Foreign Accented Speech: *Encoding and Generalization*

Shawn A. Weil

Language Perception Laboratory

Ohio State University

ASA Chicago - June 7, 2001

http://lpl.psy.ohio-state.edu

weil.17@osu.edu

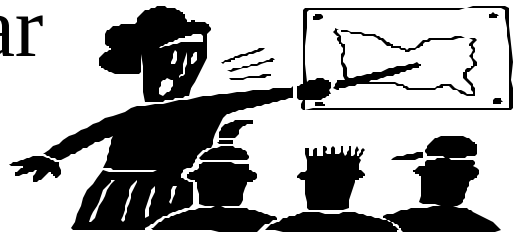
This Poster is Available At:

http://lpl.psy.ohio-state.edu

weil.17@osu.edu

Scenario

Imagine you are enrolled in a university level lecture. You find out the first day that the instructor's native language is not English, and that he/she has a noticeable accent. At first you have trouble understanding him/her, but as the semester goes on, you have less and less trouble. Why? What happens next semester, with a new instructor with a similar accent?



Why study accented speech?

- FAS miscommunications.
- Problem for speech recognition software.
- Issues in second language acquisition
- “Extreme case” of speaker variability may help in designing models of our speech perception abilities.

Questions

What type of information is encoded when listening to an unaccented talker?

What type of information is encoded when listening to an accented talker?

Can this information generalize to other accented talker, to aid speech perception?

What is an Accent?

- Similarity of phonemes between native (L1) and second (L2) language.
- L2 contrasts can be confused with L1 contrasts. (Flege, 1995)
- Individuals who share a common L1 share similar accents. (Sutter, 1980; Rogers, 1997)

Evidence for Adaptation

- Nygaard and Pisoni (1998): Listeners encode idiosyncratic talker characteristics.
- This information helps subsequent processing.
- FAS perception is just an extension of this talker variability compensation.

Nygaard, Sommers, and Pisoni (1994)

	Training	Post-Test
Length	9 Days	10th Day
Type	Feedback	No Feedback
	10 talkers	Same or Different Talkers

- New utterances spoken by familiar voices had consistently higher recognition rates (~51% versus ~42%).
- **Indexical information is stored, and helps speech perception implicitly.**

Indexical Information

- Talker idiosyncrasies (F0, speaking rate, odd pronunciations)
- Indexical Information is stored and utilized. (Nygaard, Sommers, and Pisoni,1994)
- What about the properties of accents?
  - Independent of talker idiosyncrasies?
  - Similar between accented talkers with common linguistic background.

Adaptation to Synthetic Speech

- Schwab, Nusbaum, and Pisoni (1985)
- Analogous to FAS
- Speech produced by a computer designed to emulate human performance.
- 1 Pre-test, 8 days of training, 1 post-test.
- Ss received a battery of tests on all days, and received feedback on their responses during the training days.
- Three Btwn Ss groups: No training, natural speech, and synthetic speech.

Results of Schwab, et al

- Synthetic Speech group had much improved performance compared to control groups.
- Adaptation to difficult speech can be accomplished with exposure.

