Lexical Neighborhood Influences on Phonetic Perception

James R. Sawusch
Department of Psychology,
Center for Cognitive Science
University at Buffalo

In collaboration with

Rochelle S. Newman (U. of Maryland)
Paul A. Luce (University at Buffalo)
Liza Zimack (University at Buffalo)
Lisa Incognito (University at Buffalo)
Ivy Liu (University at Buffalo)
Sunfa Kim (University at Buffalo)
Sam Soleimany (University at Buffalo)
Mary Flaherty (University at Buffalo)

Research Support

Partial support provided by NIDCD grant R01-DC000219 to the University at Buffalo

Classic Problems in Speech Perception

Variability – Speech signal varies due to talker, dialect, speaking rate and surrounding context. Listeners exhibit perceptual constancy for phonemes, syllables and words.

Segmentation - The speech signal is continuous. Listeners experience a sequence of discrete words.

Dealing with Variability

A theory of speech perception must specify the representation (units) that words are built from. Most theories use some form of phonetic representation.

Most theories of word recognition also ascribe a role for knowledge in perception. Given ambiguous phonetic information, the best matching word is "heard".

This leaves alternative possibilities for how/what knowledge and perception interact.

Knowledge and Perception

Speakers of a language know the words.

Questions:

- Does this knowledge alter listeners' perception of the "sounds" from which words are built?
- How does this influence come about? What aspects of the mental lexicon influence perception?
- What is the nature of the representation of the "sounds" that lead to word recognition?

Examples of Lexical Influences

Phoneme Restoration (work of Warren, Samuel)

Lexical Effect (work of Ganong, Fox, others)

Phoneme Restoration

A phoneme is deleted from a word (replaced by silence). A suitable "masking sound" is overlaid over the missing sound. The word is heard as intact with the missing phoneme restored.

target:	legislature with /s/ missing (cough overlaid)
foil:	legislature intact (cough overlaid)

Listeners find it very difficult to distinguish between these. The intact and missing /s/ words, by themselves, are quite distinct.

Samuel introduced a methodological refinement. Listeners were asked whether an item that thy listen to was intact (original recording with mask overlaid) or restored (recording with phoneme deleted and mask overlaid).

Samuel showed that phoneme restoration is influenced by:

- Similarity of masking sound to deleted phoneme.
- Position of phoneme in word and word length.

A periodic mask such as a tone promotes restoration (a strong illusion that the missing phoneme is present) for a periodic phoneme such as a vowel. An aperiodic (noisey) mask such as a cough promotes restoration for aperiodic phonemes such as voiceless fricatives.

This result shows that the acoustic information in the mask must “match” or substitute for the acoustic information in the missing phoneme. That is, a large part of restoration is “data driven”.

These effects show up in d’ (sensitivity) in a TSD analysis.

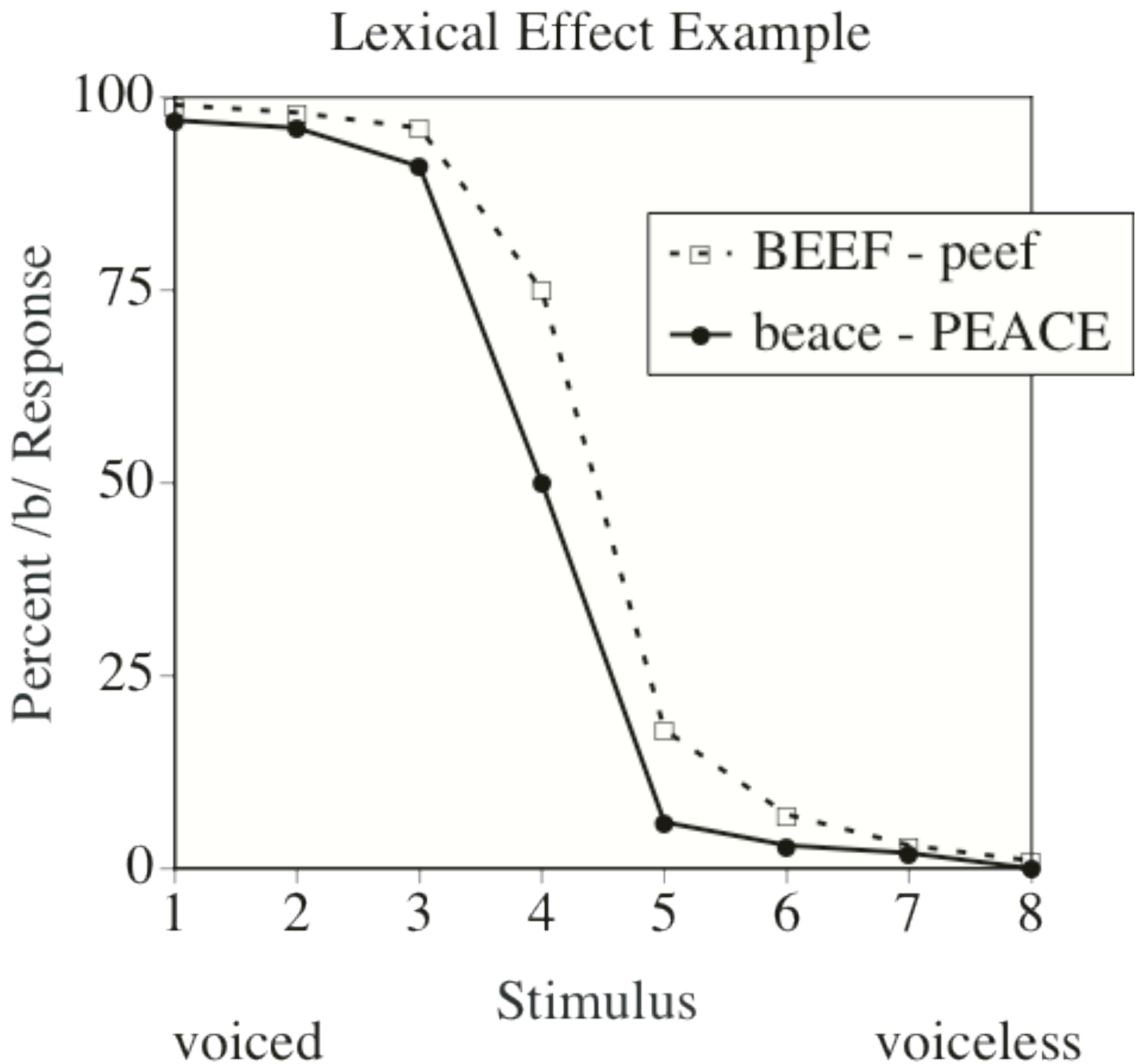
Restoration is stronger when the missing phoneme occurs later in a word relative to earlier. It is stronger in longer words. It does not occur in nonwords or in word initial position (for poly-syllabic words).

This result shows the knowledge influence. The lexical representation must be already partly active (driven from stimulus) to influence restoration.

These effects show up in d' (sensitivity) in a TSD analysis.

Lexical Effect

Two series of syllables are created in which one end of one series is a word (**beef** – peef) and the other end of the other series is a word (beace – **peace**). Using cross splicing of natural syllables, the initial consonant and vowel in each series is the same. The two series *differ only* in their final consonant.


Lexical Effect Example
100
75
50
25
0
Percent /b/ Response
BEEF - peef
beace - PEACE
1
2
3
4
5
6
7
8
Stimulus
voiced
voiceless

Lexical Influences

Connine & Clifton (1987) showed that word frequency also has an effect. Listeners will report the phoneme from the end of the continuum that makes a more common word.

Burton, Baum, and Blumstein (1989) claimed that lexical effects are not robust and do not occur with high quality, natural speech based continua. Their result is "nonreplicable" and most lexical effect studies have used high quality, natural speech.

Lexical Influences - 2

Fox (1984) examined the time course of the lexical effect. He used a reaction time task and then partitioned his data into fast, intermediate and slow responses. The lexical effect was found in slow and intermediate speed responses, but not in fast.

Lexical Influences - 3

Pitt & Samuel, 1993 did a meta-analysis of lexical effect studies and ran new series. They noted the variability in the effect, particularly for series that started in the phonemes /d/ and /t/ (e.g. **deep**-teep and deach-**teach**).

Why is the effect variable? Could it be that there are multiple "sources" of information in the lexicon that do not always vary together?

Sources of Lexical Information

- Lexical Status. Is the phonetic segment part of a word?

- Lexical Neighborhood. How many words is the carrier of the phonetic segment (e.g. syllable) similar to? Are the neighbors similar to each other?

- Probabilistic Phonotactics.

  Likelihood of a phoneme sequence (adjacent or not).

  Intrinsic probability of each phoneme.

Lexical Information and the Lexical Effect

In all prior studies of lexical effect (thru Pitt and Samuel), the sources of lexical information other than lexical status were ignored.

Studies of word recognition (Luce) show that neighborhoods and probabilistic phonotactics exert powerful influences on word recognition. Could they also influence phoneme perception?

Lexical Neighborhood & Similarity

Neighborhood *A target is similar to a word if the word can be created from the target by a one phoneme substitution, addition, or deletion*. This is the 1-phoneme rule.

Examples

pit and bit or pit and pat or pit and pin

spit and pit or spit and sit

spit and split

Lexical Neighborhood Computation

- For a target, determine the words in the dictionary that are neighbors.
- Any neighbor not likely to be in participant's lexicon is eliminated (screen for familiarity).
- Weight each neighbor according to its frequency of usage.
- Sum the frequency-weighted neighborhood density (FWND).

