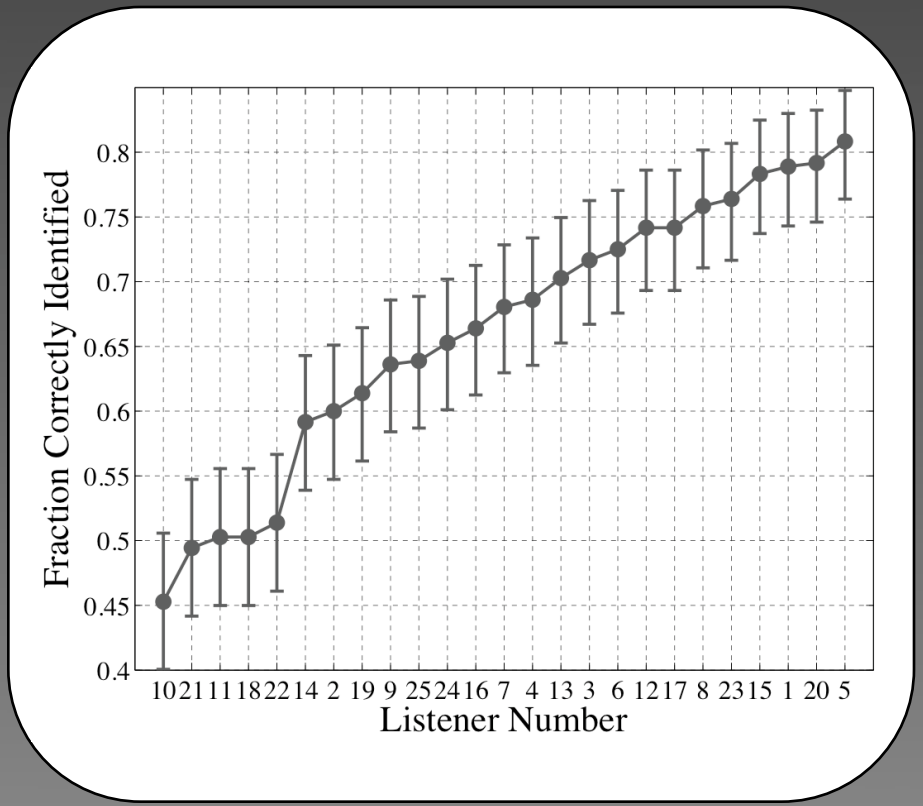
SID: Administrivia

- 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)

