

Non-linguistic sentence-length precursors affect speech perception: Implications for speaker and rate normalization

Lori L. Holt & Travis Wade

Department of Psychology & Center for the Neural Basis of Cognition
Carnegie Mellon University



INTRODUCTION

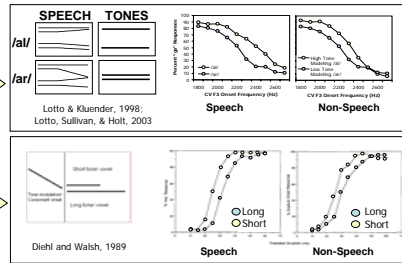
Spectral and temporal properties of speech sounds are perceived in a thoroughly context-dependent manner at both local and more protracted levels of the speech signal.

- Immediate phonetic context influences categorization of sounds differentiated by formant frequencies (Mann, 1980).
- Longer-term, speaker-specific spectral patterns also influence speech categorization (Ladefoged and Broadbent, 1957).
- Likewise, categorization of sounds based on temporal information is affected by both local and longer-term speaking rate (Miller and Liberman, 1979; Summerfield, 1981).

Accounts of these effects often invoke speech-specific normalization processes.

- These processes presumably mitigate the effects of acoustic variability introduced by coarticulation and cross-talker differences.

At least the more local of these effects might originate from general perceptual processes:



The temporal adjacency of the context-providing sounds in the experiments like those illustrated allows for the possibility that these interactions arise from well-understood local interactions in neural processing such as neural adaptation (Delgutte, 1996).

Thus, it is possible that higher-order linguistic processes or speech-perception-specific processes underlie longer-term (sentence level) speech context effects whereas low-level auditory perceptual interactions influence how local contexts affect speech categorization.

EXPERIMENT AIMS

- Investigate whether the influence of non-linguistic sounds upon speech categorization is limited to local interaction in auditory processing.
- Further, test whether the auditory system is sensitive to the statistical structure of spectral and temporal distributions of energy over time, and whether such sensitivity may sway listeners' perception of following speech.
- Observation of a context effect under these circumstances requires that linguistic and non-linguistic acoustic information interact at higher levels of auditory processing than has previously been demonstrated.
- In the present studies, sentence-length pure tone sequences serve as both spectral and temporal perceptual contexts for following speech sounds.

REFERENCES

Delgutte, B. (1996). Auditory neural processing of speech. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 505-538). Oxford: Blackwell.

Diehl, R. & Walsh, M. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustic Society of America*, 85, 2154-64.

Holt, L. L., & Lotto, A. (2002). Behavioral examinations of the level of auditory processing of speech context effects. *Hearing Research*, 167, 156-169.

Ladefoged, P. and Broadbent, D. (1957). Information conveyed by vowels. *Journal of the Acoustic Society of America*, 29, 98-104.

Lotto, A. J., & Klaunder, K. R. (1998). General contrast effects of speech perception: Effect of preceding liquid on stop consonant identification. *Perception and Psychophysics*, 60, 602-619.

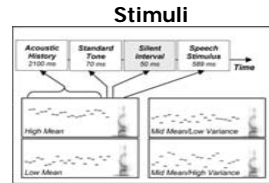
Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 28, 407-412.

Miller, J., & Liberman, A. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception and Psychophysics*, 25, 457-65.

Miller, J., O'Rourke, T., & Volaitis, L. (1997). Internal structure of phonetic categories: effects of speaking rate. *Phonetica*, 54, 121-37.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 1074-95.

SPECTRAL EXPERIMENTS

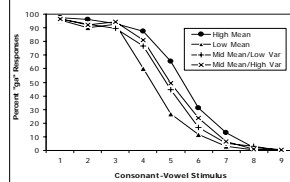


- **SPEECH TARGET STIMULUS:** Nine-step /ga/ to /da/ series created from natural productions.
- **STANDARD TONE:** 70-ms 2300-Hz sine-wave tone. Speech categorization of /ga/ to /da/ stimuli in isolation did not significantly differ from categorization of the syllables preceded by the standard tone ($t_{(9)}=1.35, p=.21$).
- **ACOUSTIC HISTORIES:** 21 70-ms sine-wave tones (30-ms silent intervals) with unique frequencies. Tone order was randomized on a trial-by-trial basis.
 - **HIGH MEAN:** 2300-3300 Hz in 50-Hz steps
 - **LOW MEAN:** 1300-2300 Hz in 50-Hz steps
 - **MID MEAN:** 1800-2800 Hz in 50-Hz steps
 - **MID MEAN/HIGH VAR:** 1300-3300 in 100-Hz steps

Procedure

Stimuli were mixed across conditions. Listeners categorized the speech targets as /ga/ or /da/.

Experiment 1: Do Acoustic Histories Influence Speech Categorization?



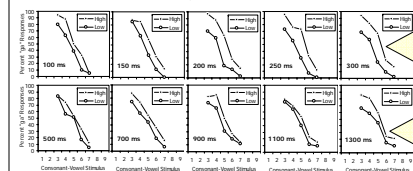
Results

Although the surface acoustic characteristics of the Acoustic Histories varied on a trial-by-trial basis and the final Standard Tone of each Acoustic History was constant across conditions ... they exerted a significant influence on speech categorization, $F_{(3,9)}=29.3, p<.001$. This influence was **spectrally contrastive**. Acoustic histories drawn from the distribution with a **HIGH** mean frequency led listeners to more often categorize following speech targets as /ga/, the alternative with a **LOWER**-frequency F3 onset.