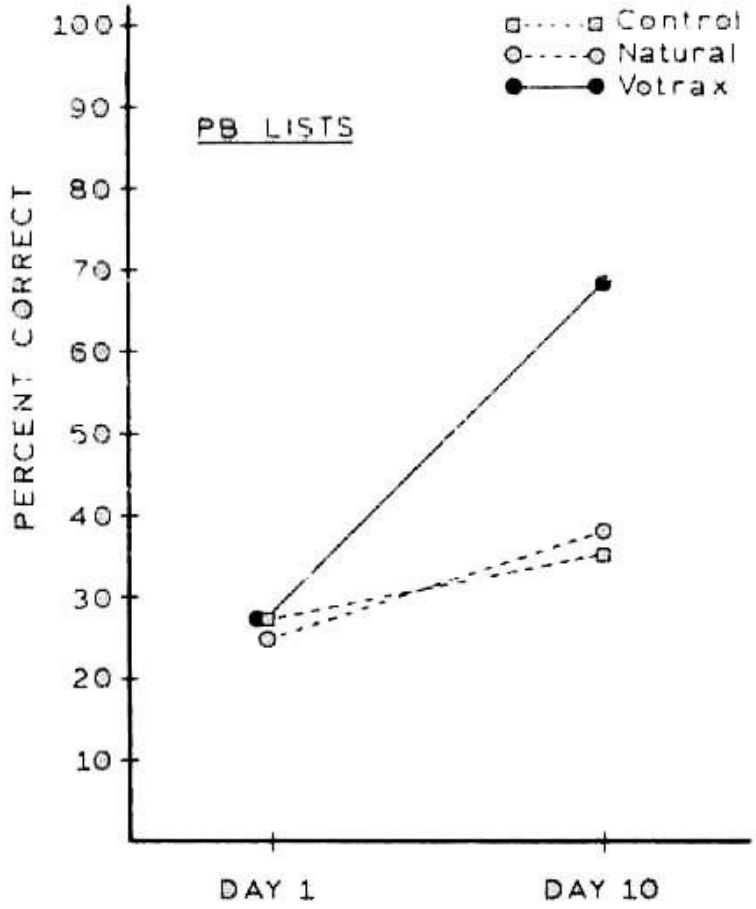


Figure 1. *Mean accuracy for Day 1 (pretest) and Day 10 (posttest) for transcribing isolated synthetic words (PB Lists).*

So…

- Listeners utilize indexical information, and this helps during subsequent speech perception (Nygaard et al,1994).
- Exposure to non-typical speech (i.e., synthetic speech) resulted in improved word identification and transcription. (Schwab et al,1985).

Does this generalize between similar talkers?

Evidence for Generalization

Goldinger (1996)

- Subjects trained on set of talkers.
- Tested on second set of novel talkers.
- Better performance for novel talkers that are close to trained talkers in multidimensional space.

Conclusion: Talkers who share indexical characteristics lead to generalization.

Current Investigation

- Five Days
- Testing Days: 1 and 5 (Signal Correlated Noise)
- Training Days: 2 – 4
- 2 x 3 Between Subjects Design:
  - Experience: “Training” v “No Training”
  - Day 5 Talker: M1, M2, R1
- Talkers: 2 Marathi Talkers
  1 Russian Talker
- Listeners: 105 Native English Speakers
- Tasks: Transcription of Sound and Anomalous Words and Sentences. Schwab et al (1985)

Overall Design

Experience	Talker: Days 1-4	Talker: Day 5	n
“Training”	M1	M1	14
	M1	M2	15
	M1	R1	14
“No Training”	none	M1	22
	none	M2	21
	none	R1	19

Talker Information

Talker	Age	Native Language	Other Languages	AOA	Age of First English Instruction
M1	26	Marathi	Hindi	25	6
M2	26	Marathi	Hindi	24	3.5
R1	21	Russian	n/a	17	7

Battery of tests

1) Harvard Sentences:
   - The **source** of the **huge river** is the **clear spring**.
2) Haskins Sentences:
   - The **old corn cost** the **blood**.
3) Phonetically Balanced (PB) Words.
   - Fifty monosyllabic words, balanced for English phonology.
4) Modified Rhyme Test (MRT):
   - dig dip did dim dill din
5) Prose Passages

