32. PERCEPTUAL INSERTION OF PHONETIC SEGMENT INDUCED BY SPEECH-SHAPED NOISE

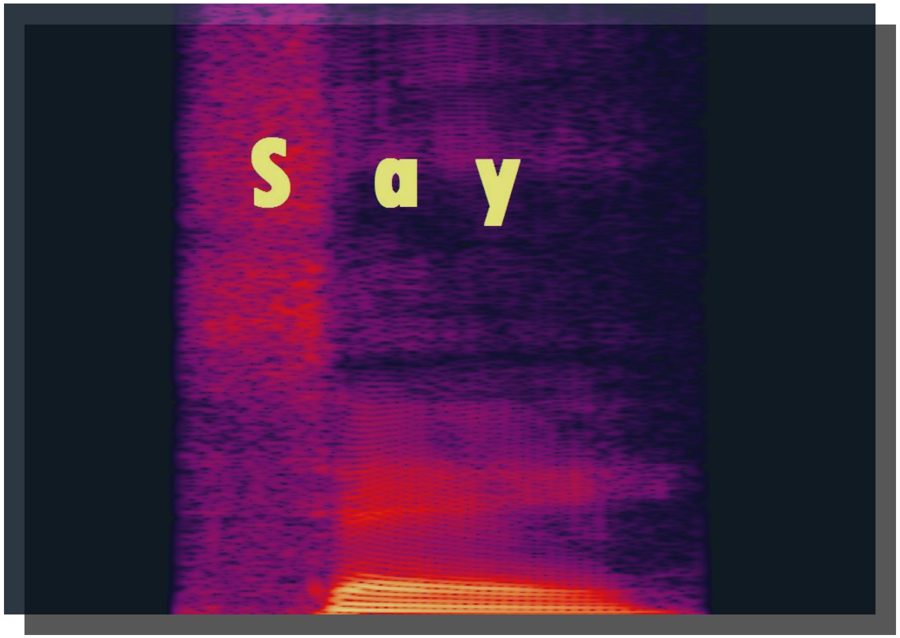
Andrew Lotto, Ph.D, Antonia Vitela, B.Sc., and Sarah Sullivan, M.A.
University of Arizona, Tucson, AZ

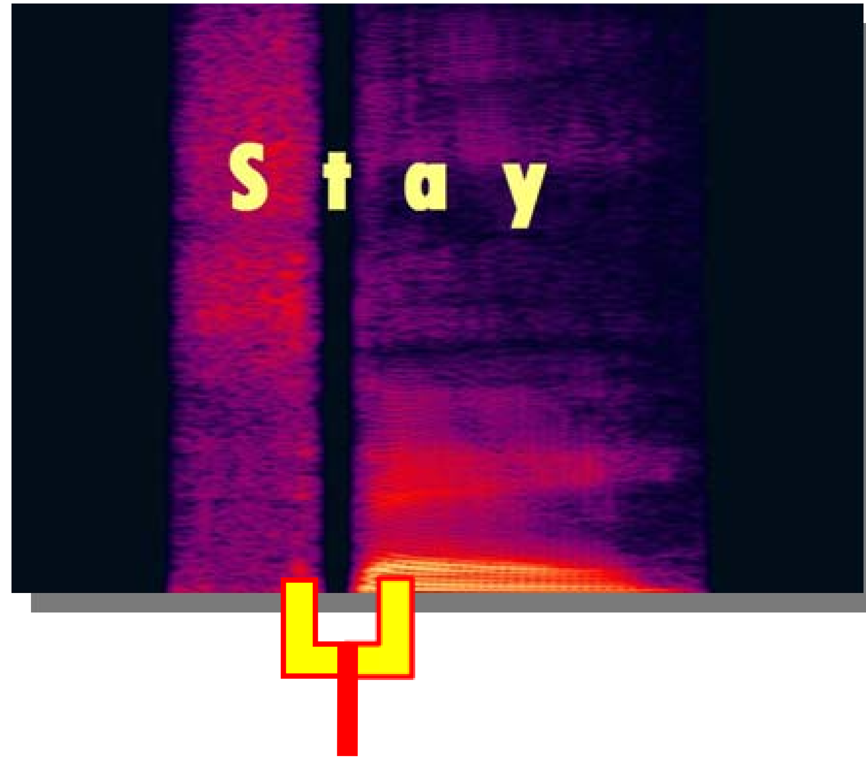
Lori Leibold, Ph.D.
University of North Carolina, Chapel Hill, NC