Lexical Neighborhood and Lexical Effect

In most lexical effect studies, lexical status and neighborhood are correlated:

Series	**gift**	-	kift	giss	-	**kiss**
FWND	**11.1**		10.7	26.5		**30.6**
Lexical Effect?				Yes		

Lexical Neighborhood and Lexical Effect - 2

In all /d/ - /t/ series, the non-word end has the higher FWND.

Series	**deep** -	teep	deach -	**teach**
FWND	31.1	**49.4**	**29.8**	28.3
Lexical Effect?		No		

How to Separate?

- Two series of word-nonword and nonword-word. Choose such that neighborhood is controlled.
- Two series of nonwords. In one series, the syllable at one end must be similar to more, higher frequency words while in the other series, the syllable at the other end must be similar to more, higher frequency words.
- Data. Show that ambiguous phonemes are classified by listeners with label similar to real words/denser neighborhoods.

Pure lexical effect

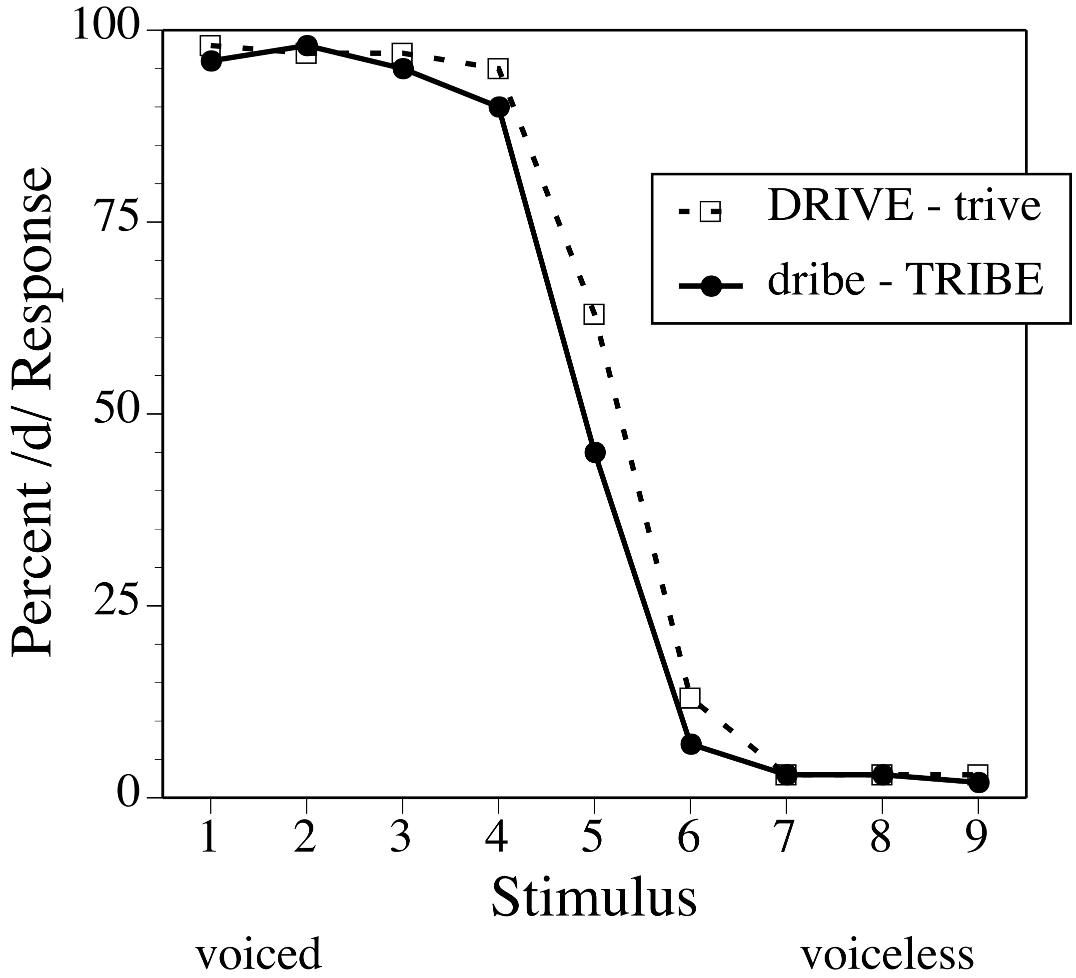
drive-tribe

	drive	-	trive	dribe	-	**tribe**
	/draɪv/	-	/traɪv/	/draɪb/	-	**/traɪb/**
FWND	16.3		**19.3**	11.8		**15.2**

both-dose

	both	-	doth	bose	-	**dose**
	/doθ/	-	/doθ/	/bos/	-	**/dos/**
FWND	**23.1**		22.9	**32.0**		14.9

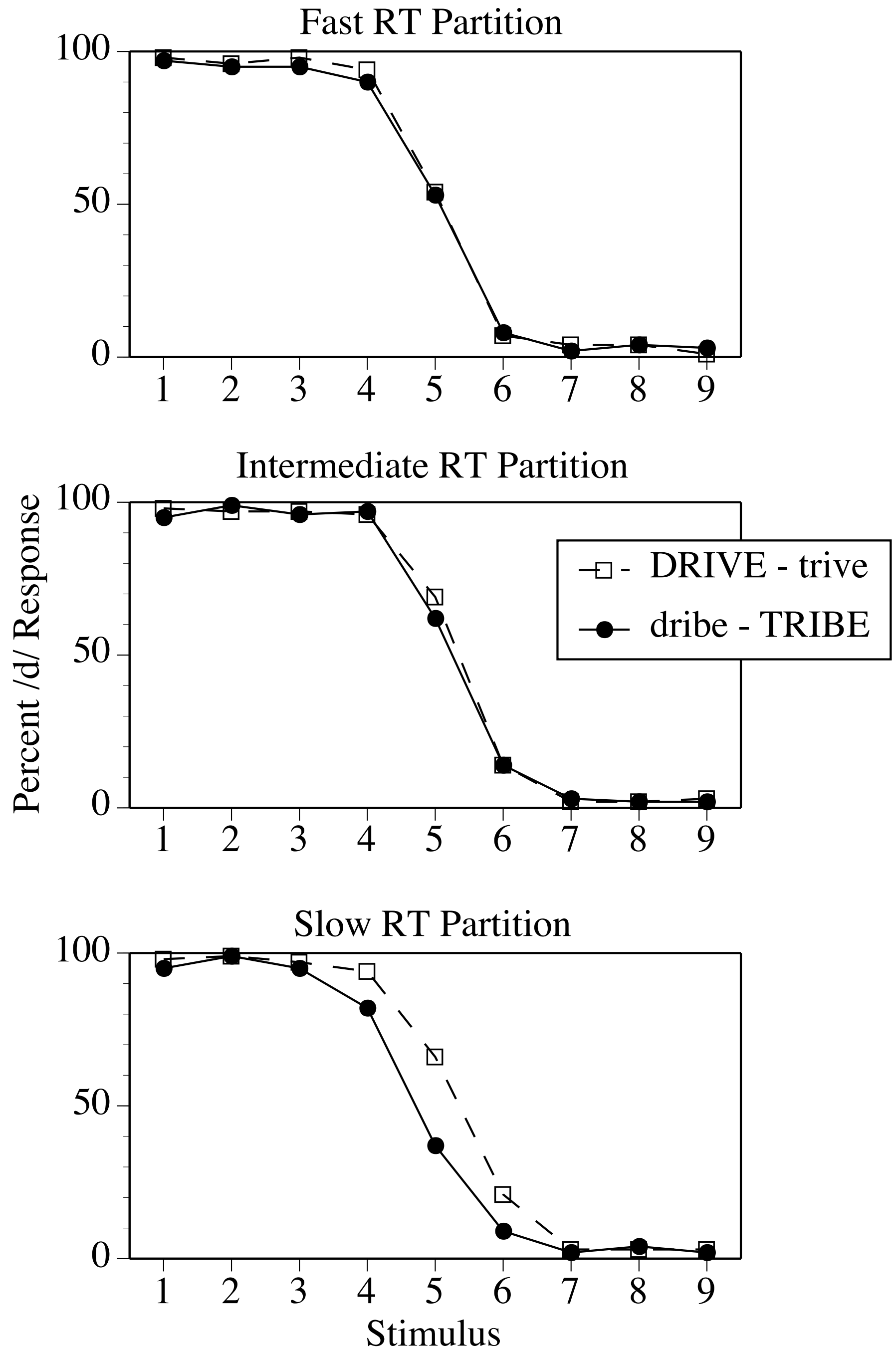
Overall Lexical Status Effect



Partitioned Data

- Divide up data based on listeners' RT. For each listener, separate into fast, intermediate and slowest thirds for each syllable.
- Based on prior work, expect influence of lexical status in slow or slow and intermediate speed responses.

Lexical Status Effect - RT Partitions



Lexical Status Summary

- Once other lexical influences are controlled, lexical status does influence phoneme perception, even for /d/ - /t/.
- Like in previous studies, shows up in slower responses.
- Effect reflects lexical access. Appears as changes in bias parameter (criterion change) in TSD analysis.

Neighborhoods?

- Does neighborhood influence phonetic perception?
- Is this influence also "bias"?
- If an effect is found, is it neighborhood or cohort?
- What is the segmental representation that is being influenced?

