4pSC19.

Investigation of the neural bases of context-dependent speech categorization.

Erika J.C. Laing[1], Lori L. Holt[1], & Anto Bagic[2]

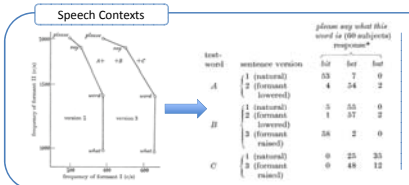
[1] Department of Psychology and Center for the Neural Basis of Cognition, Carnegie Mellon University;
[2] Department of Neurology, University of Pittsburgh Medical Center

Background

Objective: What are the neural mechanisms supporting context-dependent speech categorization?

Prior Research

Speech Contexts



Ladefoged & Broadbent (1957)

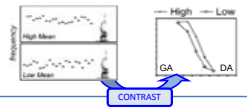
Preceding sentence contexts influence vowel perception.

NonSpeech Contexts

Holt (2005) (Holt & Kluender, 2000; Holt & Lotto, 2002; Holt et al., 2000; Huang & Holt, in press; Lotto & Kluender, 1998; Lotto et al. 2003; Stephens & Holt, 2003)

- Nonspeech tones sampling high vs. low mean frequencies shift consonant categorization.
- These frequency shifts mirror the manipulations Ladefoged & Broadbent (1957) made to introduce "talker" differences in synthesized speech, yet carry no speech information.

Is it possible that general auditory mechanisms play a role in talker normalization?



Perceptual shift appears to operate via a spectrally contrastive mechanism not specific to speech

The Current Study

Specific Aims
To replicate and extend Ladefoged and Broadbent (1957) to consonant contexts to examine the influence of sentence contexts vs. spectrally matched Nonspeech contexts on consonant categorization

LTAS Hypothesis (Long Term Average Spectrum)

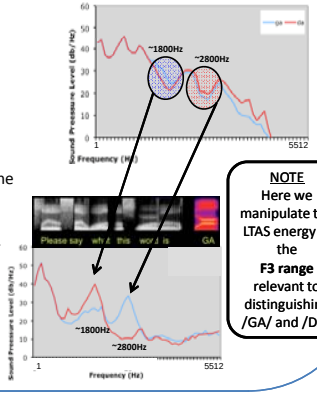
What is LTAS?

LTAS measures the energy present at each frequency in a given sound.

Hypothesis

- Results of the NonSpeech studies described above are consistent with a general auditory mechanism that tracks the long-term average spectrum (LTAS) of a sound and perceives subsequent sounds relative to, and contrastive with, the LTAS.
- This effect should only be relevant to frequency ranges particular to the speech targets being tested.

Observe: LTAS for /GA/ and /DA/ Targets



NOTE
Here we manipulate the LTAS energy in the **F3 range** relevant to distinguishing /GA/ and /DA/

Behavioral Task

Procedure

- Participants listened to either a Speech or a NonSpeech context, followed by a Speech Target, which was randomly selected from a continuum from /GA/ to /DA/, morphed on F3 onset. Participants were asked to categorize the speech target as either a /GA/ or a /DA/.
- Participants were tested with contexts for which the Long Term Average Spectrum (LTAS) was manipulated in the F3 frequency range (High and Low) to test whether LTAS peaks at critical frequencies alter perception.
- Participants were also tested with contexts for which LTAS was manipulated in the F1 range (High and Low), with the prediction that this manipulation will not elicit perceptual shifts.

Stimuli

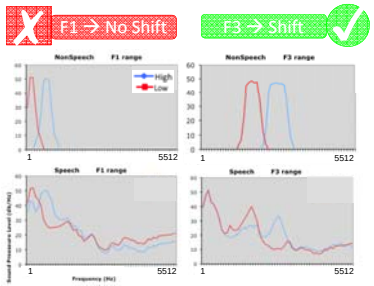
• LTAS peaks	for both speech and nonspeech	
F1 range	Low	300 Hz
	High	800 Hz
F3 range	Low	1800 Hz
	High	2800 Hz

• Speech context was synthesized using Klatt

Predictions

Long Term Average Spectrum (LTAS)

- The F1 vs. F3 Range manipulation will test whether LTAS differences are specific to the region of the spectrum significant for categorization (F3) and also whether perceived talker differences across contexts are sufficient to influence target categorization.
- Both F1 Range and F3 Range LTAS manipulations produce a perceived change in talker.