Abstract

A stimulus series varying perceptually from "say" to "stay" was created by inserting a silent gap between the initial fricative and the vowel of "say" (gap durations varied from 0 to 75 ms in 5-ms steps). Adult listeners with normal hearing were presented members of this series in quiet and 0 dB S/N and asked to label the sounds as "say" or "stay". One might predict that noise would mask the silent gap and result in more "say" responses. In fact, the opposite result was obtained. More "stay" responses were obtained in noise. Response functions in noise maintained steep slopes suggesting that the shift in responses was not due to a decrement in performance or guessing. This result demonstrates that degraded listening conditions do not just result in lost or distorted phonetic segments but can also lead to perceptual insertion of segments. From a strict audibility perspective this is surprising, but not from the perspective of complex pattern recognition processes.

“Say” Versus “Stay”

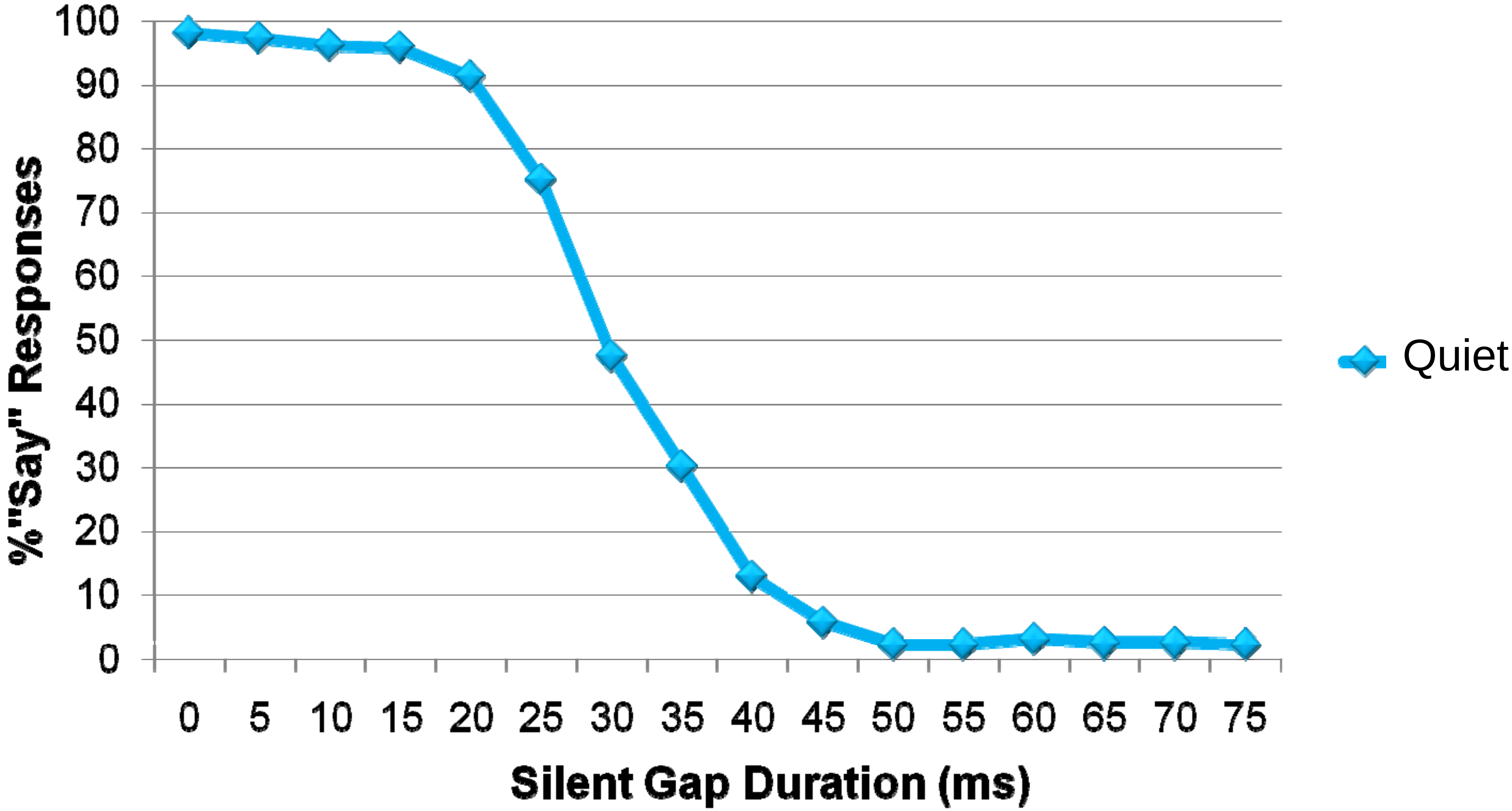




Distinguished acoustically by duration of silent gap

16 stimuli were created by manipulating the duration of the silent gap from 0 to 75 ms in 5-ms steps.

%“Say” Response in Quiet



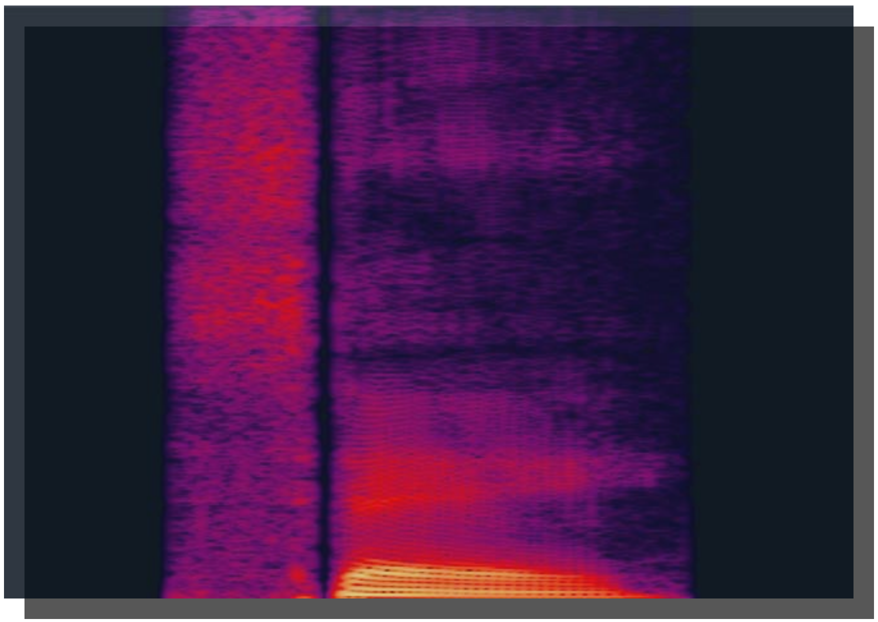
“Say-Stay” as Gap Detection

- This “say-stay” series is analogous to a gap detection experiment if you presume that listeners respond “stay” when they detect any gap or a gap discrimination experiment if you presume that listeners respond “stay” when they perceive a gap of a particular duration.