Example of Series for Examining Lexical Neighborhood Effects

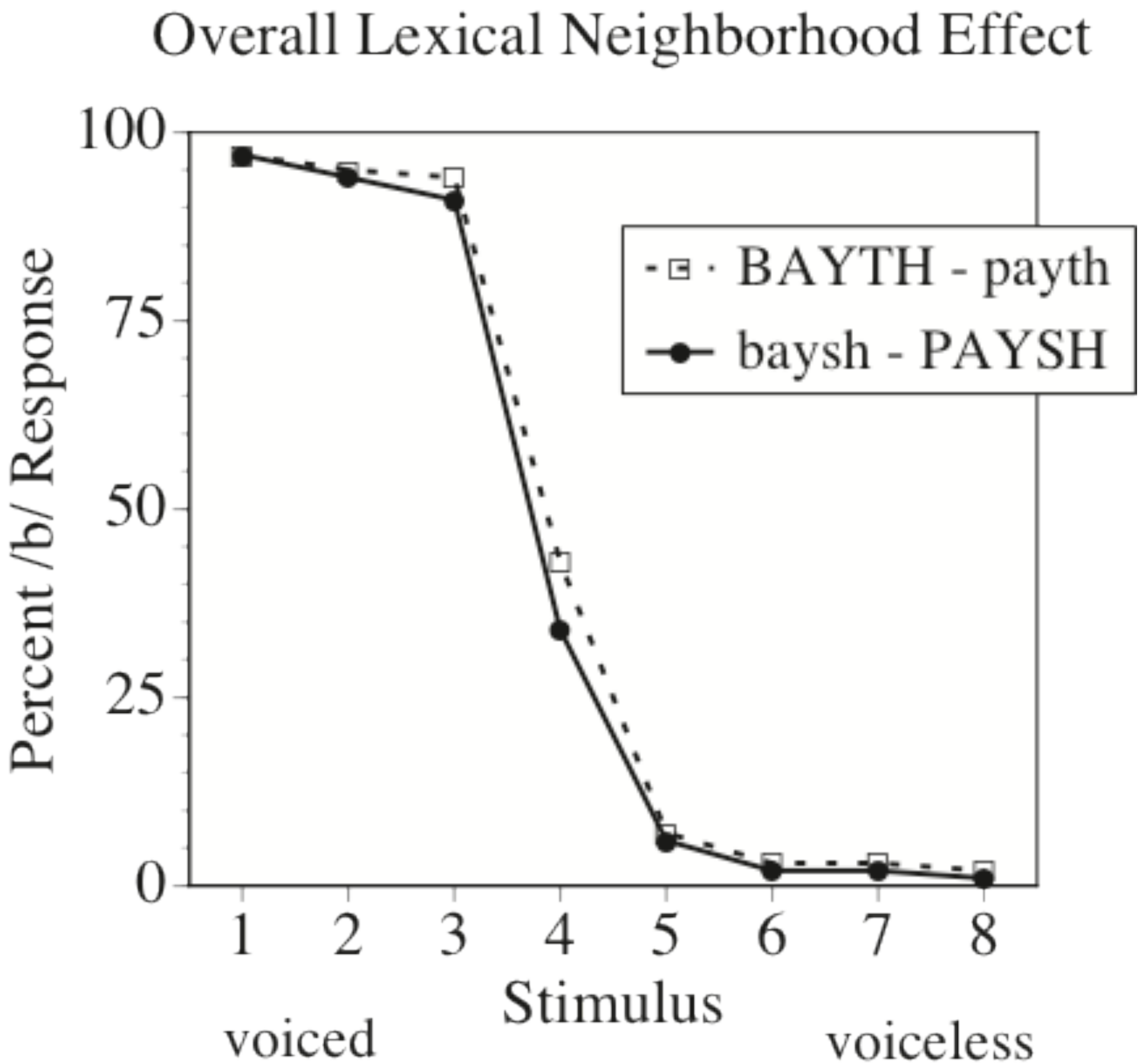
Series	bayth - payth		baysh - paysh	
	/beθ/ -	/peθ/	/beʃ/ -	**/peʃ/**
FWND (density)	**29.7**	25.6	18.9	**23.5**

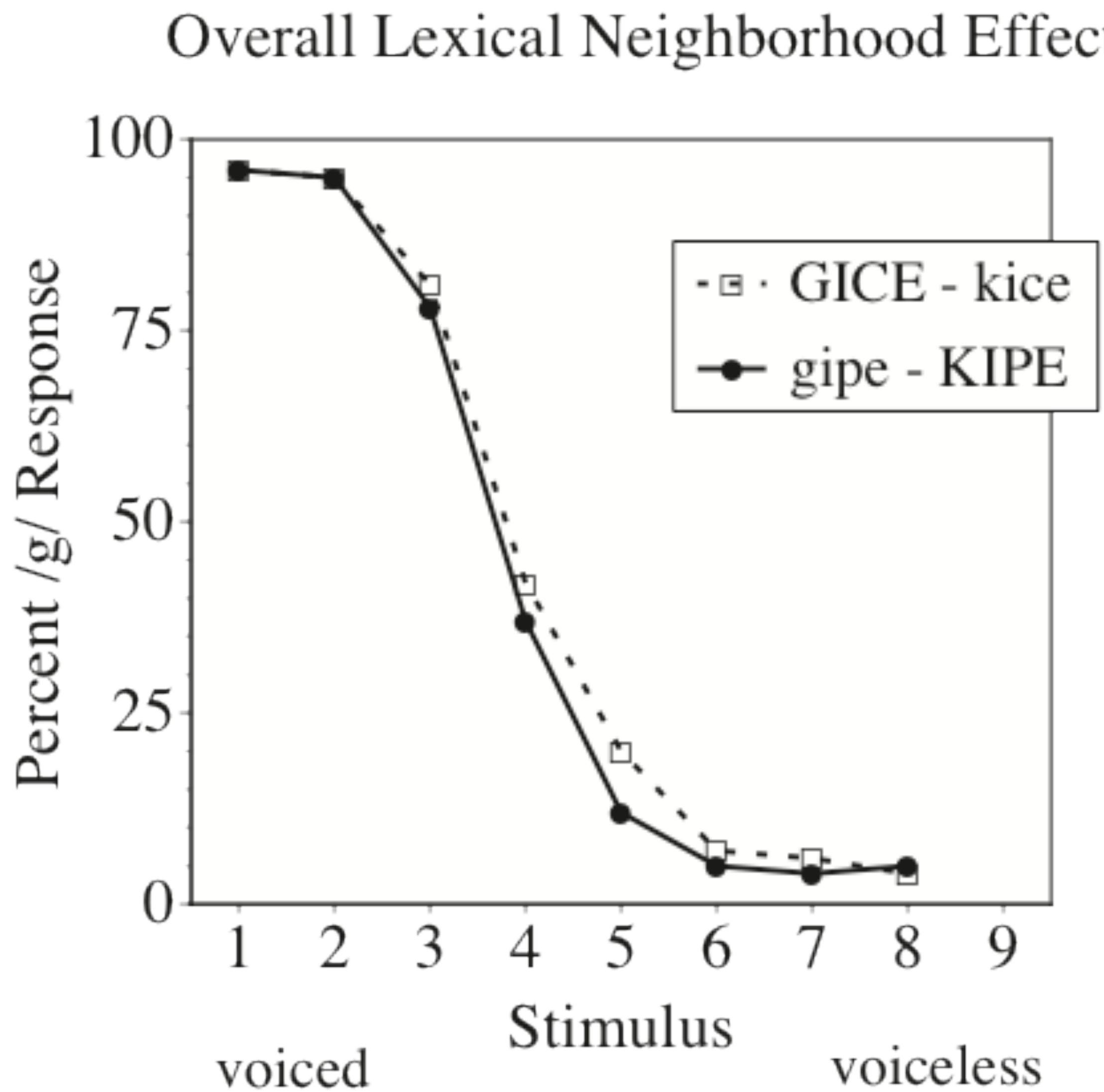
A Lexical Neighborhood

Neighborhood of nonword paysh - /peʃ/

word	frequency	log10(freq x 10)	familiarity
pace	43	2.63	6.7
page	66	2.82	7.0
paid	145	3.16	7.0
pail	62	2.79	7.0
pain	91	2.96	6.9
pave	2	1.30	7.0
pay	172	3.24	7.0
pair	1	1.00	7.0
posh	1	1.00	6.0
push	37	2.57	6.9

$\sum \log10(\text{freq} \times 10) = 23.47$


Overall Lexical Neighborhood Effect
100
75
50
25
0
Percent /b/ Response
BAYTH - payth
baysh - PAYSH
1 2 3 4 5 6 7 8
Stimulus
voiced
voiceless