Tasks

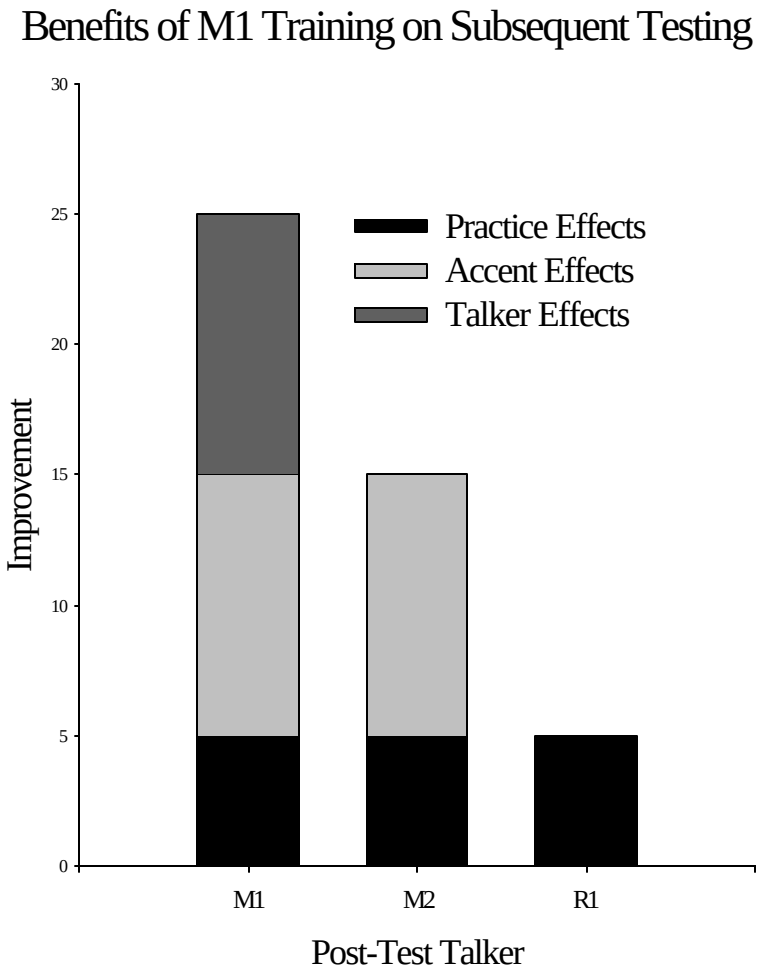
Task	Response Type	Responses	Dependent Measure
1 PB List	Free Transcription	50	Correct Transcription
1 Haskins List	Free Transcription	40	Correct Transcription
4 Prose Passages	True/False	20	Correct Comprehension
1 Harvard List	Free Transcription	50	Correct Transcription
2 MRT Lists*	6 AFC	100	Correct Identification