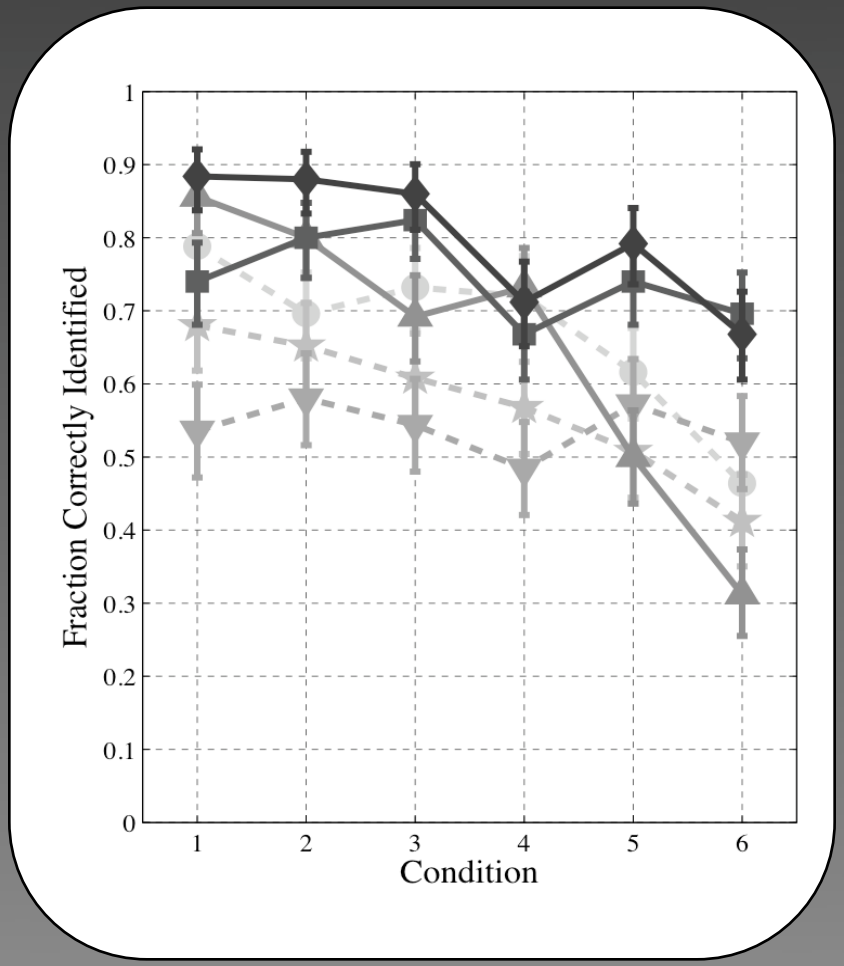
SID: Results

- 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

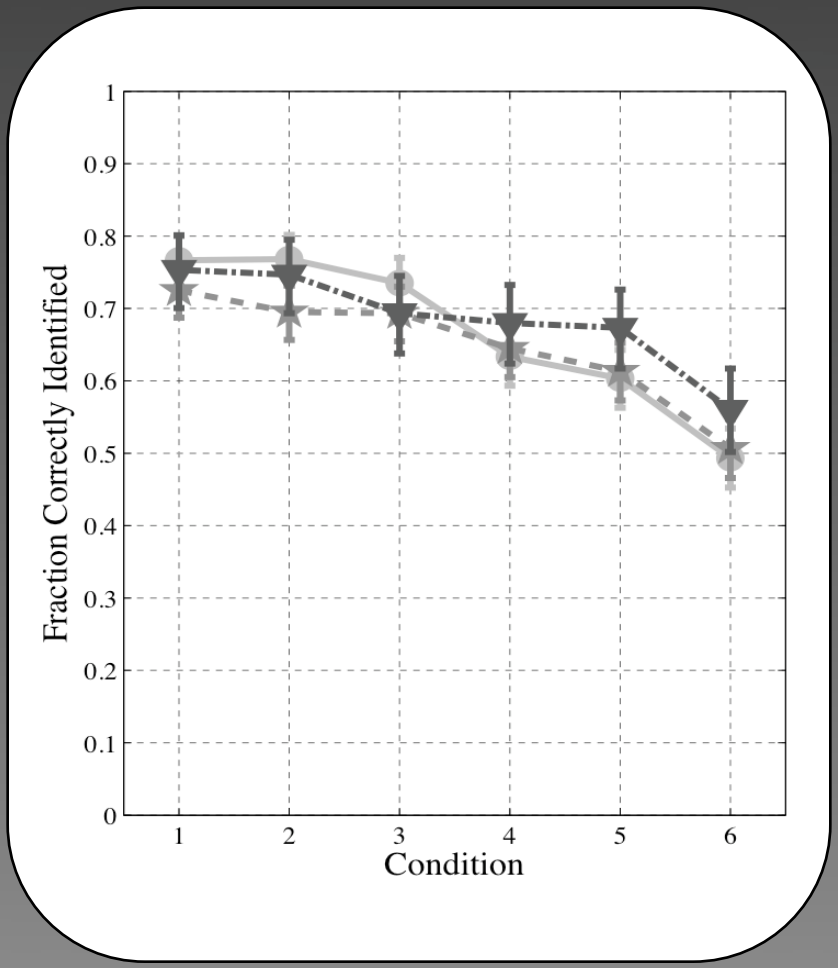
- Confusion Matrix
  - Male-female confusion is very low
  - Males 2 and 3 most often confused
  - Females 2 and 3 most easily recognized

	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.