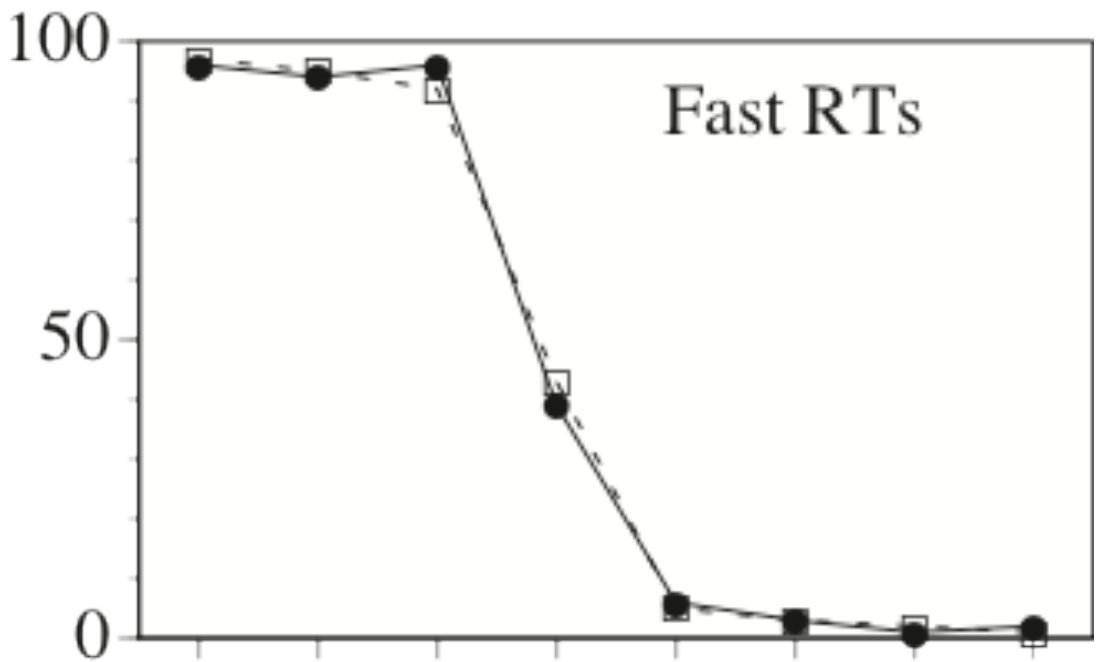
Overall Lexical Neighborhood Effect
100
75
50
25
0
Percent /g/ Response
GICE - kice
gipe - KIPE
1 2 3 4 5 6 7 8 9
Stimulus
voiced
voiceless

Partitioning RT Data

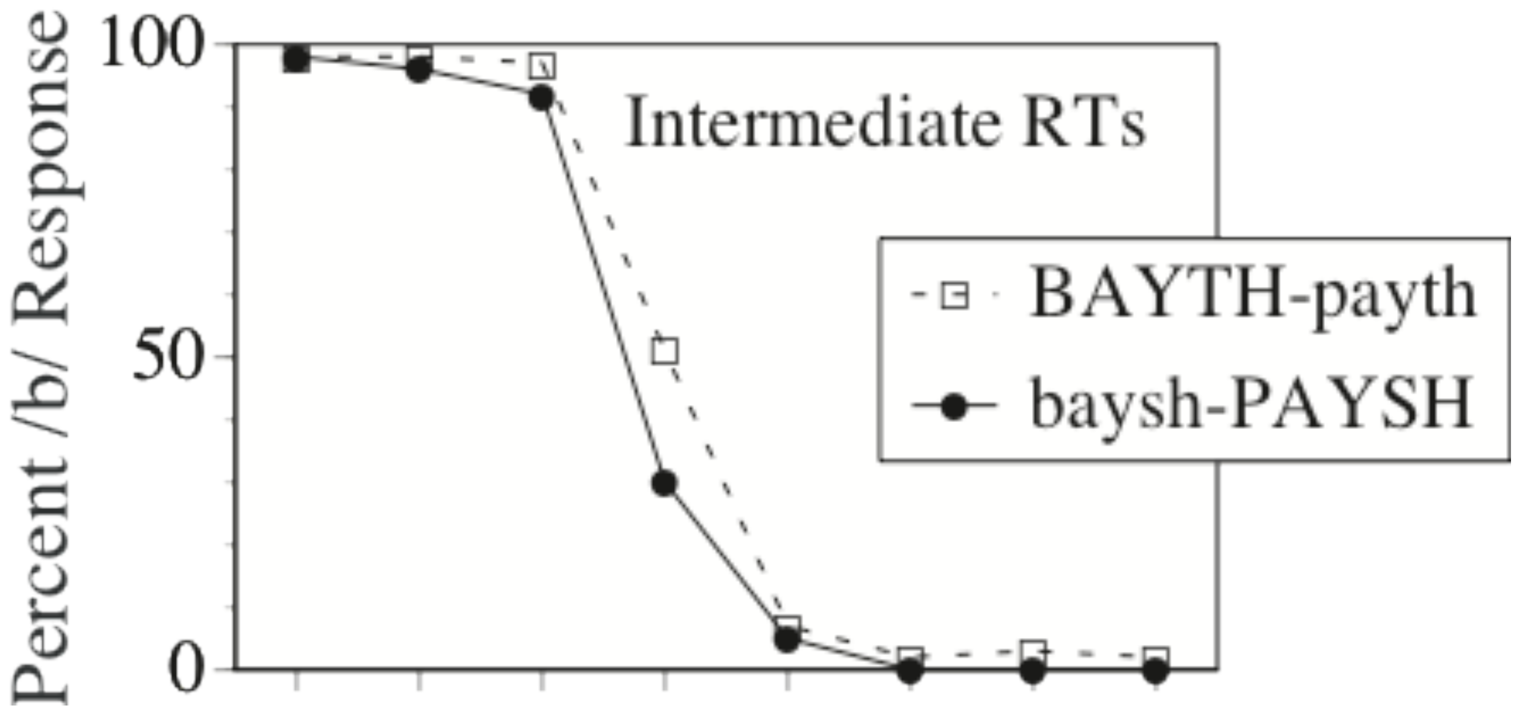
Listeners were asked to respond as rapidly and accurately as possible.

The data for each listener were partitioned into fast, intermediate and slow responses (for each syllable) to examine the effect of neighborhood over the course of perceptual processing (cf. Fox, 1984).

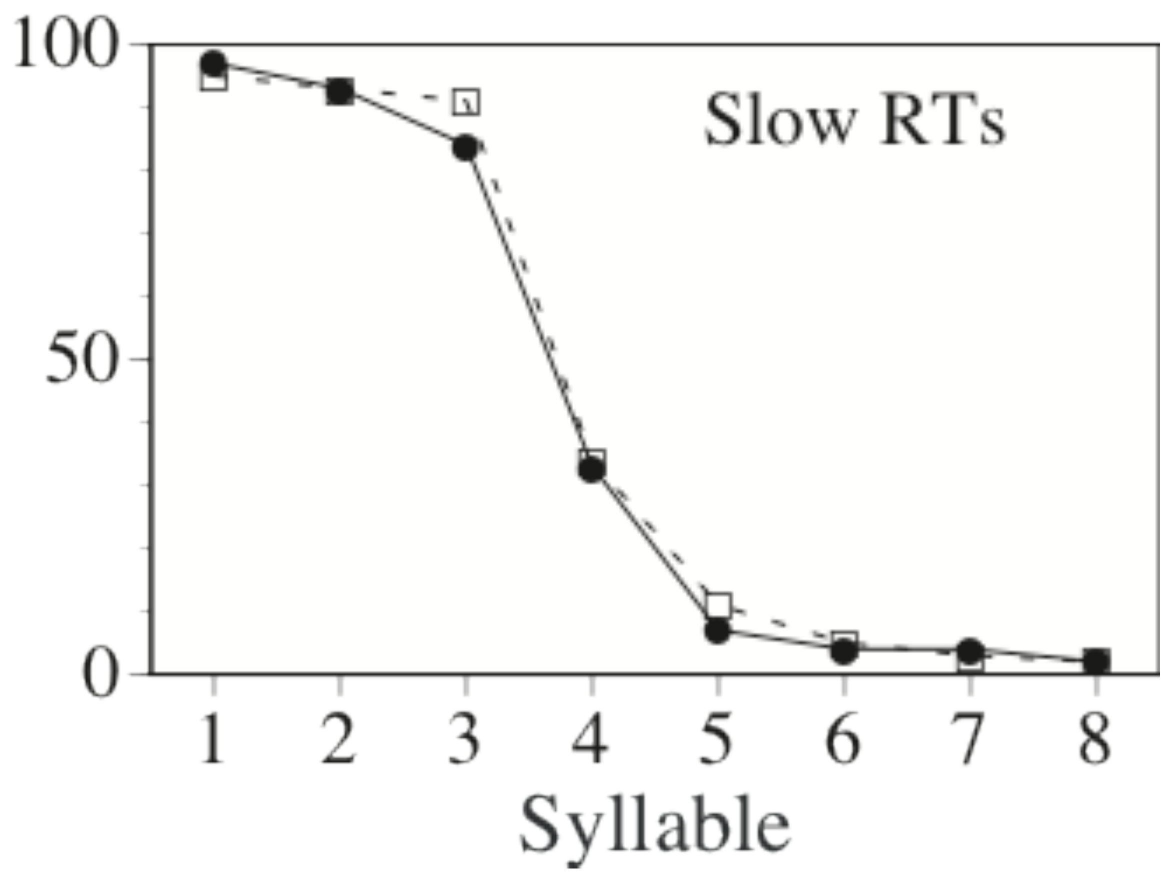
Partitioned bayth-paysh Data 1



Partitioned bayth-paysh Data 2



Partitioned bayth-paysh Data 3



Summary of Basic Results

- Lexical neighborhood influences phonetic perception.

  Have found effect for 7 different series.

  Have found effect with different talkers.

- The effect is transient, takes time to accumulate then dissipates.

- Effect shows up as change in sensitivity in TSD analyses.

- One-phoneme rule, with its assumptions, works well to characterize the similarity among words.

Next Question: Neighborhood versus Cohort

Initial position. These studies examined consonant perception in syllable initial position in CVC syllables.

For initial position, the predictions of a neighborhood activation system and a cohort activation system are largely identical.