**MRT task only presented on Days 1 and 5*

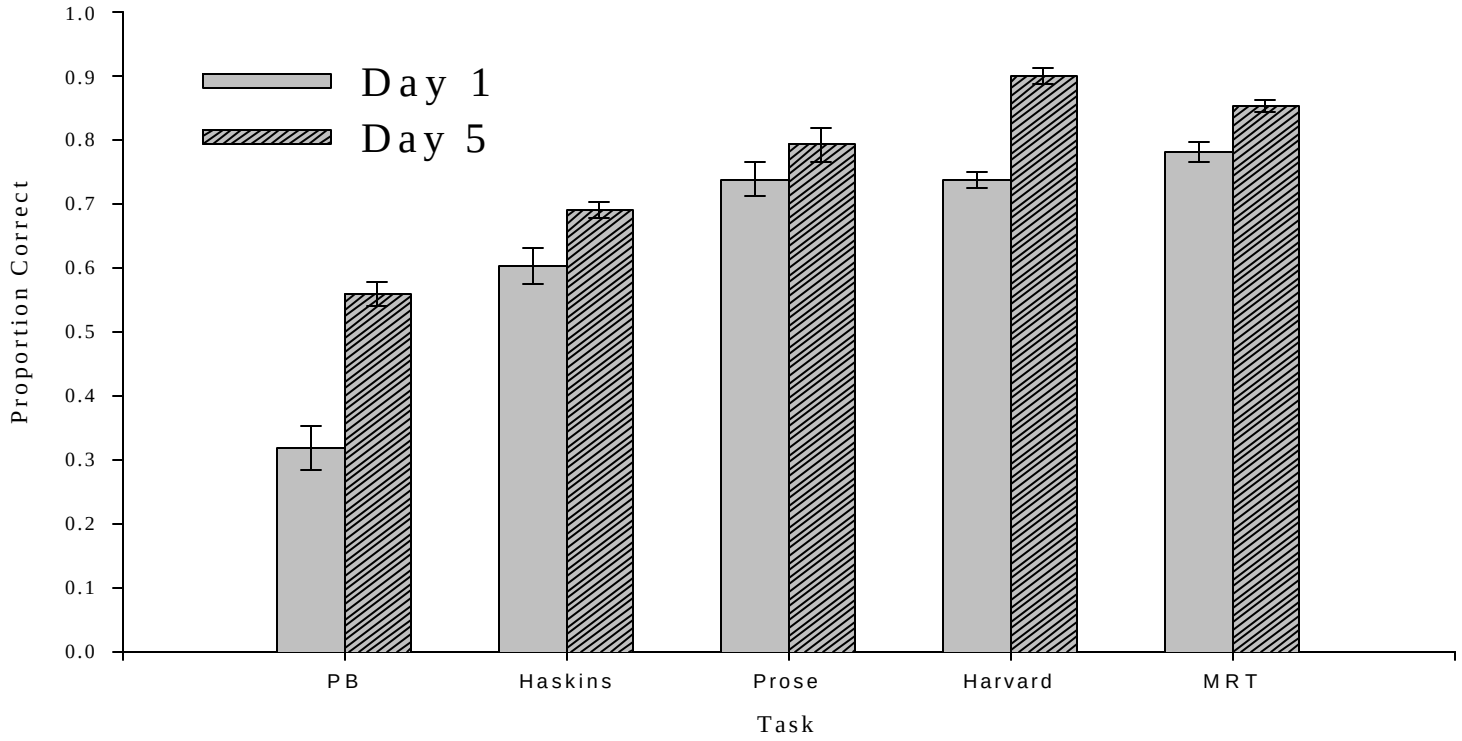
Expected Results

- Three reasons for improvement :
  1) Practice Effects
  2) Talker Effects
  3) Accent Effects

- For each task, take the *difference* between "Training" and "No Training" conditions.

