

Perceptual Normalization for Reduced Speech: Phonetic Tuning vs. Spectral Contrast

Antonia D. Vitela, B. Sc.

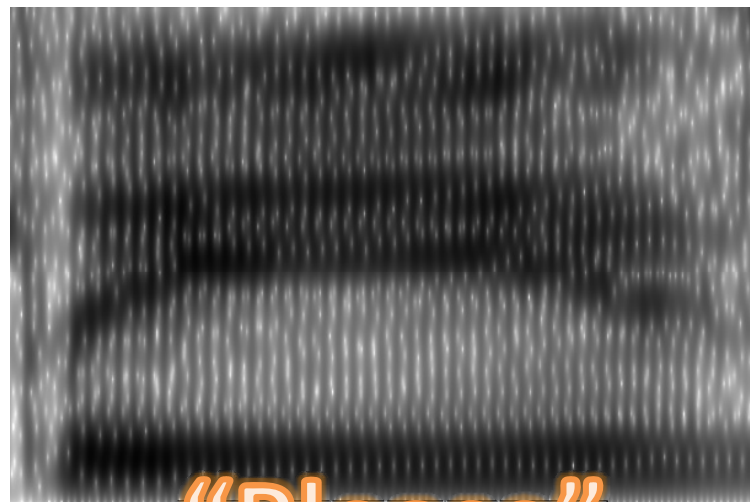
Andrew J. Lotto, Ph. D.

University of Arizona, Tucson, AZ



Reduced vs. Clear Speech

- Speaking styles are typically identified as being clear or reduced.
- According to Lindblom's H&H Theory, there is a continuum of production styles from clear (hyper) to reduced (hypo).
- These styles affect the acoustic realizations of phonemes.
- Even within a speaker, the formants corresponding to a particular vowel can change drastically from clear to reduced.



"Please"





Our Questions

- How do listeners accommodate for this variability and decide what the intended word is?
- Does the listener change what "counts" as a particular vowel based on the speaking style of the speaker?
- Does listeners' categorization of target vowels shift as a function of carrier phrase speaking style?



Three Possibilities

1) Expectations

- The listener hears reduced speech and then, changes their expectations to a “reduced speech mode”.

2) Mapping

- The listener maps out the speaker’s vowel space – this space is shrunk for reduced speech (Ladefoged and Broadbent, 1957).

3) Spectral Average

- The listener extracts the average of the spectrum and any phonetic decisions are made relative to that average (General Auditory Approach).
- All three predict that you should identify more centralized vowels as more peripheral vowels in the reduced speaking condition.
- However, the **General Auditory Approach** predicts that the shift could go in the opposite direction than what would be expected or could have no effect at all, depending on the average spectral content.



Experiment Set-up

- Phrases: Each produced by the same speaker in a clear and reduced style.

- Exp. 1 (Beat-Bit):

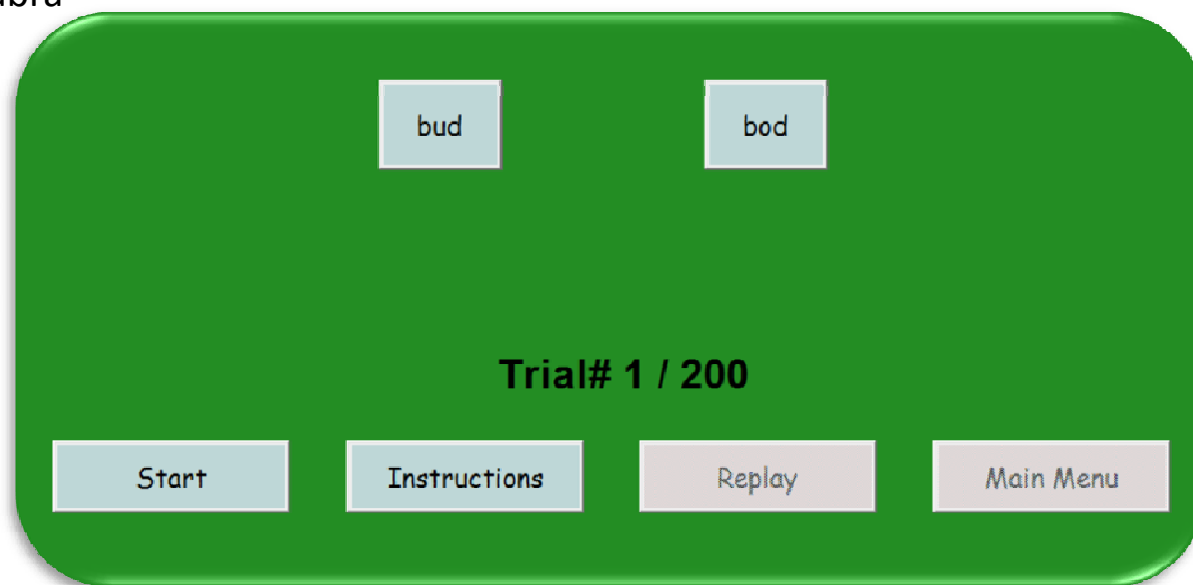
- 1) Please press the key that matches what I say.
- 2) Please would you choose the label most like the vowel I say.