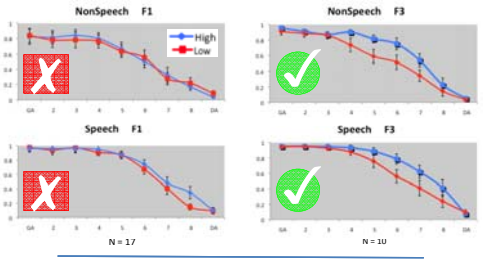


• Speech target series created by manipulating natural /ga/-/da/ (Holt, 2005)

• 50ms gap between context and speech target

Results

LTAS is sufficient to shift consonant perception, but only in the frequency range relevant to the speech target categorization, without regard to whether the context was Speech or NonSpeech.



	Same Speaker	Different Speaker
F1 Range	96%	97%
F3 Range	99%	93%

In a second task, participants easily discriminated the 'voices' of the Speech stimuli as coming from different talkers.

MEG Task

Can we use MEG neuroimaging to study context-dependent speech categorization?

Procedure and Stimuli

- F3 range context stimuli only
- Other parameters identical

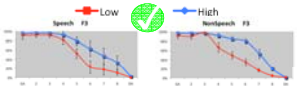
Analysis Procedures

- Acquisition Device
  - Elekta Neuromag, 306 sensor Vectorview
  - Sampling Rate = 1000 kHz
- Spatial Filtering
  - Signal Space Separation (SSS)
- Frequency Filtering
  - Bandpass: 0.5 - 40 Hz
- Rejection Parameters
  - Gradiometers: 3000 fT/cm
  - EOG: 150 µV
- Averaging
  - Time-locked to presentation of speech target

Preliminary Results

N = 6

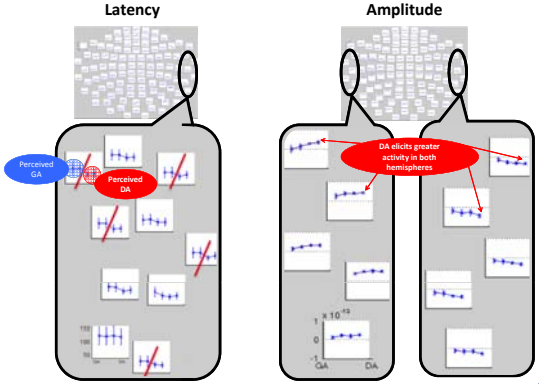
Behavioral



- Initial analysis of M100 signal (peak 50-150ms), collapsed across all conditions and grouped according to how the participant categorized the target, as judged from behavioral responses.
- Latency of M100 reveals a distinction between targets perceived as /GA/ vs. /DA/, even for the middle physically ambiguous stimuli. /GA/ M100 comes online a bit later than /DA/ in the right temporal magnetometer sensors.
- Amplitude of the M100 signal is higher for /DA/ than for /GA/, in both hemispheres.

Preliminary analysis suggests that the M100 may be able to be used as a neural marker for perceived /DA/ and /GA/ speech sounds.

MEG -- M100 Signal



Conclusions

- **Extension of the classic Ladefoged & Broadbent (1957) results to consonants.**
- **Long Term Average Spectrum (LTAS), and not perceived talker differences, predicted the results.**
- **Speech categorization is strongly influenced by shifts in the LTAS of a context stimulus when the affected frequencies are relevant to target categorization.**
- **Preliminary MEG results suggest the M100 as a potential neural marker of /GA/-/DA/ perception. Future analyses are planned to observe how this marker is modulated by context.**

References and Acknowledgements

Broadbent, D. E., & Ladefoged, P. (1960). Vowel judgments and adaptation level. Proceedings of the Royal Society of London, Series B: Biological Sciences, 151.

Holt, L. L. & Kluender, K. R. (2000). General auditory processes contribute to perceptual accommodation of coarticulation. Phonetica, 57, 170-180. Invited contribution to a special issue of Phonetica on Speech Communication and Language Development.

Holt, L. L. (2005). Temporally non-adjacent non-linguistic sounds affect speech categorization. Psychological Science, 16, 305-312.

Holt, L. L., & Lotto, A. J. (2002). Behavioral examinations of the neural mechanisms of speech context effects. Hearing Research, 167, 156-169.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. Journal of the Acoustical Society of America, 29, 98-104.

Lotto, A. J., Kluender, K. R., & Holt, L. L (1998). Depolarizing the perceptual magnet effect. Journal of the Acoustical Society of America, 103, 3648-3655

Lotto, A. J., Sullivan, S. C., & Holt, L. L. (2003). Central locus for nonspeech context effects on phonetic identification. Journal of the Acoustical Society of America, 113, 53-56.

Stephens, J. D. W., & Holt, L. L. (2003). Preceding phonetic context affects perception of nonspeech. Journal of the Acoustical Society of America, 114, 3036-3039.

Work supported by National Institute of Health (NIH) and The National Organization for Hearing Research Foundation (NOHR)