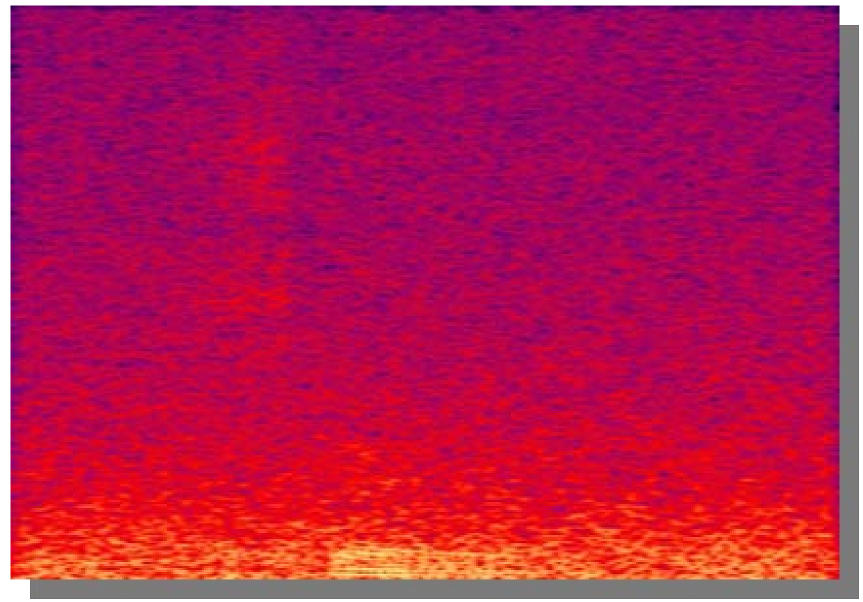
- If the syllables were presented in noise, one may predict that the ability to detect or discriminate the gap will be compromised leading to more “Say” responses.