- Exp. 2 (Bud-Bod):

- 1) Touch the button for what comes up.
- 2) Abracadabra

- Each phrase was attached to a target sound (either from the /i/ to /I/ series or the /[^]/ to /a/)

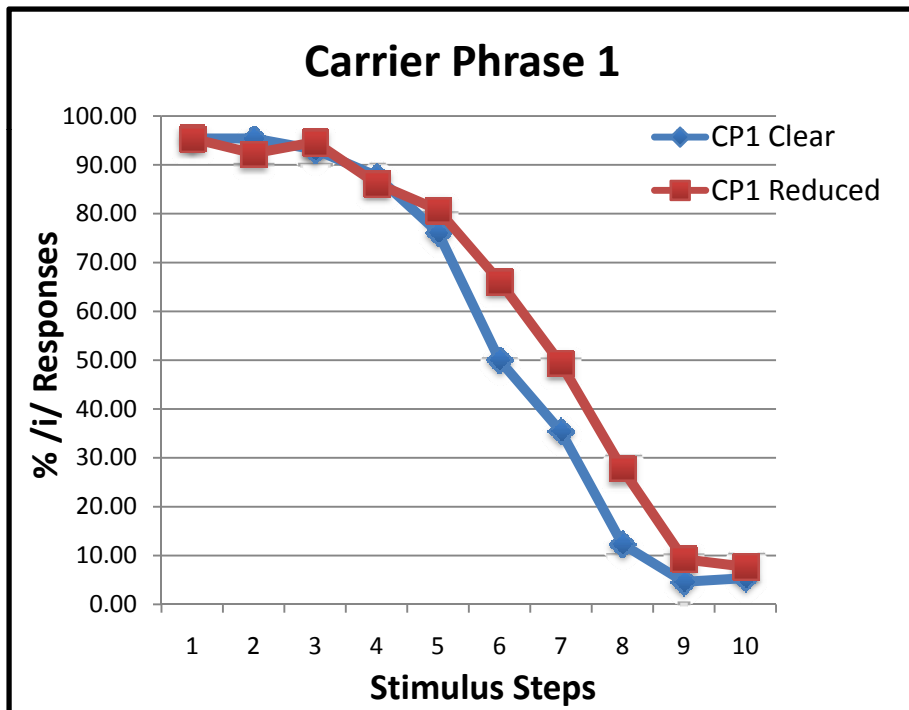
- 2 phrases x 2 speaking styles x 10 targets x 5 repetitions per block x 2 blocks = 400 trials.