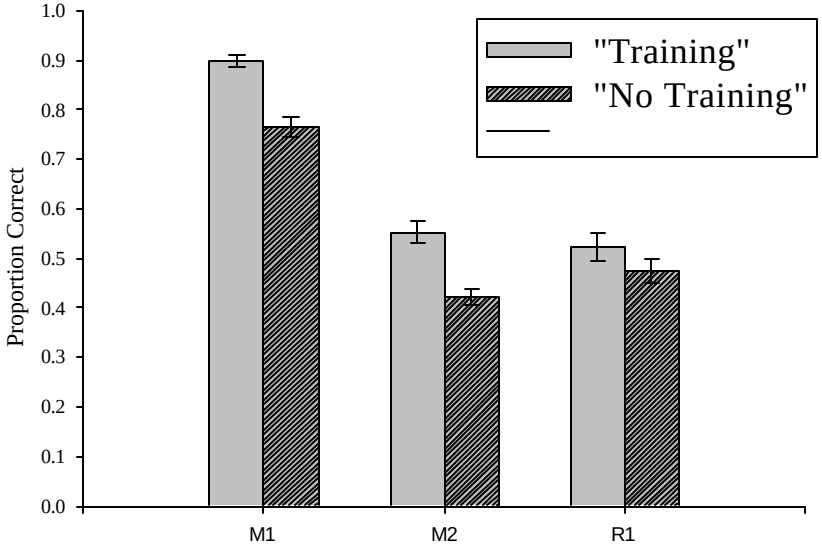


Adaptation Results

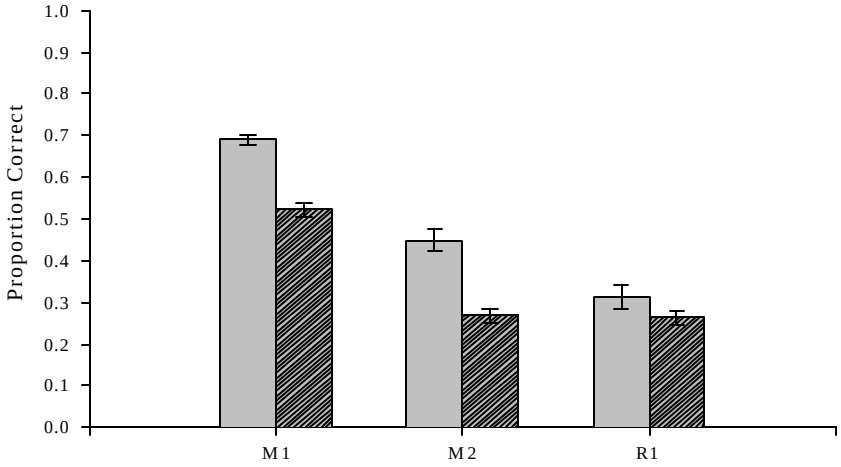


Within Subject Improvement for M1 "Training" Subjects

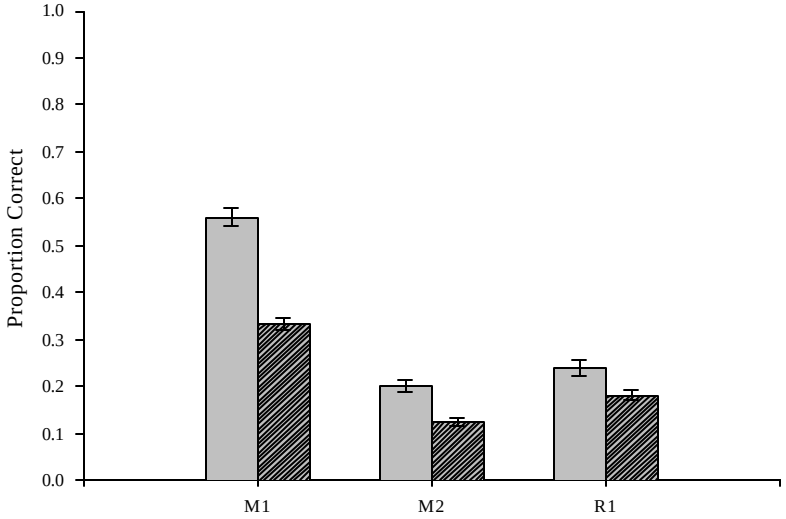
Harvard Results



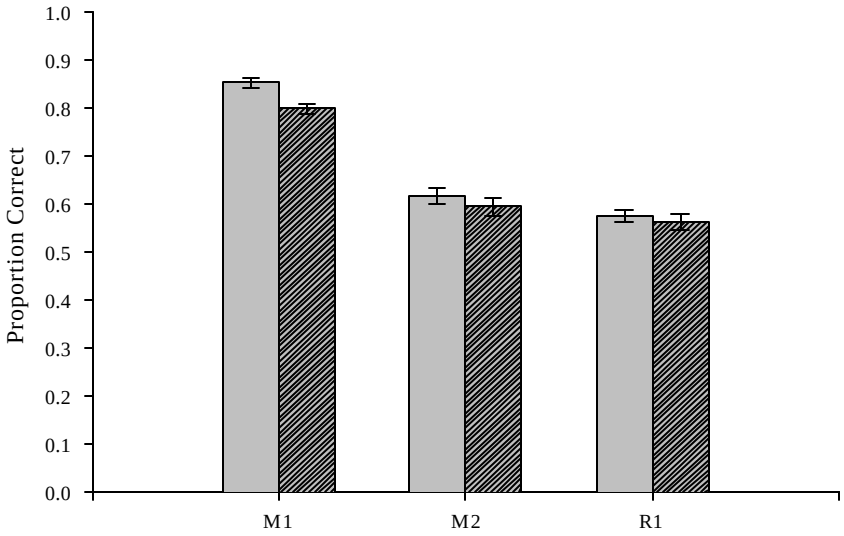
Haskins Results



PB Results

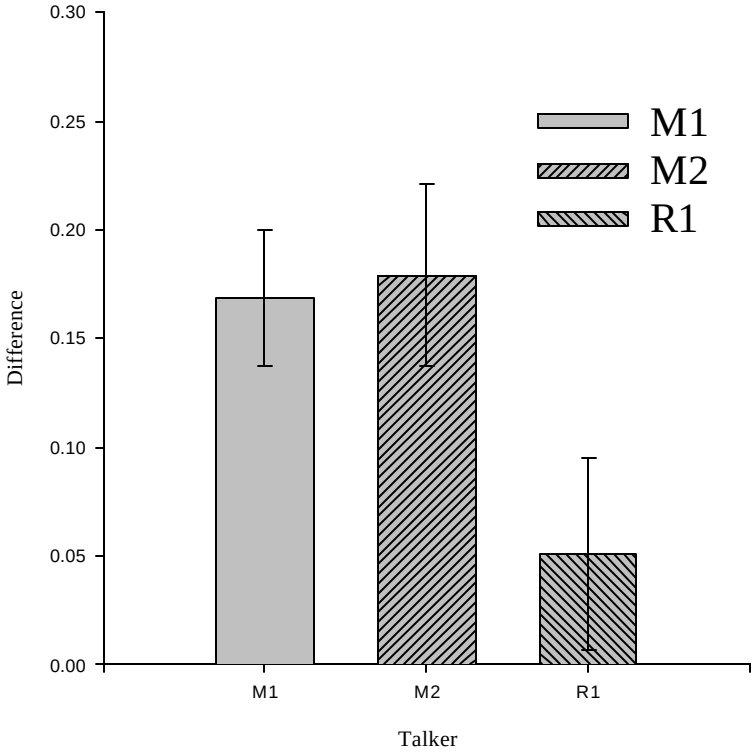


MRT

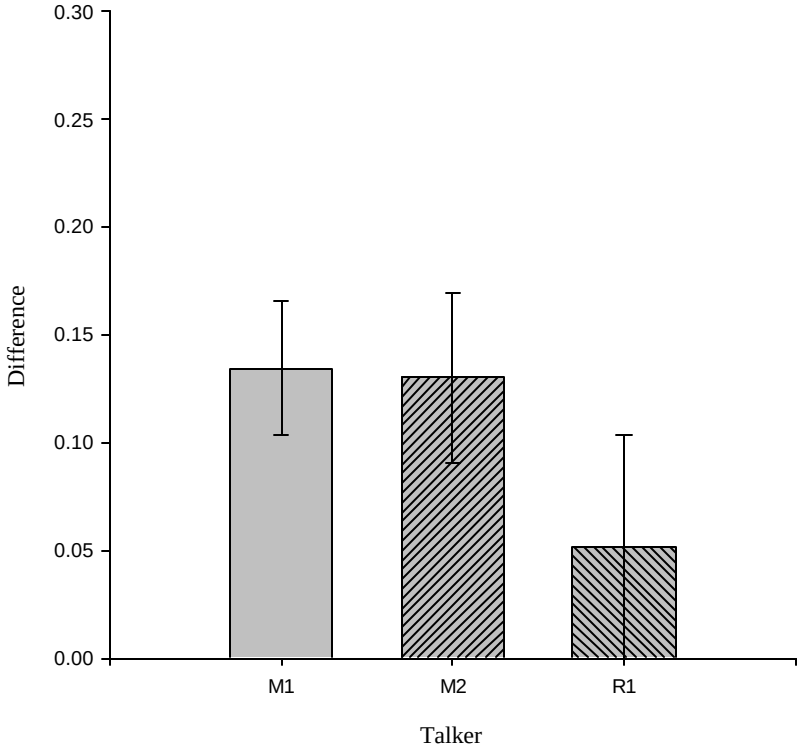


Sentence Task Improvement

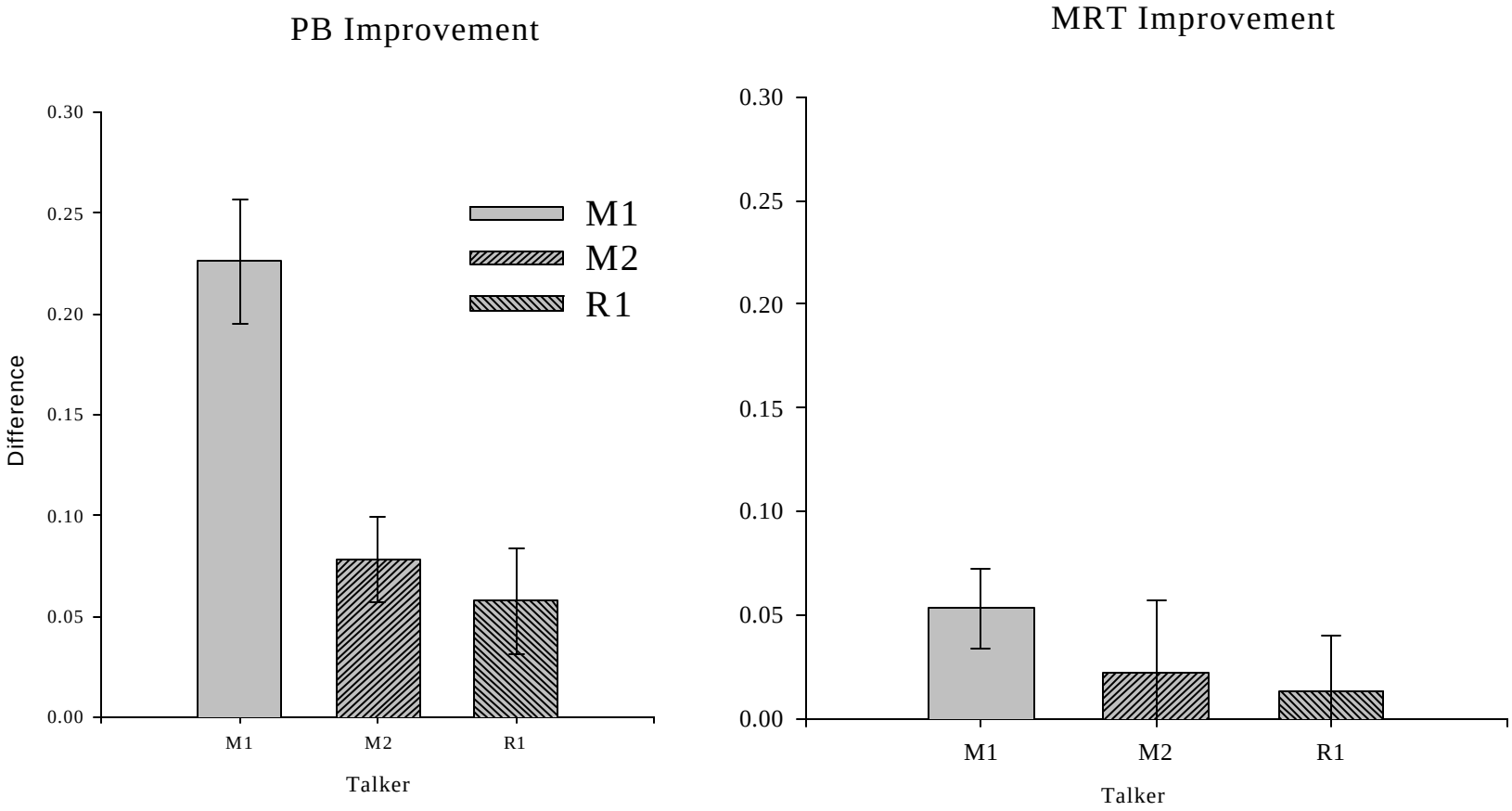
Haskins Differences



Harvard Differences



Word Task Improvement



Summary of Results

- Listener's encode both talker and accent information, and this information is utilized in subsequent speech perception.
- As expected, M1 post-test performance is the highest, R1 is the lowest.
- This ability may be limited by context; Sentence tasks showed more *accent effects* when compared to word tasks.

What do these results mean?

- M2 performance, which was expected to be in between M1 and R1, seemed to change as a function of the type of task.
  - Contextual Cues in Sentence tasks?
  - Interaction of SCN and phonetic cues?
  - Increased salience of prosody in longer sentence task?

The Future

- Can accented speakers utilize these results for more efficient communication?
- What would RT data tell us that transcription data cannot. Better sensitivity?
- Study the effects of signal correlated noise
- Prosodic analysis of talkers.

Acknowledgments

Center for Cognitive Science
Graduate Research Fellowship

Dr. Keith Johnson
Dr. Mark Pitt
Dr. Mari Jones

Everyone who listened to me *kvetch*.

Thank You