Quiet vs. Noise

Quiet



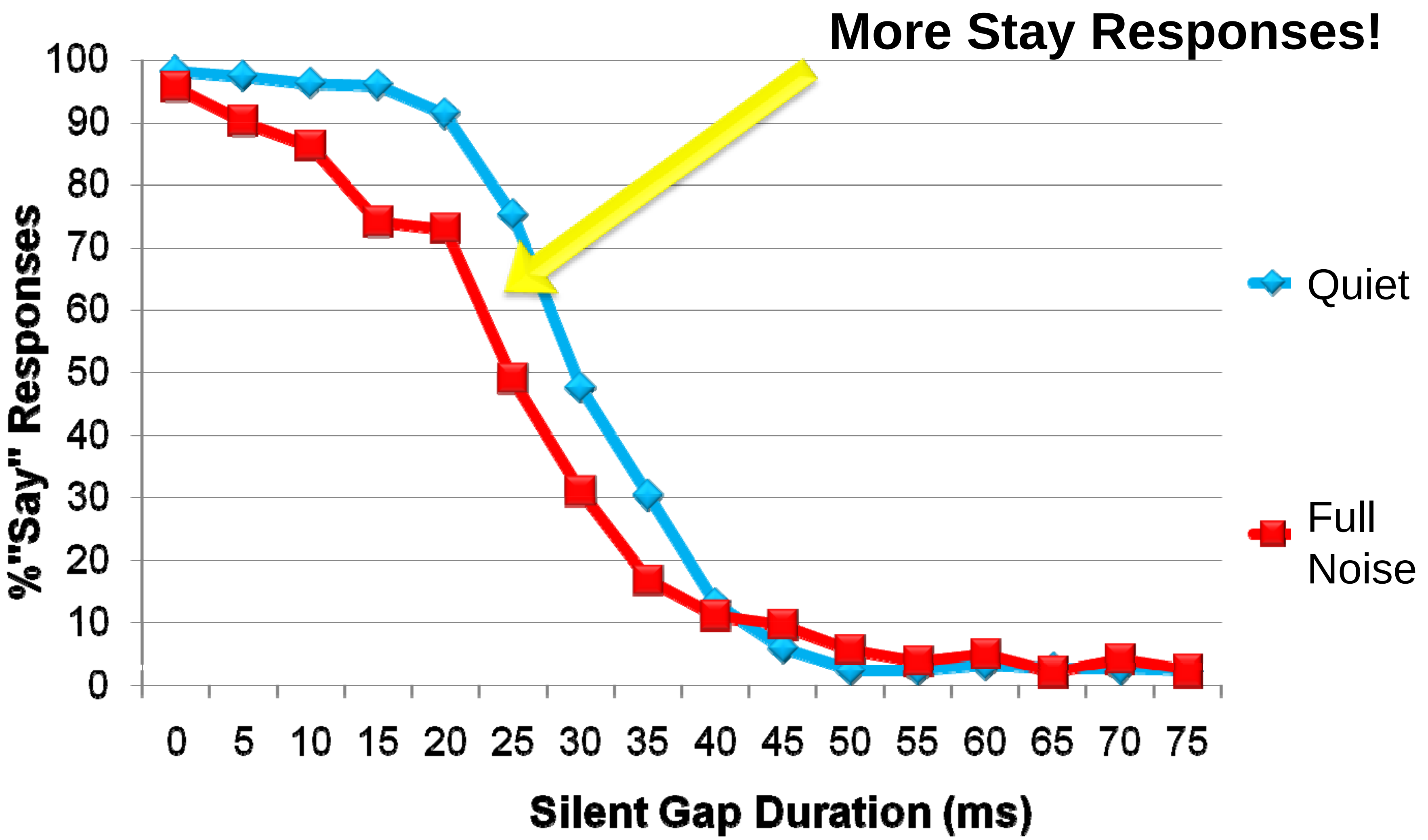
Speech-Shaped Noise



Mask detection of gap → **Predict** → More "Say"

Ambiguous "say-stay" stimulus embedded in quiet versus speech-shaped noise (0 dB S/N).

% “Say” Response in Noise and Quiet

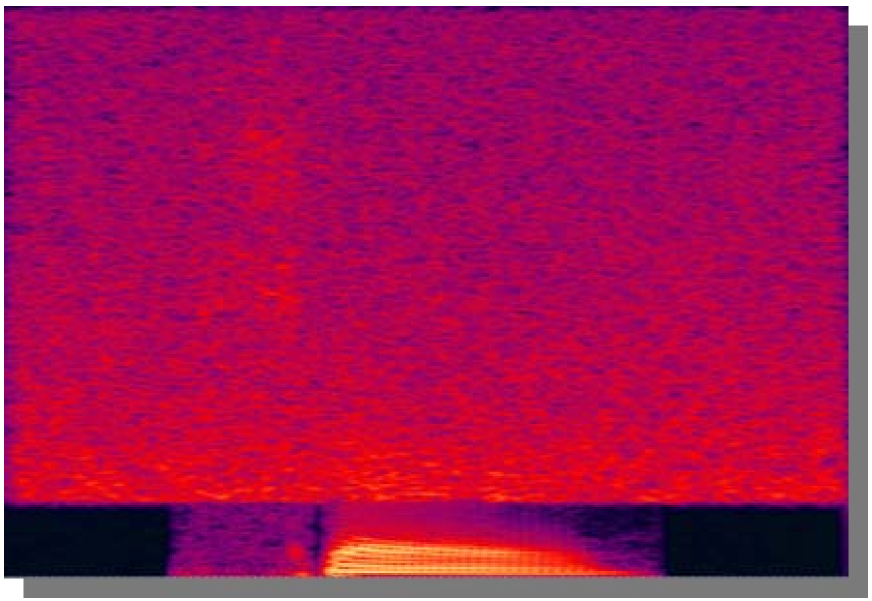


Why more “stay” responses?

- Two potential cues for the “say-stay” distinction are the silent gap and the formant transitions (Best, Morrongiello, & Robson, 1981).
- One possible explanation is that the noise changes the perception of the frequency slope or amplitude slope of those transitions.
- To test this hypothesis, four more conditions were created that presented noise in different frequency regions of the spectrum. If the noise is affecting perception of the transitions than one would predict different results for conditions in which the formants are presented in noise (the high pass conditions) from those conditions in which the formants are “uncovered” (the low pass conditions).