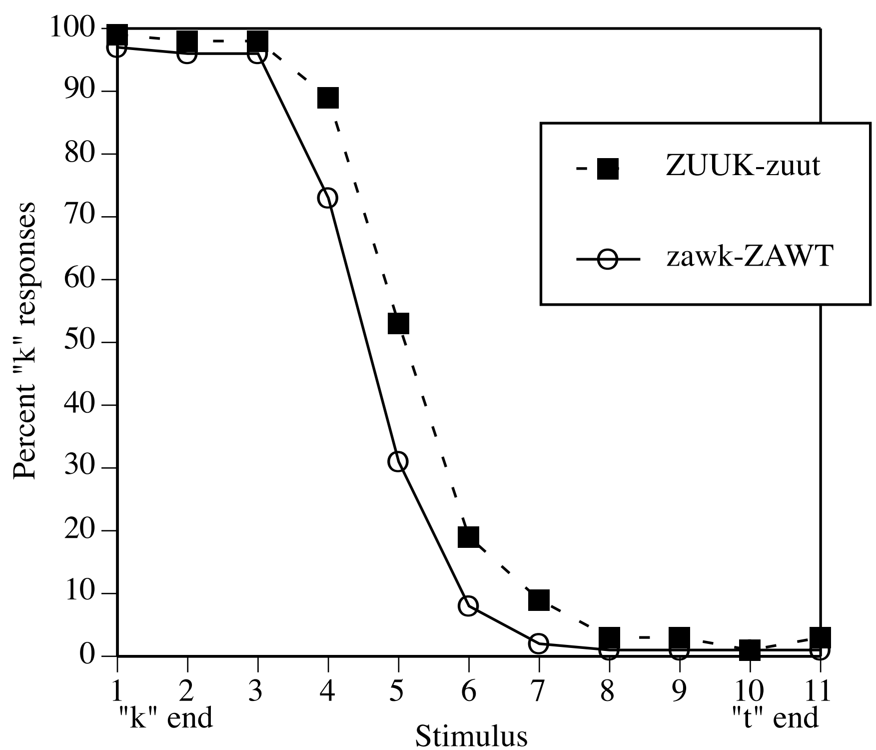
Final position. In syllable final position, the predictions of neighborhoods and cohorts can be separated.

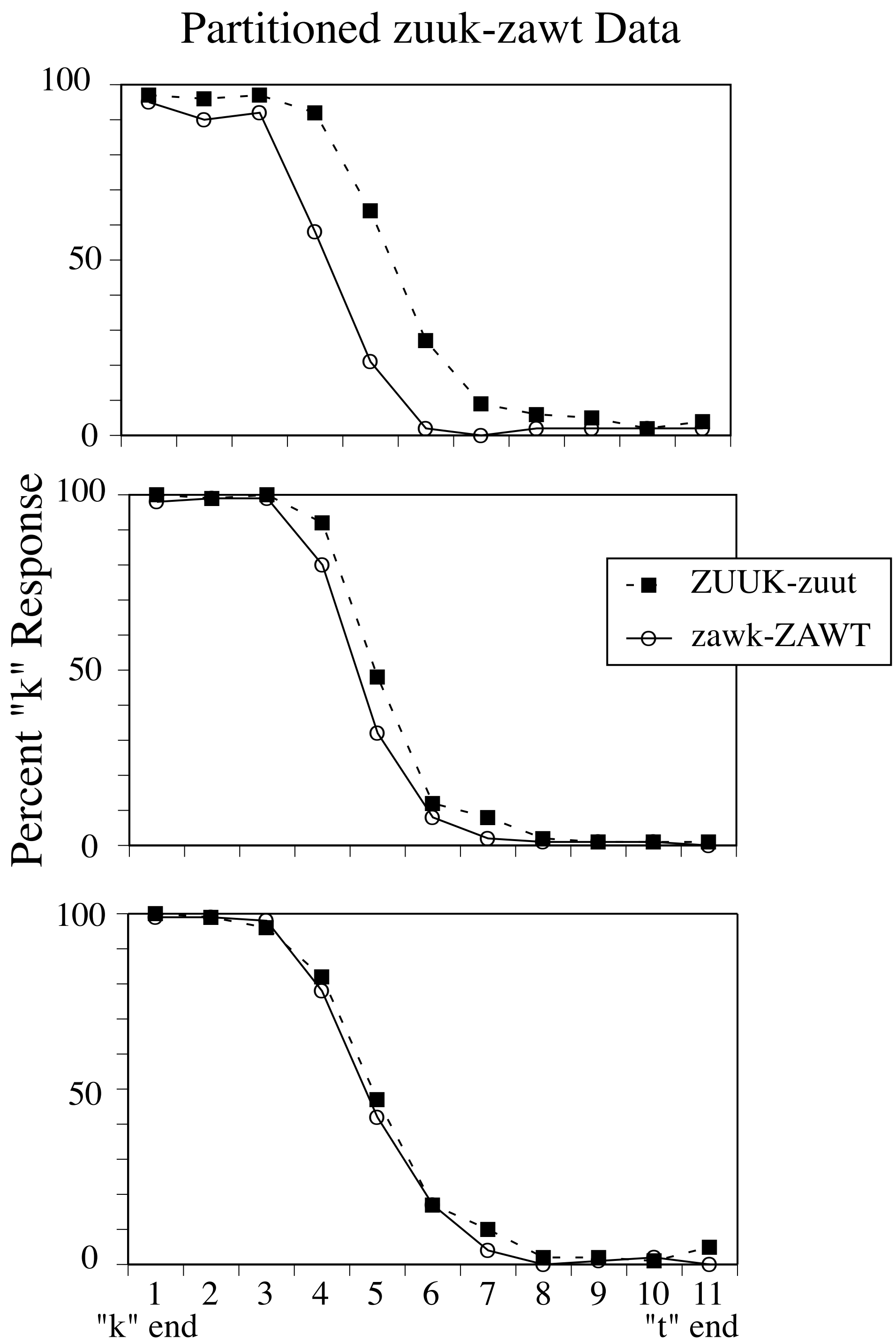
/zʊk/ - /zʊt/ and /zɔk/ - /zɔt/

In this series, neighborhood differences are not cohort differences. None of neighbors for either series start with /z/.

/zʊk/	-	/zʊt/	/zɔk/	-	/zɔt/
17.6		7.5	11.8		20.6

Results from Experiment 4:
"zuuk" - "zawt" stimuli





Partitioned zuuk-zawt Data
Percent "k" Response
100
50
0
ZUUK-zuut
zawk-ZAWT
1 2 3 4 5 6 7 8 9 10 11
"k" end
"t" end

Summary

Clearly, neighborhood can influence phonetic perception separately from cohort.

Influence is in fast partition since context precedes target and has already partly activated neighborhood.

As seen in initial position, influence is transient.

Next Question

Is it just the size of the neighborhood that influences phonetic perception? In word recognition, the interactions among the neighbors influence word recognition.

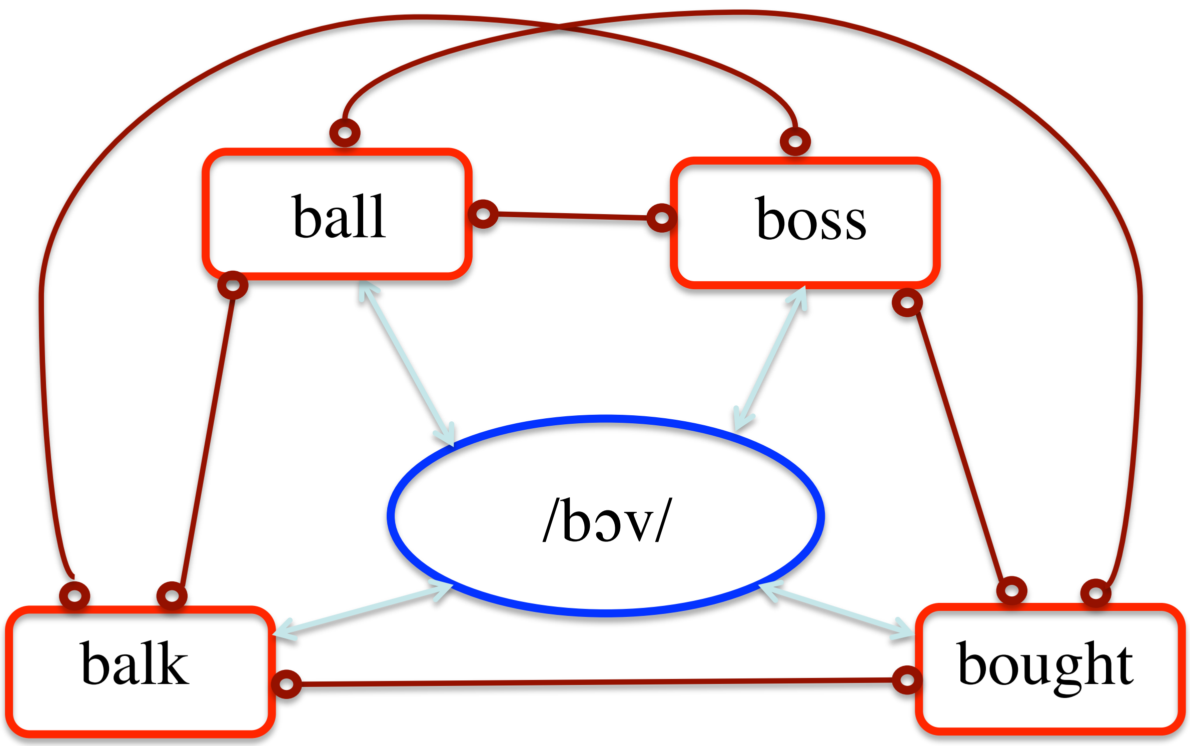
Chan & Vitivitch showed that when a words neighbors are also mutual neighbors, it slows word recognition. Might this mutual interaction also influence phonetic perception?