Since these melodies were distinct, ended at the same frequency level, and comprised wide, overlapping frequency and duration ranges, any influence they showed upon categorization of following speech was expected to demonstrate listeners' sensitivity to their long-term spectral or temporal distribution and not merely to the simple acoustic characteristics of any particular segment.

Experiments 2a-b: What is the Time Course of the Influence of Acoustic Histories?

The silent interval separating the Standard Tone and the Speech Target was manipulated in two experiments.



Effect of Acoustic History, $p<.001$
No Effect of Silent Interval, $F<1$
Acoustic History x Silent Interval Interaction, $p<.001$

Context Effect Gets Larger with Longer Silent Intervals

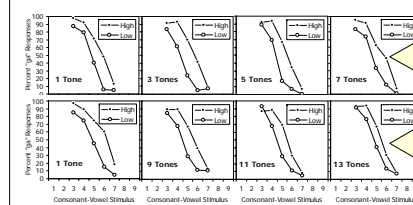
Effect of Acoustic History, $p<.001$
No Effect of Silent Interval, $F<1$
Acoustic History x Silent Interval Interaction, $p=.14$

Equivalent Across Conditions

There is a significant effect of Acoustic History Mean... even when 1300 ms of silence separated tone sequences and speech targets. This is significantly longer than has been observed previously.

Experiments 3a-b: What are the Boundaries of Temporal Adjacency?

The Standard Tone was repeated 1-13 times across two experiments. This final context was constant across conditions, so any effect of context would need to persist across intervening acoustic stimuli.



Effect of Acoustic History, $p<.0001$
Effect of Repetition, $p=.01$
No Acoustic History x Repetition Interaction, $F<1$

Context Effect was Equivalent Across Conditions

Effect of Acoustic History, $p<.0001$
Effect of Repetition, $p=.01$
No Acoustic History x Repetition Interaction, $p=.12$

Context Effect was Equivalent Across Conditions

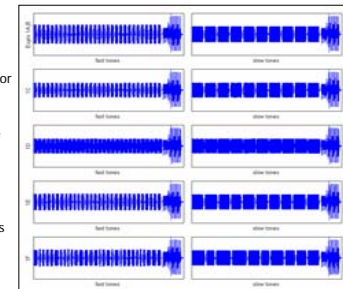
The influence of Acoustic Histories persists even when 1300 ms of intervening constant acoustic stimulation separates Acoustic Histories and Speech Targets.

TEMPORAL EXPERIMENTS

Stimuli

- **SPEECH TARGET STIMULUS:** Eleven-step synthesized /ba/ to /wa/ series varying only in F1, F2 transition duration.
- **ACOUSTIC HISTORIES:** 1.2-s pure tone precursors sequences with 30 fast or 10 slow tones.
 - **FAST MEAN:** 40ms tone-to-tone: 1 continuum member shorter than most /ba/-like initial formant pattern
 - **SLOW MEAN:** 120ms tone-to-tone: 1 continuum member longer than most /wa/-like initial formant pattern

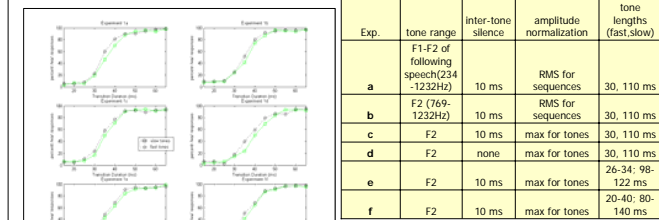
Tone order was randomized on a trial-by-trial basis



Procedure

Stimuli were mixed across conditions. Listeners categorized the speech targets as /ba/ or /wa/ (Experiment 1) or rated their goodness as /wa/ (Experiment 2).

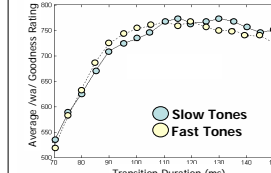
Experiments 1a-f: Does the Rate of Acoustic Histories Influence Categorization?



Effect of tone rate on speech categorization for all six conditions, $p<.025$
No effects or interactions involving tone manipulations

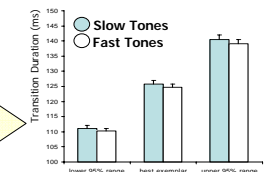
The rate of sentence-length non-speech precursors influences speech categorization. This influence was **temporally contrastive** such that **FASTER** acoustic histories led listeners to more often categorize following speech targets as /wa/, the alternative with a **SLOWER** formant-frequency transition.

Experiment 2: Does the Context Affect Internal Category Structure?



Target: /ba/ - /wa/ continuum extended from Exp. 1 to include exaggerated /w/ transitions
Context: pure tone precursors similar to Exp. 1c

Quadratic functions fit to response curves for stimuli with 50% overall /wa/ rating
Precursor rate effect observed for location of maximum estimated /wa/ rating ($F_{(1,26)}=4.40, p=0.046$), lower ($F_{(1,26)}=4.32, p=0.048$) and upper limits of 95% best exemplar range: ($F_{(1,26)}=4.31, p=0.048$).



The rate of the Acoustic Histories significantly influenced internal category structure.