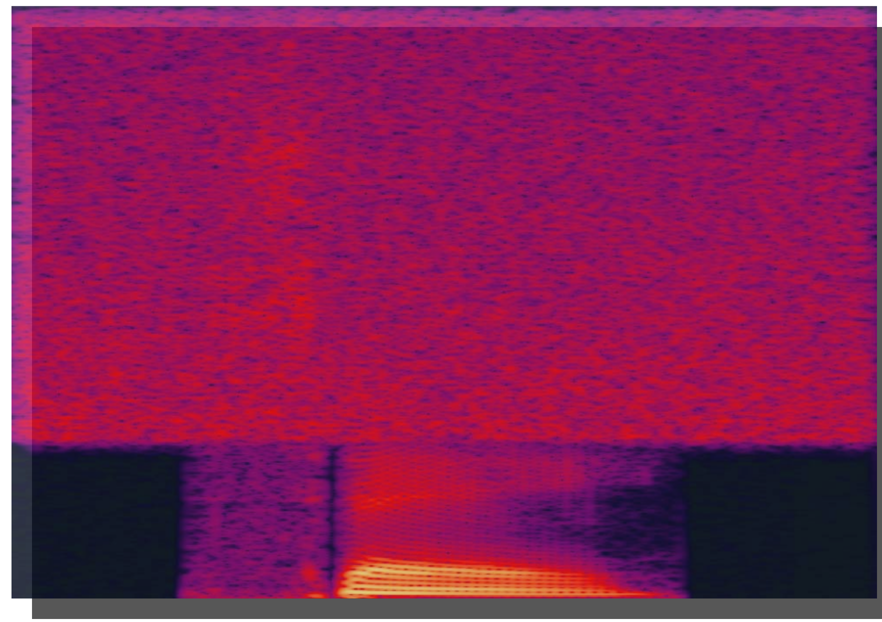
High Pass Stimuli

HP 1500



High pass filtered at 1500 Hz
Formant 1 is uncovered.

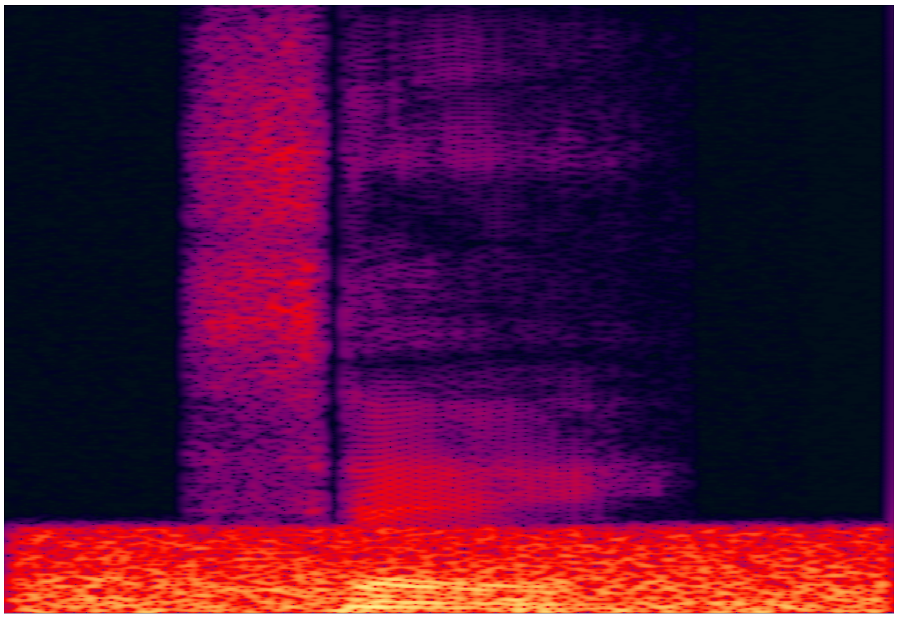
HP 3000



High pass filtered at 3000 Hz
Formants 1, 2, and 3 are uncovered.