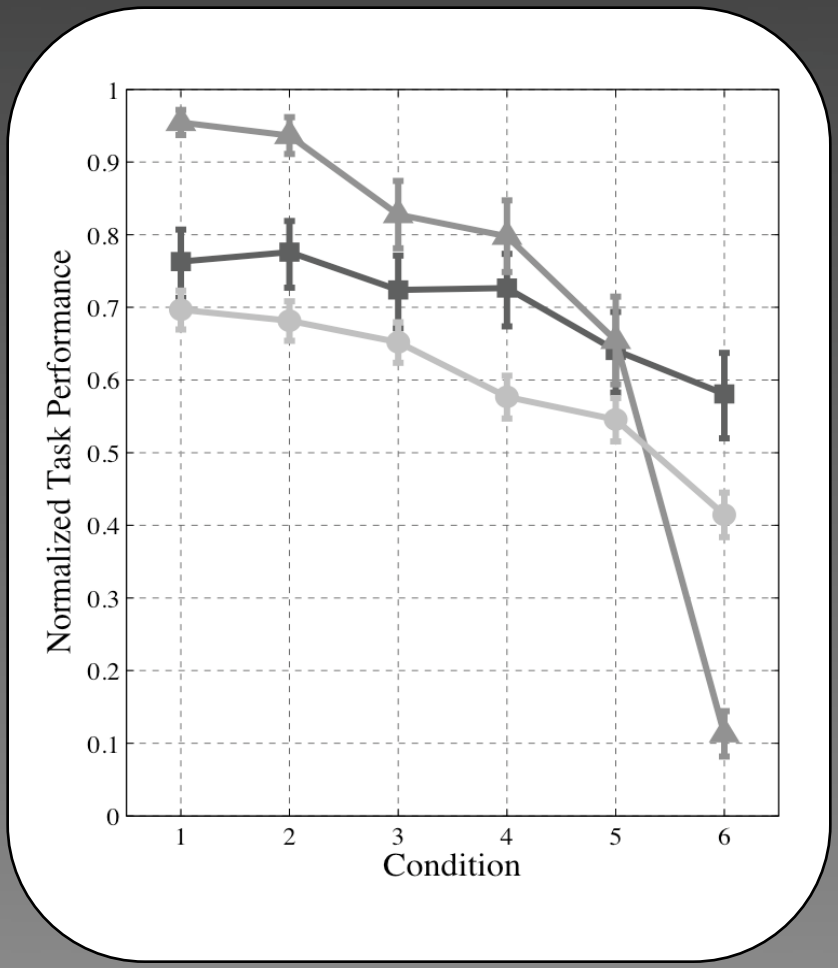
SID: Results

- 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

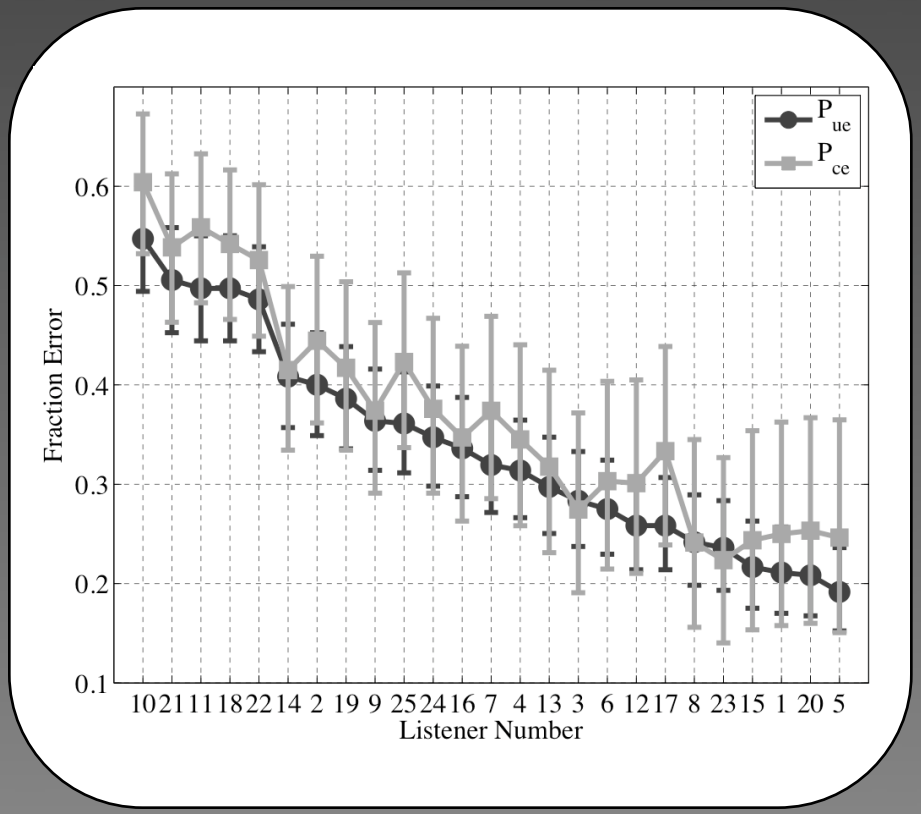


SID: Post-Hoc Work

- 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

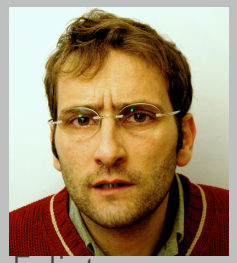

3 listeners


3 listeners


3 listeners


3 listeners

M2


5 listeners


4 listeners


3 listeners

M3


3 listeners


2 listeners

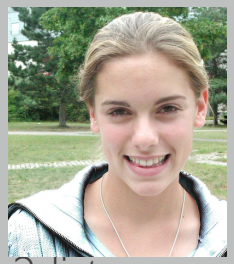