Results: Experiment 1

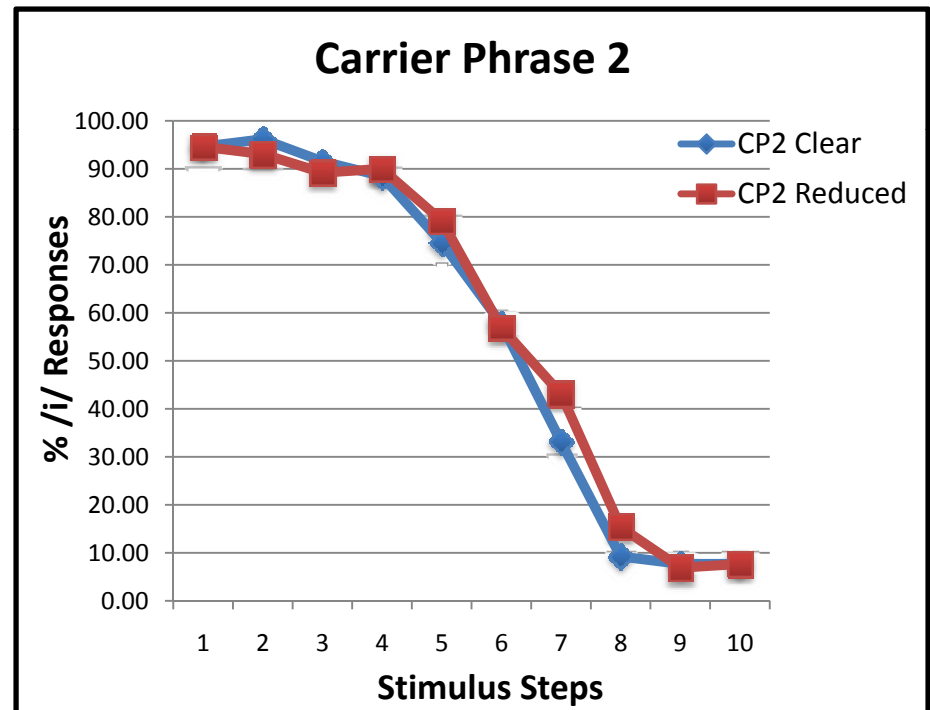
Beat-Bit: Percentage of /i/ Responses



$t(12) = -2.72, p < .05^*$

As predicted!

(more peripheral /i/ responses with reduced carrier)



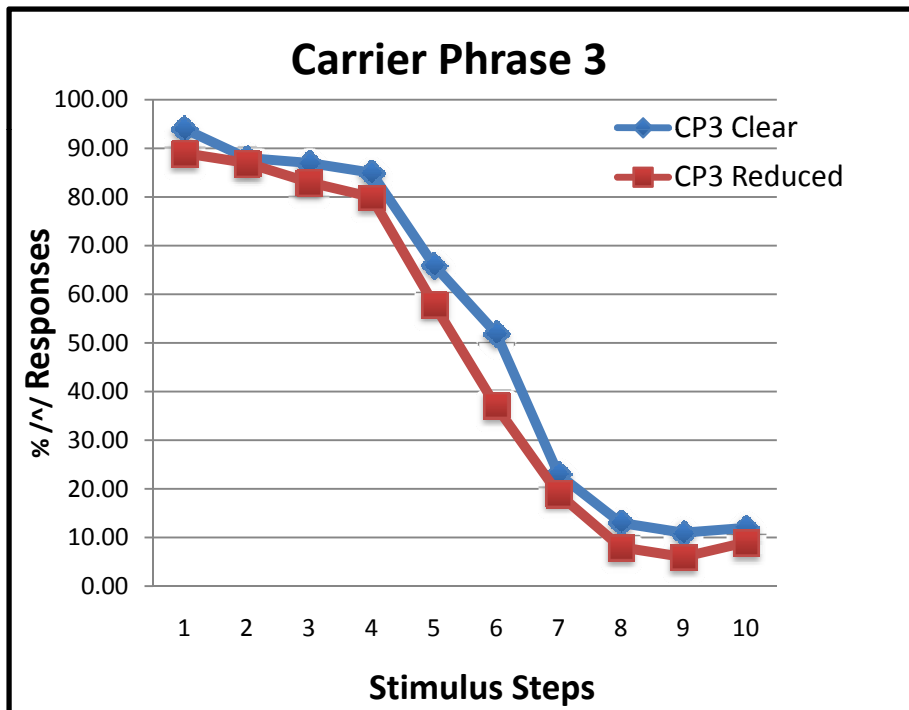
$t(12) = -.90, p > .05$

No effect...