Cluster Coefficient

The cluster coefficient is an index of whether an items neighbors are also neighbors of each other. It varies from 0 (no mutual neighbors) to 1 (all neighbors are mutual).

Can we create a set of series where neighborhood density is controlled but the neighborhoods differ in their cluster coefficient? Will this influence phoneme perception?

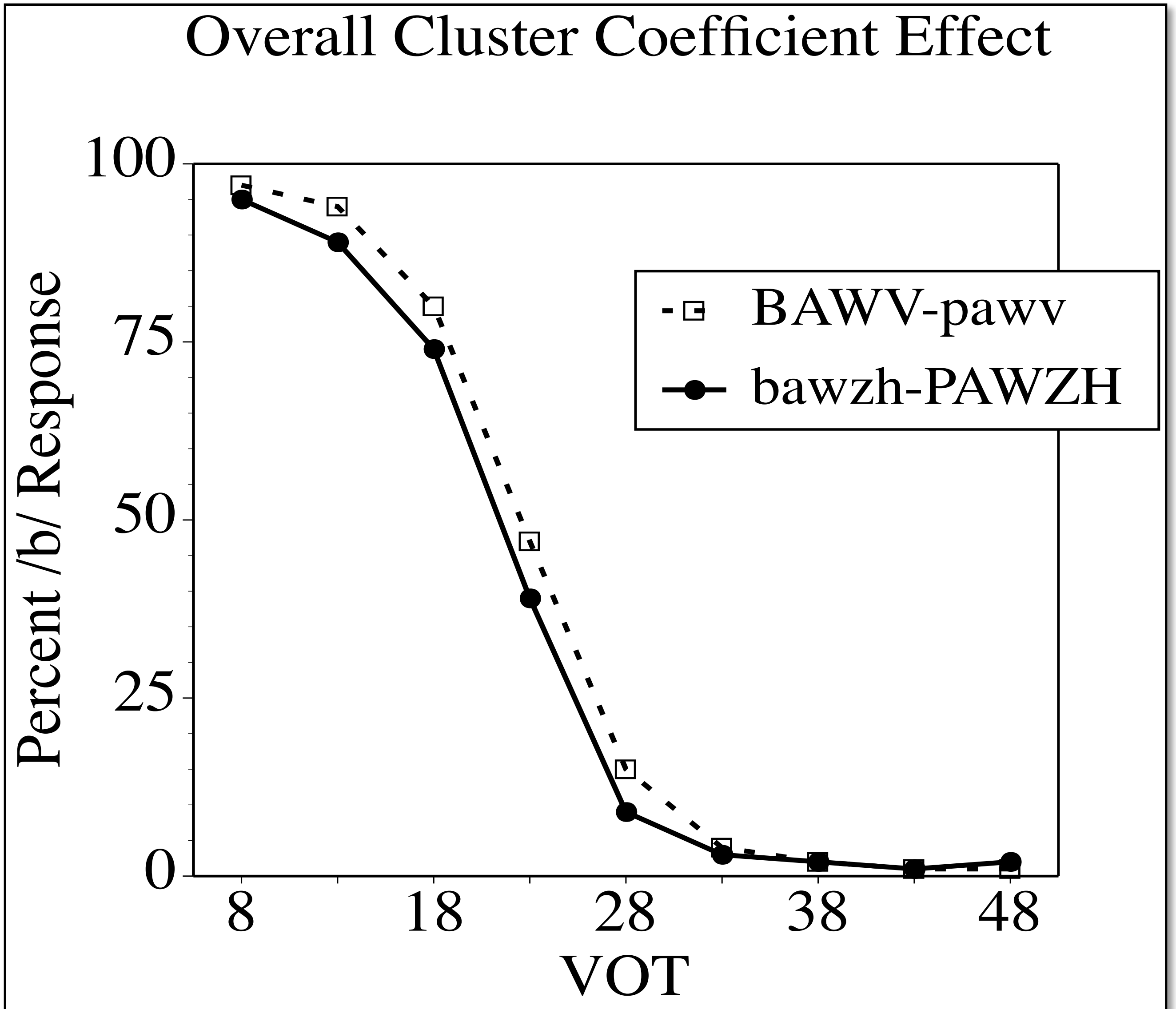
Cluster Coefficient Example

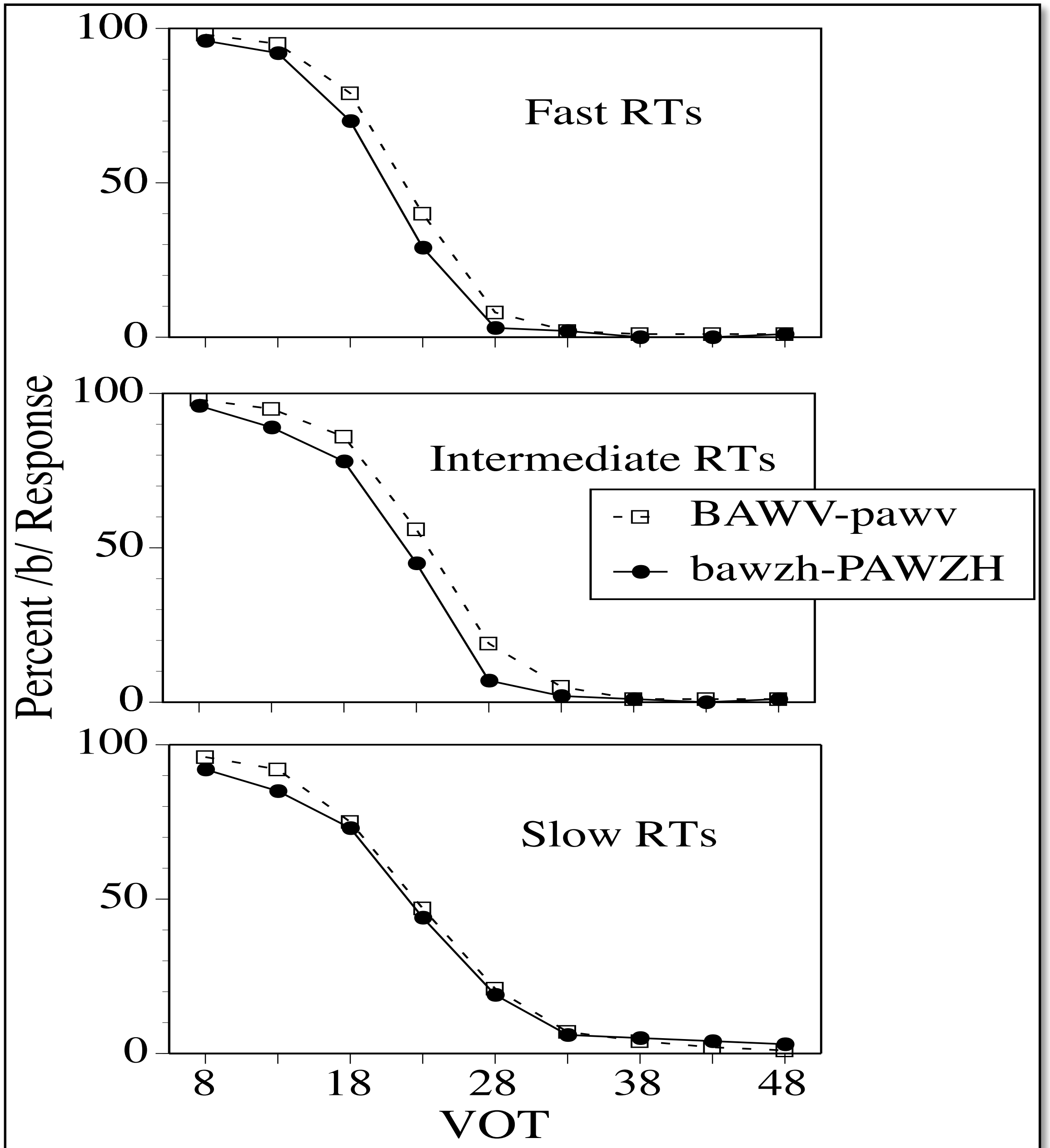


Cluster Coefficient Series

	/bɔv/	-	/pɔv/	/bɔʒ/	-	/pɔʒ/
FWND	9.1		9.0	10.1		7.7
CC	1.0		0.6	0.6		1.0

Neighborhoods are controlled and, if anything, predict an influence opposite to that of the cluster coefficient. Since a higher CC seems to increase lexical competition in word recognition, it is predicted to alter phoneme perception in much the same way as FWND.


Overall Cluster Coefficient Effect
100
75
50
25
0
Percent /b/ Response
BAWV-pawv
bawzh-PAWZH
8
18
28
38
48
VOT


Fast RTs
Intermediate RTs
Slow RTs
BAWV-pawv
bawzh-PAWZH
Percent /b/ Response
VOT
100
50
0
8
18
28
38
48

Cluster Coefficient Summary

The pattern of mutual interaction among neighbors also appears to influence phonetic perception. When this interaction among mutual neighbors is stronger, it has a larger influence on phonetic perception.

Next Question

Singleton phonemes. These studies examined consonant perception in syllable initial position in CVC syllables.