2 listeners

SID: Post-Hoc Work

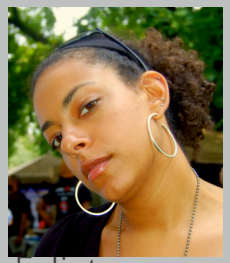
F1


7 listeners

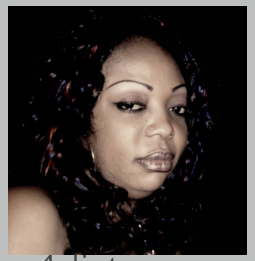
4 listeners


2 listeners

F2

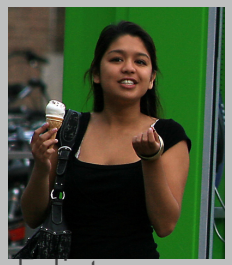

5 listeners

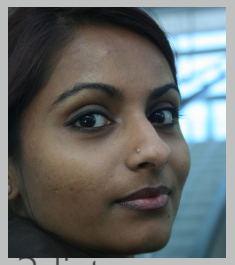

4 listeners


4 listeners


4 listeners

F3


5 listeners

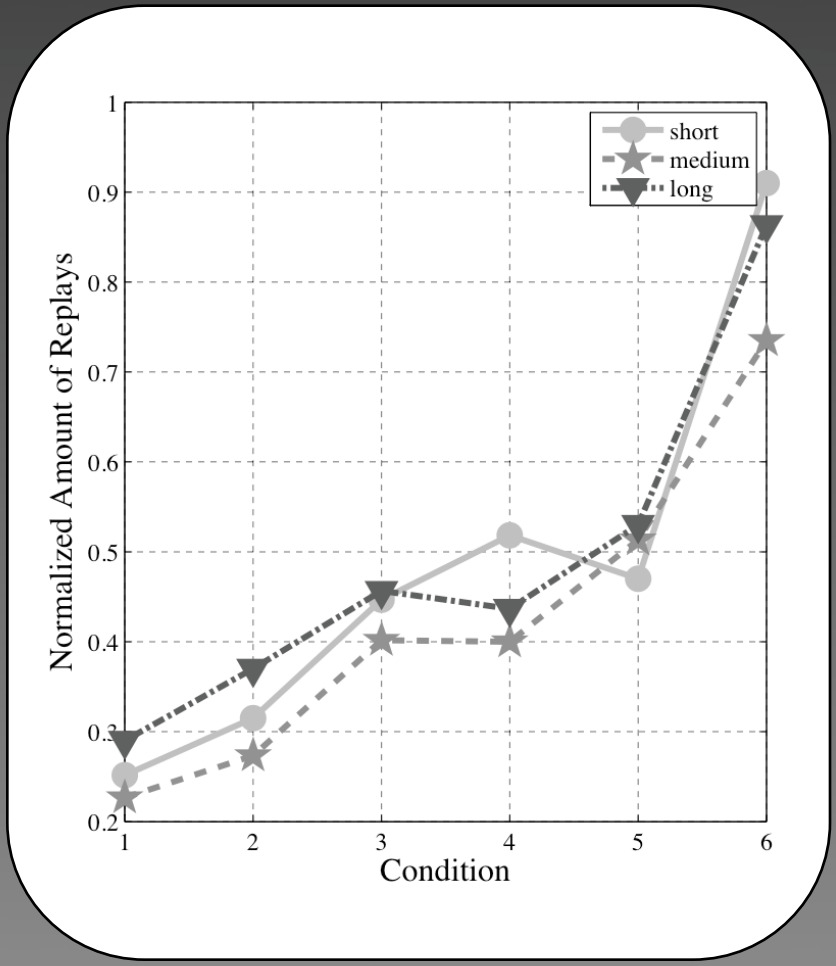

3 listeners


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