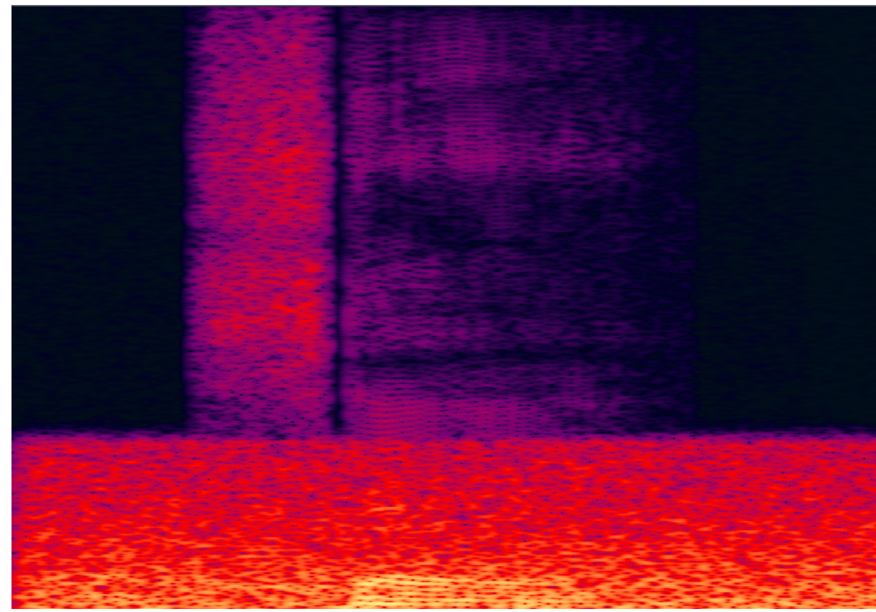
Low Pass Stimuli

LP 1500



Low pass filtered at 1500 Hz
Formant 1 is presented in noise.

LP 3000



Low pass filtered at 3000 Hz
Formants 1, 2, and 3 are presented in noise.