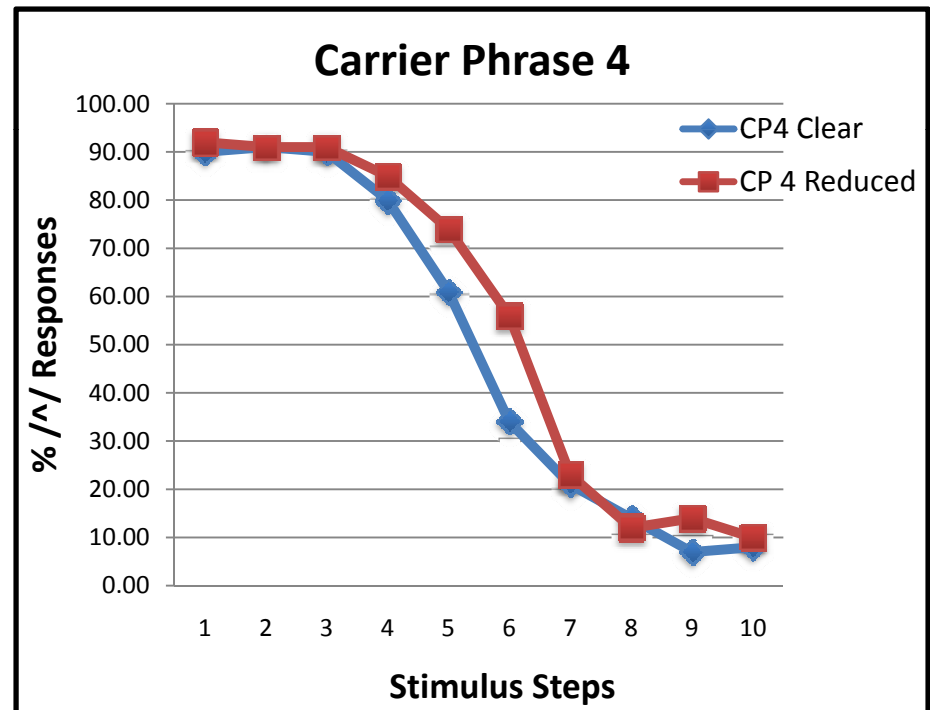


Results: Experiment 2

Bud-Bod: Percentage of /ʌ/ Responses



$t(9) = 2.58, p < .05^*$



$t(9) = -4.3, p < .005^*$