What about a consonant in a cluster – the /b/ or /l/ in black?

If words are built from phonemes as we have assumed, we should see similar effects.

…But, what is the alternative?

The Cluster Rule

Suppose that all of the consonants in a cluster are grouped together, as a single unit. That is, the /g/ and /r/ in /grin/ (*green*) or the /s/, /p/, and /r/ and the final /n/ and /t/ of /sprInt/ (*sprint*) are grouped into single units.

In this case, both *green* and *sprint* have 3 units (CVC). A one-unit substitution rule (similar to the one-phoneme rule) can be used to find sets of nonwords where neighborhoods are controlled by one rule and vary according to the other.

Now, we can pit the two rules against one another.

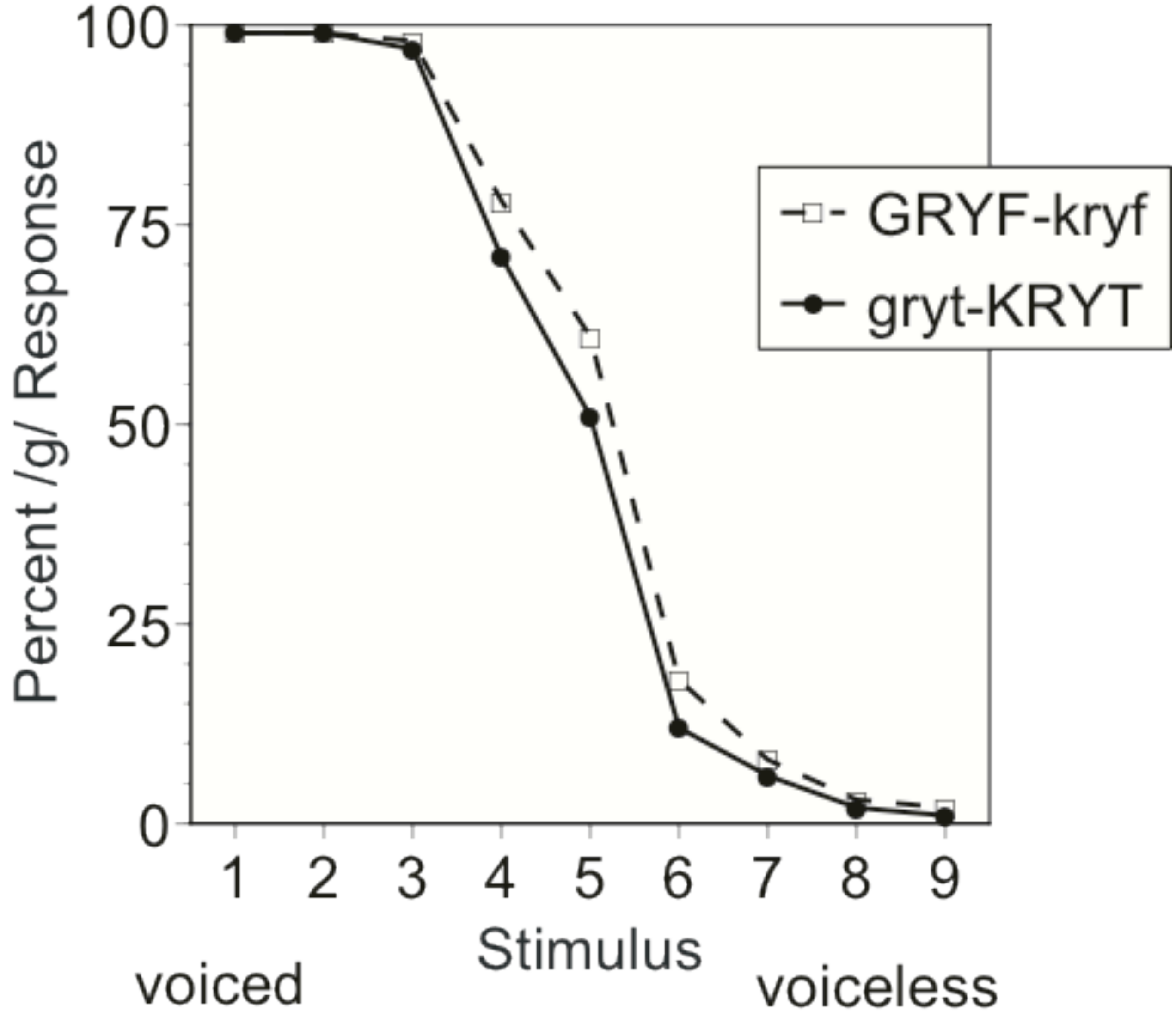
Initial Position
One-Phoneme Rule Predicts Difference

Rule	**/graIf/**	- /kraIf/	/graIt/ -	**/kraIt/**
phoneme – FWND	6.2	6.2	18.1	**25.4**
N	**4**	3	9	**12**
cluster – FWND	19.4	21.1	60.7	62.2
N	9	8	25	25

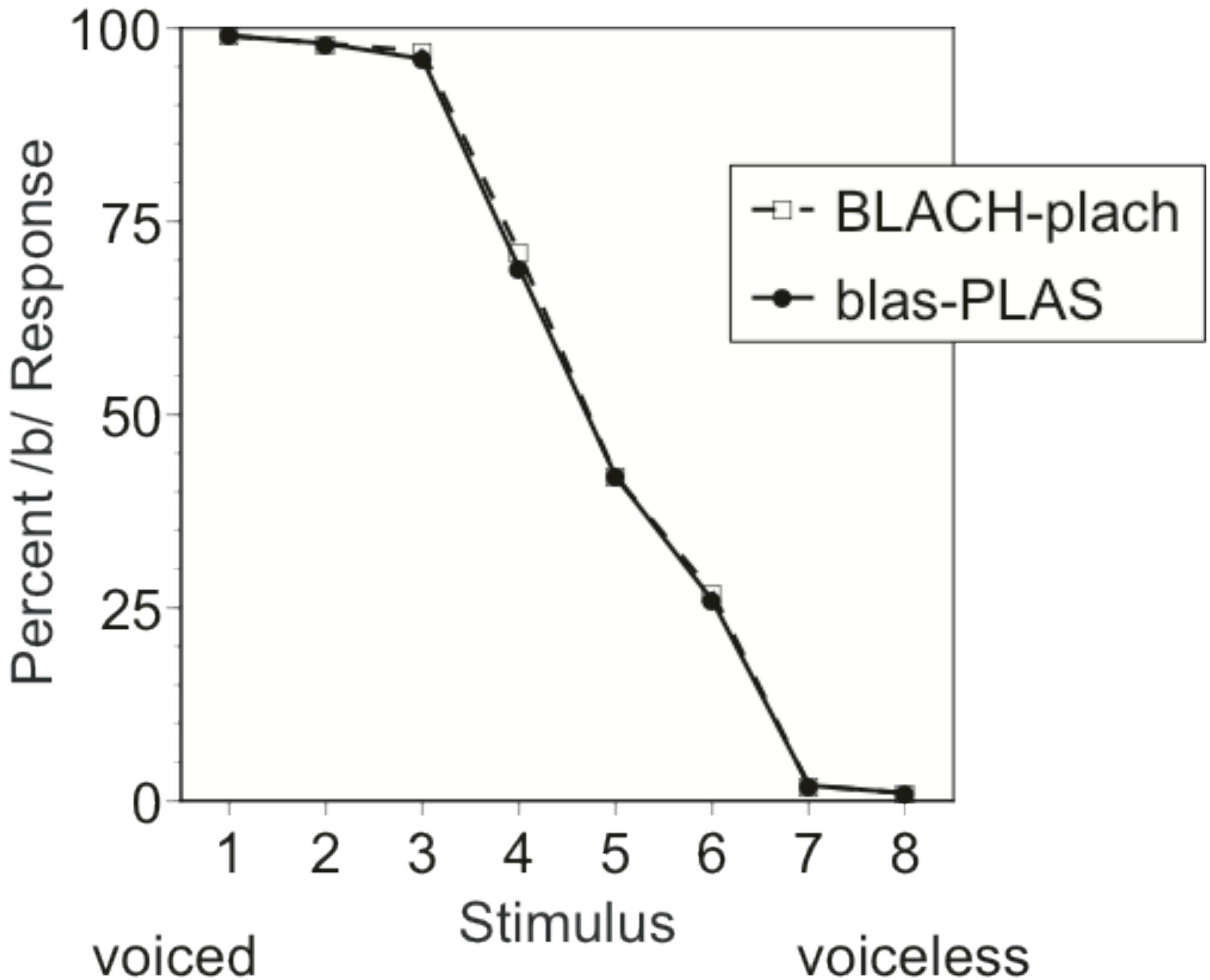
Initial Position
Cluster Unit Rule Predicts Difference

Rule	**/blætʃ/**	/plætʃ/	/blæs/	**/plæs/**
phoneme – FWND	9.7	9.4	20.2	21.5
N	6	5	9	8
cluster – FWND	**28.2**	26.6	37.3	**41.4**
N	**15**	13	16	16

Overall Lexical Neighborhood Effect Difference Predicted by Phoneme Rule



Overall Lexical Neighborhood Effect Difference Predicted by Cluster Rule



Initial Position Consonants - Summary

The results for initial consonants in clusters mirror those for single initial consonants. The one-phoneme change rule accurately predicts the influence of neighborhoods on phoneme identification.

The cluster rule fails to predict the observed effects and effects are observed when it predicts no effect.