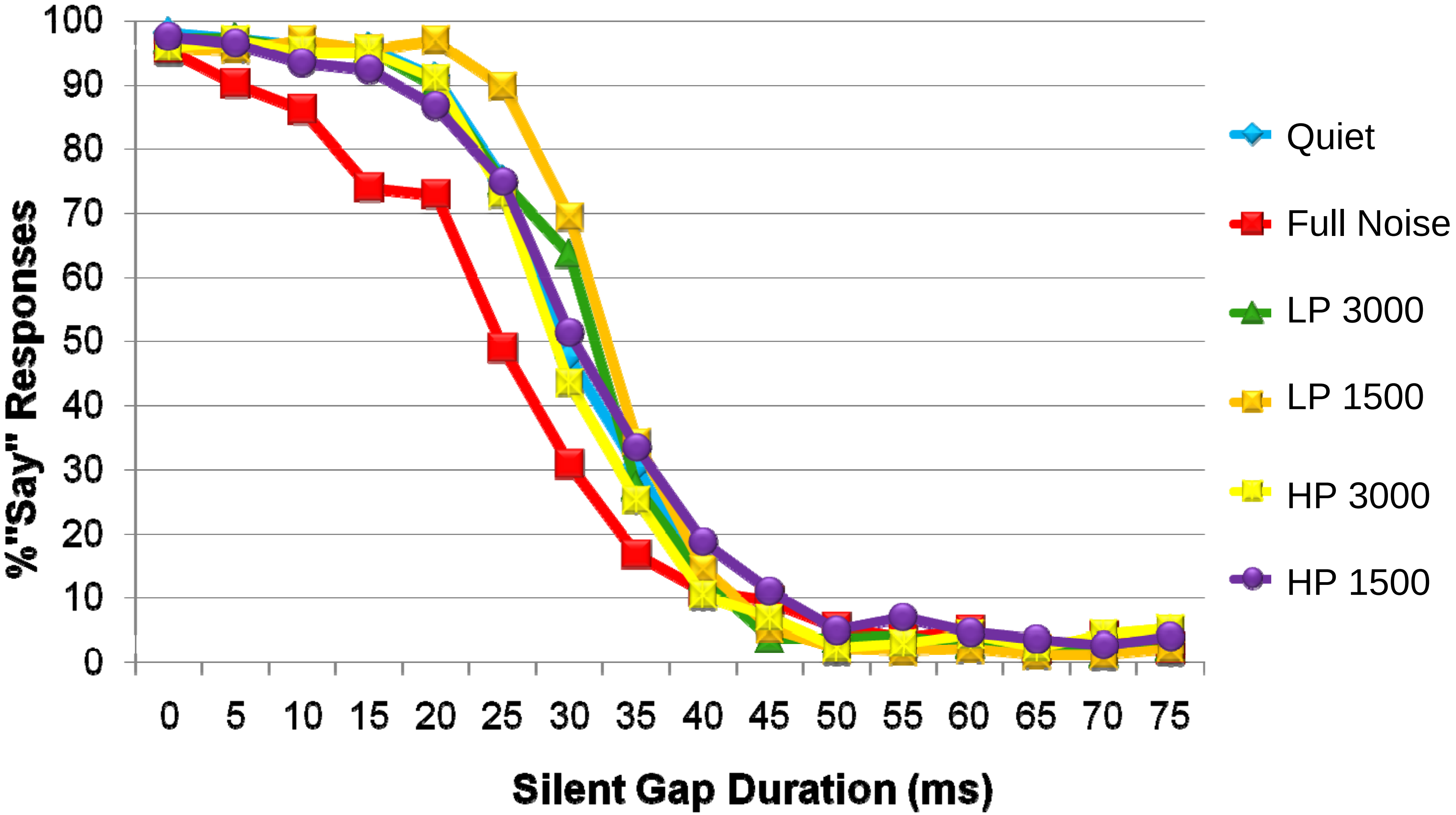
Experiment Set-up

Condition	Speech-Shaped Noise
1	No noise
2	0-10,000 Hz
3	0-3,000 Hz
4	0-1,500 Hz
5	3,000-10,000 Hz
6	1,500-10,000 Hz

12 presentations of stimuli from each conditions were presented randomly to listeners for a total of 1152 trials (16 stimuli x 12 presentations x 6 conditions). The entire experiment took ~50 minutes.

The stimuli were presented binaurally over headphones to 29 normal-hearing adults in a quiet room. Stimulus presentation and data collections using ALVIN (Hillenbrand & Gayvert 2003).

%"Say" Responses All Conditions



Results

“Say” responses (collapsed across gap duration):

Significant effect of Condition (6 levels) [p<.001]

Post-hoc comparisons (Bonferroni-adjusted):

- Full Noise Condition differs from all other conditions
- All Filtered Conditions equivalent to Quiet

Boundary values (Probit analysis):

Significant effect of Condition (6 levels) [p<.001]

Post-hoc comparisons (Bonferroni-adjusted):

- Full Noise Condition differs from all other conditions
- All Filtered Conditions equivalent to Quiet

Slopes (Probit analysis):

Significant effect of Condition (6 levels) [p<.001]

Post-hoc comparisons (Bonferroni-adjusted):

- Full Noise shallower than Quiet
- All filtered conditions have slopes between Quiet and Full Noise
  (differ from both except LP1500 similar to Quiet)

Pieces of the puzzle…

1. Full-spectrum speech-shaped noise leads to more "Stay" responses.
2. All filtered noise conditions look like quiet
3. Noise condition slopes are shallower, but still relatively steep
4. Children (5-10 yrs) show the opposite shift in response - more "say" in noise (see Leibold and Lotto presentation on Saturday morning)