As predicted!

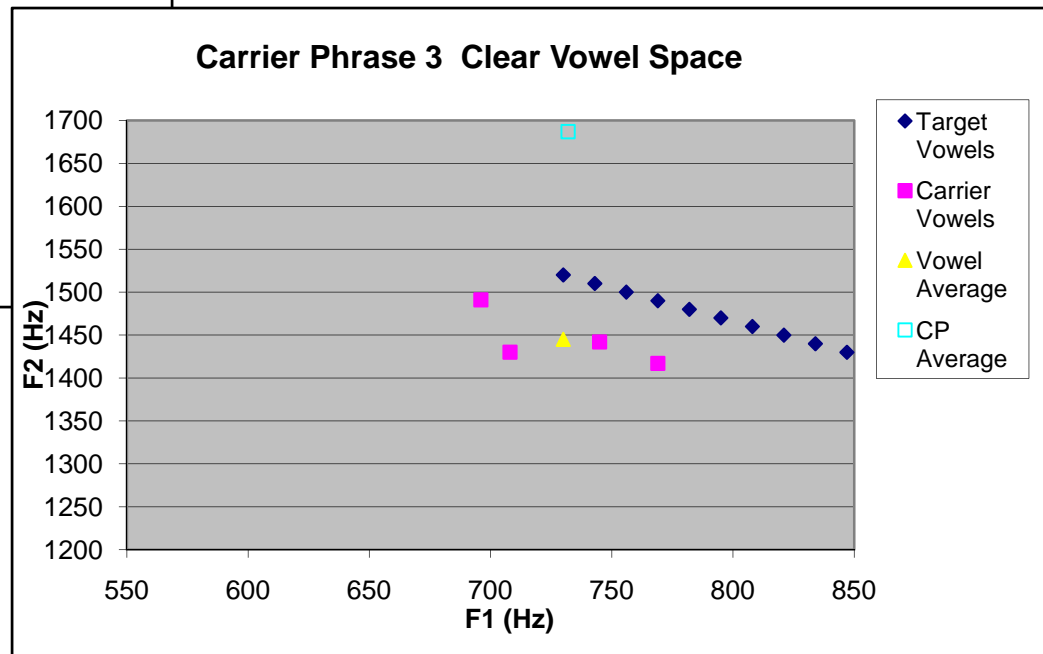
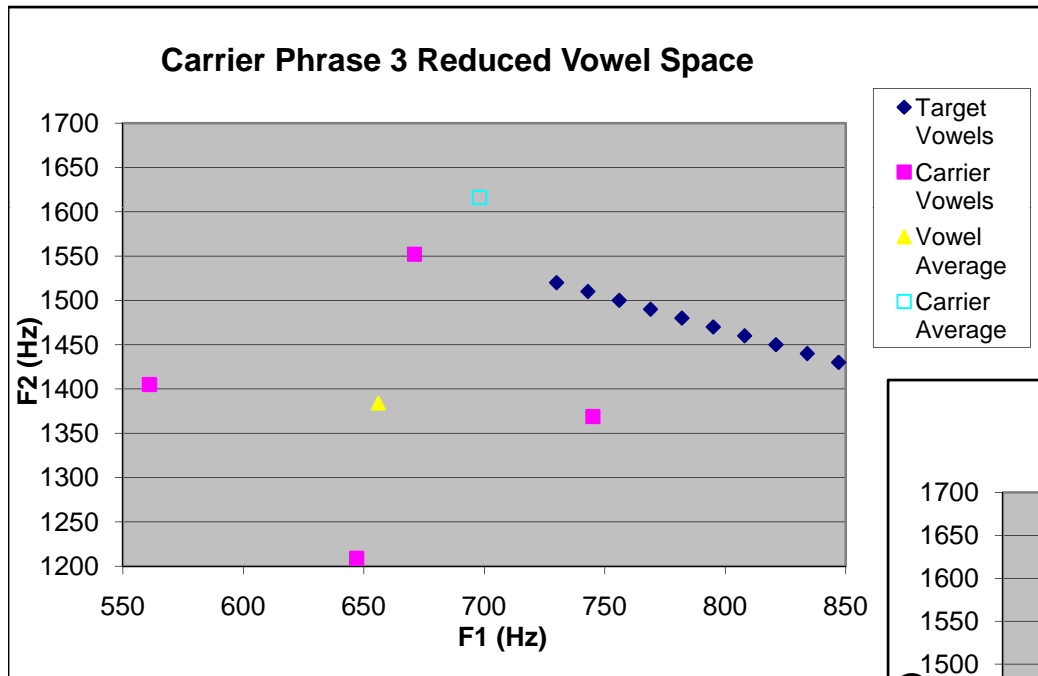
(more peripheral /a/ responses for reduced carrier)