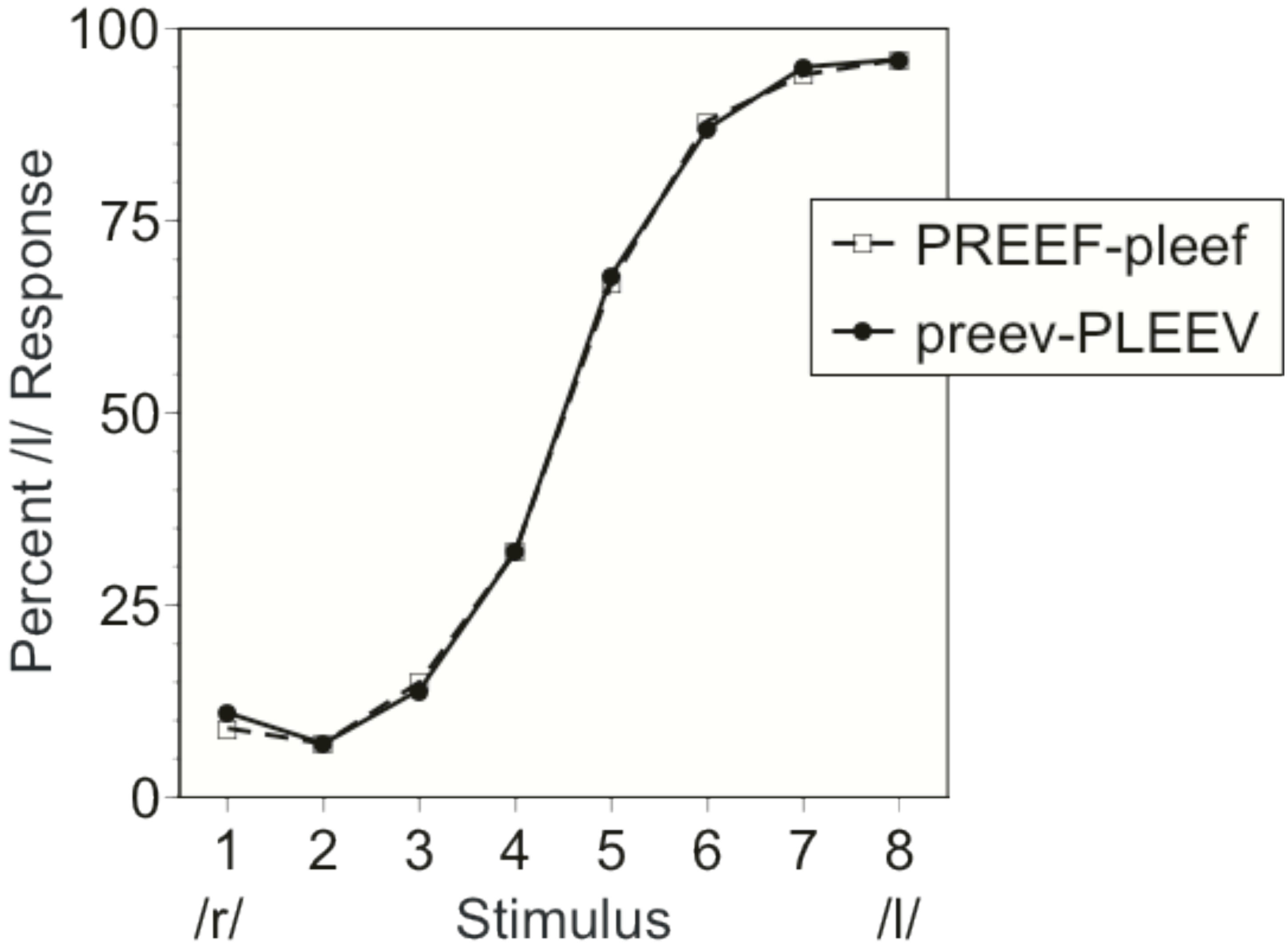
Second Position
One-Phoneme Rule Predicts Difference

Rule	**/prif/** - /plif/		/priv/ - **/pliv/**	
phoneme – FWND	**11.4**	9.6	5.6	**12.9**
N	5	5	3	**6**
cluster – FWND	23.2	24.0	19.4	20.1
N	10	11	10	11

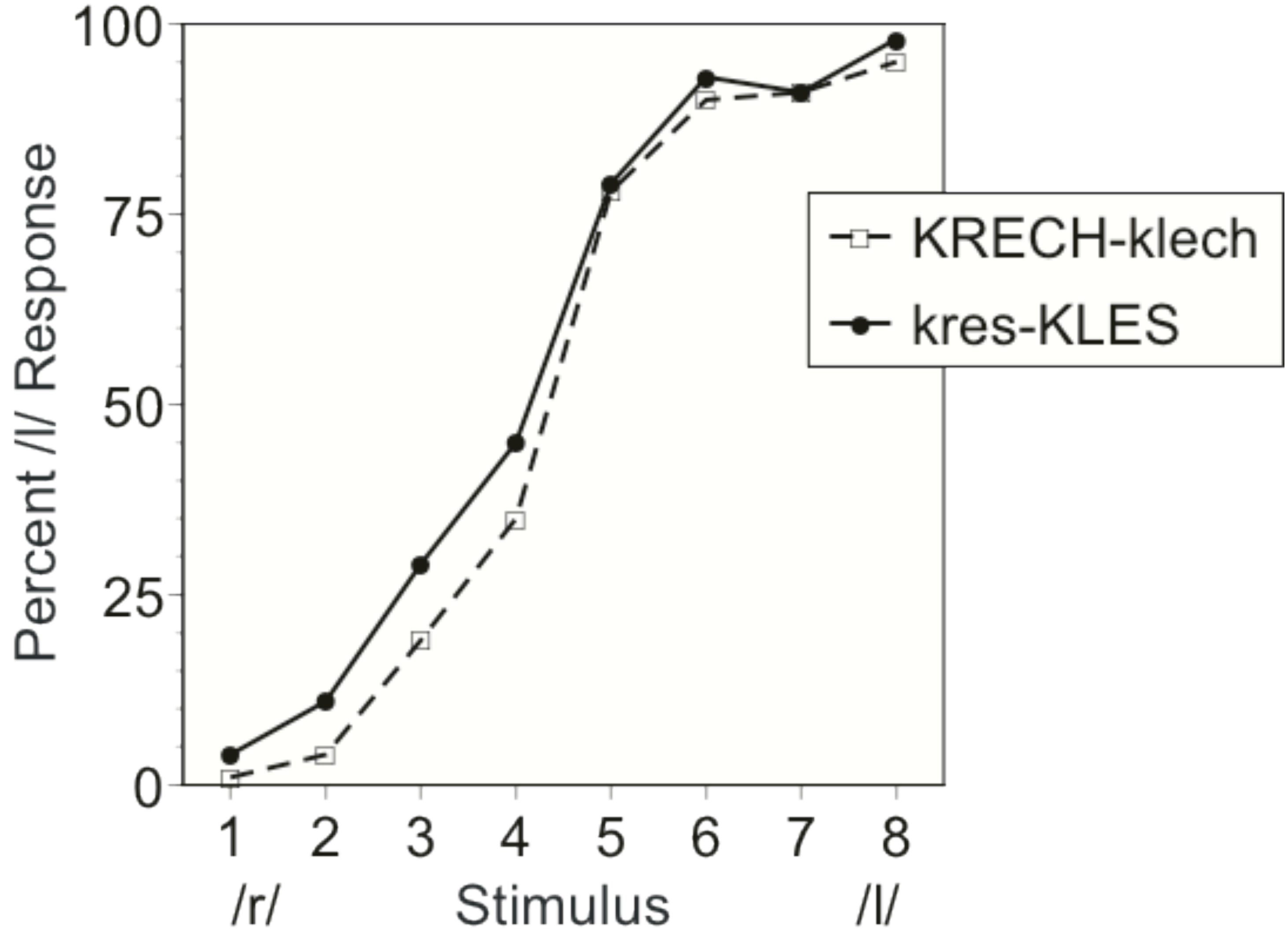
Second Position
Cluster Rule Predicts Difference

Rule	**/kretʃ/** - /kletʃ/		/kres/ - **/kles/**	
phoneme – FWND	7.7	7.7	25.0	21.8
N	3	3	13	7
cluster – FWND	**13.3**	10.6	49.1	**51.4**
N	**9**	4	19	**20**

Overall Lexical Neighborhood Effect
Difference Predicted by One Phoneme Rule



Overall Lexical Neighborhood Effect
Difference Predicted by Cluster Unit Rule



Second Position Consonants - Summary

In second position in a syllable initial consonant cluster, the one-phoneme change rule fails to predict the observed effects. This implies that this rule does not accurately predict neighborhoods for second position in an initial cluster.

The cluster rule accurately predicted the presence and absence of neighborhood effects for the /r/ - /l/ contrast in second position. This implies that clusters may not be decompositional in second (or third) position.

Theory?

The data seem to imply that as speech unfolds from left (early) to right (late), successive units (phonemes) are grouped to form larger units (cf. Luce and the Adaptive Resonance approach to word recognition). It is these larger units that are influenced by lexical neighborhood during speech perception.

Conclusions

- The one-phoneme change rule for describing the similarity (neighborhoods) of words appears to accurately predict the influence of lexical neighborhood on the perception of syllable initial (and final) consonants. The mutual interactions among neighbors are also influential.

- Consonants in initial clusters that are not syllable initial do not appear to follow the one-phoneme change rule. This implies that they are represented, at some point in perception, as clusters rather than as a sequence of consonants.