5. Preliminary evidence that white noise leads to opposite shift in adults – more "say" responses (though slopes are much shallower suggesting that it may be due to guessing).

Possible Explanations

- Instead of presuming that “Say” is the default and evidence of a gap must be present to hear “Stay”, perhaps “Stay” is the default (due to frequencies in the input?) and a lack of gap must be detected to hear “say”. If noise is interfering with precise perception of gap, then one would expect more of the default “stay” response. Perhaps children have not yet learned this default.
- Adult listeners rely on the gap information whenever there is reliable evidence for it any frequency region, but when the entire spectrum is degraded they shift some “weighting” to formant based cues that result in more “Stay” responses.
- Either of these explanations suggests that one cannot predict speech categorization by looking simply at the audibility of a dominant acoustic cue. Listeners are highly adaptive and rely on many sources of information. In order to meaningfully predict speech perception performance, we must account for these adaptive processes and rely on more complex models of pattern recognition than provided by an acoustic-feature detection account.

References and Acknowledgements

- Best, C.T., Morrongiello, B., and Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. Perception and Psychophysics, 29(3), 191-211.
- Hillenbrand, J. M. & Gayvert, R. T. (2003). Open-source software for experiment design and control. JSLHR, 48, 45-60.
- Leibold,L. J., & Lotto, A. J. (2008). Effects of noise on children's speech identification. AAS 2008 Presentation, March 8, 11:30-11:45

- Work sponsored by grants from NIH-NIDCD