Backwards...



Why do we get this shift?

Carrier 3: "Touch the button for what comes up ____."

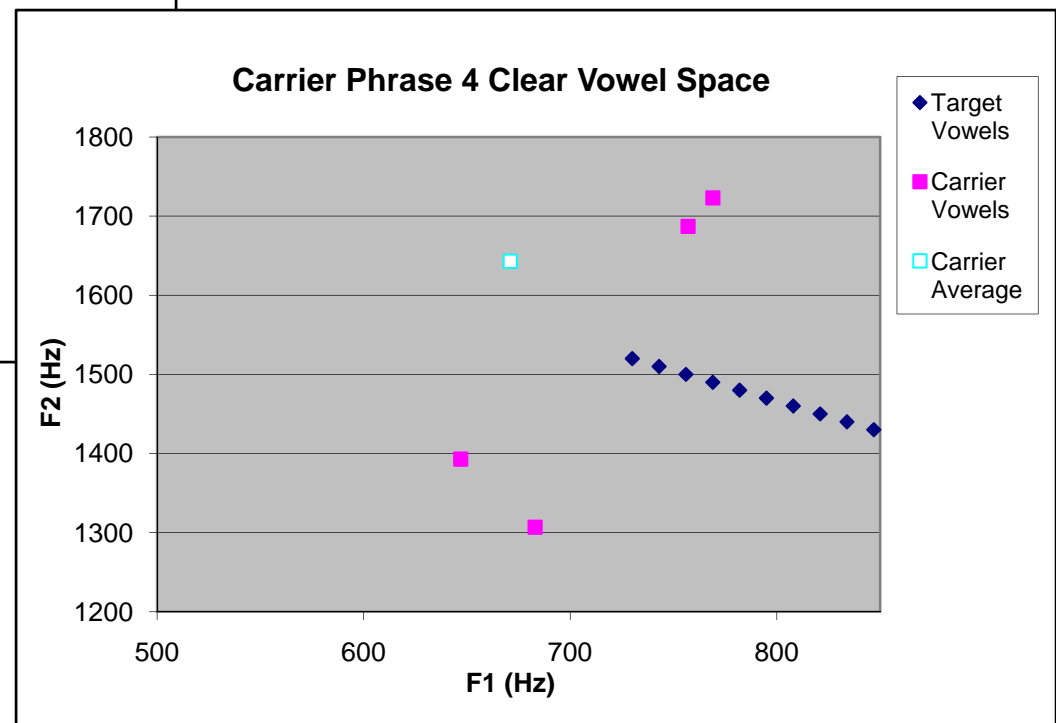
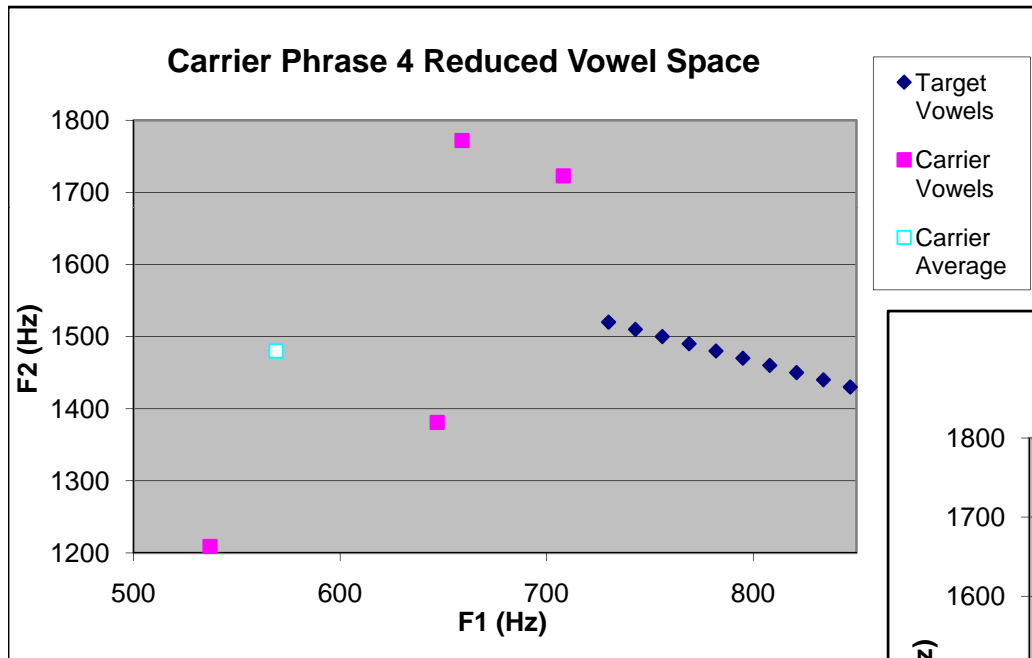


Target Vowels = synthesized vowels
Carrier Vowels = measured values for /[^]/ vowels in carrier phrase
Vowel Average = Average values for carrier vowels
Carrier Average = F1 & F2 averages estimated across all voiced portions of carrier



Why do we get this shift?

Carrier 4: "Abracadabra ____."



Target Vowels = synthesized vowels
Carrier Vowels = measured values for /[^]/ and /ae/ vowels in carrier phrase
Carrier Average = F1 & F2 averages estimated across all voiced portions of carrier



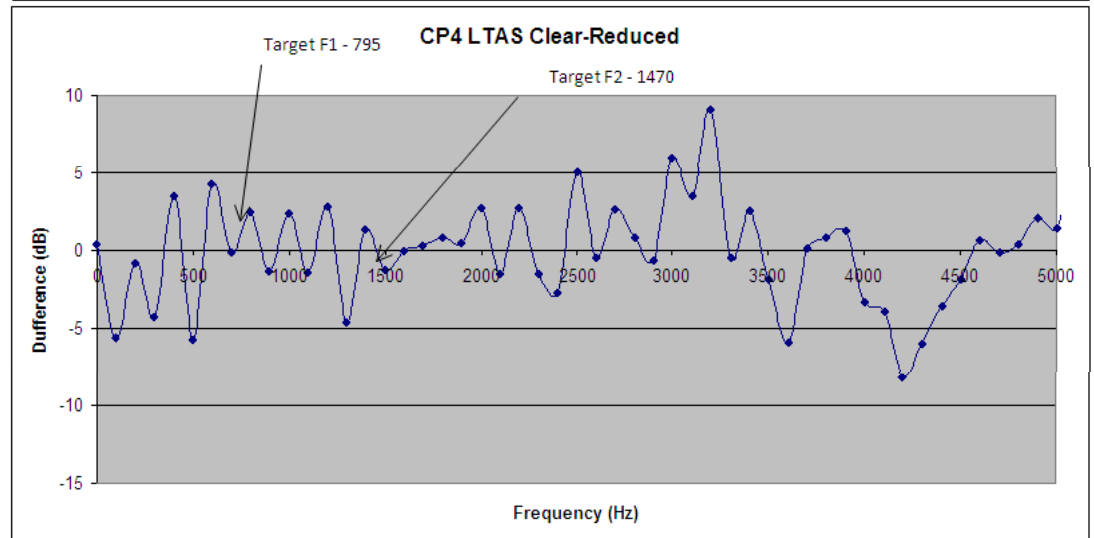
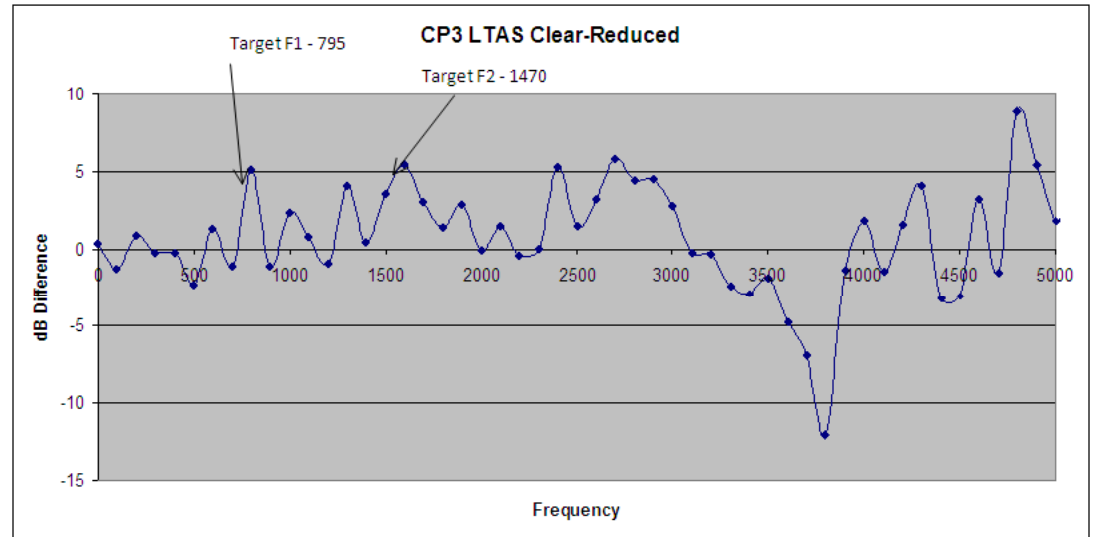
Discussion

- Categorization of target vowels WAS shifted as a function of the speaking style of the carrier phrase.
- Two of the 4 conditions went in the direction of more peripheral vowels for reduced speech as predicted by:
 - Expectations of reduced speech
 - Mapping (Ladefoged and Broadbent, 1957)
 - Spectral Contrast (General Auditory Approach)
- Two of the four did not show this shift with the largest overall shift actually backwards
 - Not predictable from a switch of "modes" since "clear"- "reduced" differences were obvious for each phrase set
 - Not predictable from vowel spaces
 - Can we predict the 4 shifts from the average spectra?



Spectral Differences

- Lotto & Sullivan (2007) and Holt (2006) suggest that the average spectra of the carrier phrases may affect target categorization in a contrastive manner.
- To the right are the DIFFERENCES between the average spectra of clear and reduced for carrier phrases 3 & 4 with formant values for an ambiguous target vowel.





References and Acknowledgments

- Holt, L. L. (2006). The mean matters: Effects of statistically-defined non-speech spectral distributions on speech categorization. *Journal of the Acoustical Society of America*.
- Ladefoged, P., & Broadbent, D. E. (1957). Information Conveyed by Vowels. *Journal of the Acoustical Society of America*, 29(1), 98-104.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H Theory. In W.J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (403-439). New York, NY: Springer.
- Lotto, A.J. & Sullivan, S.C. (2007). Speech as a sound source. In W.A. Yost, R.R. Fay, & A.N. Popper (Eds.), *Springer Handbook of Auditory Research: Auditory Perception of Sound Sources*.
- Work sponsored by grants from NIH-NIDCD

