Effects of Attention on the Strength of Lexical Influences on Speech Perception: Behavioral Experiments and Computational Mechanisms

Daniel Mirman[a], James L. McClelland[b], Lori L. Holt[c], James S. Magnuson[a]

[a]*Department of Psychology, University of Connecticut, and Haskins Laboratories*
[b]*Department of Psychology and Center for Mind, Brain, and Computation, Stanford University*
[c]*Department of Psychology and Center for the Neural Basis of Cognition, Carnegie Mellon University*

Abstract

The effects of lexical context on phonological processing are pervasive and there have been indications that such effects may be modulated by attention. However, attentional modulation in speech processing is neither well documented nor well understood. Experiment 1 demonstrated attentional modulation of lexical facilitation of speech sound recognition when task and critical stimuli were identical across attention conditions. We propose modulation of lexical activation as a neurophysiologically plausible computational mechanism that can account for this type of modulation. Contrary to the claims of critics, this mechanism can account for attentional modulation without violating the principle of interactive processing. Simulations of the interactive TRACE model extended to include two different ways of modulating lexical activation showed that each can account for attentional modulation of lexical feedback effects. Experiment 2 tested conflicting predictions from the two implementations and provided evidence that is consistent with bias input as the mechanism of attentional control of lexical activation.

Keywords: Speech perception; Attention; Lexical feedback; Neural networks; Interactive processing; Phoneme recognition; Human experimentation

1. Introduction

Lexical knowledge can influence listeners' recognition of speech sounds. As just one example, speech sounds in words are recognized more quickly than speech sounds in non-words ("word advantage"; Rubin, Turvey, & van Gelder, 1976; see also Mirman, McClelland,

& Holt, 2005). Some researchers have suggested that these lexical effects are modulated by degree of lexical, as opposed to prelexical, attention. For example, lexical effects emerge more strongly when listeners perform a lexical task than when they perform a nonlexical task (Eimas, Hornstein, & Payton, 1990; Vitevitch & Luce, 1999), suggesting that task demands modulate the activation of lexical representations. Another study (Cutler, Mehler, Norris, & Segui, 1987) found that the word advantage emerged when stimulus lists were heterogeneous with respect to consonant-vowel structure of the stimuli but not when the lists were homogenous (only consonant-vowel-consonant stimuli). However, stimulus variability need not entail greater attention to lexical information and task differences are not just a matter of attention. One important factor is that tasks that take longer to perform would be expected to show bigger lexical effects simply due to allowing more time for lexical information to affect lower levels of processing. An experimental paradigm in which the critical items and task are matched across attention conditions is required for a clear assessment of attentional modulation of lexical effects on speech processing.

Attention can be manipulated explicitly (e.g., by instructions to participants) or implicitly (e.g., by the demands of the task or stimuli). In the cases of implicit attention manipulation, the key assumption is that participants will focus attention on information that is useful to task performance rather than information that is not useful to task performance; for example, participants will reduce their attention to lexical information when the proportion of words is low because the lexical information is irrelevant or even misleading with respect to task performance. In previous studies, when the proportion of words relative to non-words in an experimental block was reduced, the use of lexical level information was also reduced in a wide range of tasks, including single word reading (increased errors on inconsistent words; Monsell, Patterson, Graham, Hughes, & Milroy, 1992), speech production (decreased word bias in speech errors; Hartsuiker, Corley, & Martensen, 2005), spoken word recognition (decreased lexical neighborhood effects; Vitevitch, 2003), and verbal short-term memory (increased errors on words; Jeffries, Frankish, & Lambon Ralph, 2006). In each of these cases the results were consistent with the assumption that participants will focus attention on information that is useful to task performance: lexical information is useful to word reading, speech production, verbal short-term memory, etc., only when most stimuli are words; if most of the stimuli are non-words, lexical information is not useful and is not attended. Experiments that manipulate the proportion of words among the filler items (and thus the overall proportion of words in the experimental block) provide a paradigm in which critical items and task can be controlled while attention is manipulated. Experiment 1 adapted this paradigm to examine attentional modulation of lexical feedback effects when critical stimuli and task are matched across attention conditions.

Investigating effects of attention on speech perception allows both a test and an extension of current models of speech perception. Lexical effects on speech sound recognition are consistent with both interactive (e.g., TRACE, McClelland, & Elman, 1986) and autonomous (e.g., Merge; Norris, McQueen, & Cutler, 2000) models of speech perception. However, Norris et al. (2000) argued that feedback effects are obligatory in interactive models and, thus, that attentional modulation of lexical effects is inconsistent with interactive models. This alleged failure of interactive models formed one of the key arguments in favor of autonomous

models of speech perception (Norris et al., 2000). Many of the arguments against interactive processes in speech perception have been discredited (see McClelland, Mirman, & Holt, 2006, for a review), but the issue of attention modulation has not been addressed. In fact, although researchers have appealed to attention as an account of their findings, there has been no attempt to empirically test a computational framework for attentional influences on speech perception.

The present work seeks to answer three questions: (a) Can attention (manipulated by proportion of words) modulate lexical feedback effects when critical stimuli and task are identical across attention conditions? (b) Is there a computational account of the effects of attention that is consistent with interactive models of speech perception (contrary to the criticisms of Norris et al., 2000)? (c) What are the consequences of reduction of lexical attention on the dynamics of lexical activation and competition and how does this inform the possible computational implementations of attentional modulation? We begin with findings from a phoneme detection experiment in which attention was manipulated by the proportion of words in a block in order to provide a well-controlled test of attentional modulation of lexical feedback effects. We then describe a general mechanism of attentional modulation that is consistent with interactive principles and extend the framework of the interactive TRACE model with two possible implementations of this general mechanism to account for basic attentional modulation of lexical feedback effects. The attention mechanism is quite general with many specific implementations possible, but behavioral data from a follow-up experiment are consistent with only one of the two implementations we tested.

2. Experiment 1

Experiment 1 was designed to test whether manipulation of proportion of words can modulate the word advantage on phoneme detection. Participants had to detect a phoneme target in words and non-words with the proportion of words manipulated between participants. If proportion of words affects lexical attention, then the word advantage in phoneme detection should be smaller when the proportion of words is low. Importantly, manipulation of proportion of words allows the critical stimuli and task to be identical across attention conditions.

2.1. Methods

2.1.1. Materials
Table 1 shows an example of each type of stimulus and the number of stimuli of each type in a block.

2.1.1.1. Critical stimuli: The critical stimuli were 40 word–non-word pairs (20 /t/-final and 20 /k/-final) equated for phonotactic probabilities and containing phoneme targets in the final position. Non-words were created by swapping consonant onsets between the words (e.g., "hemlock," "logic" and "lemlock," "hogic"). Since the non-words only differ from words by their onsets, controlling for intra-word phonotactic effects was accomplished by matching the onset-vowel rates of occurrence between the words and non-words. To minimize prelexical consequences of swapping onsets, critical words were constrained to have consonant onsets,

Table 1
Examples (number of trials) of stimuli used in Experiment 1

		Critical	Filler	Attention Shifters		Total (High)	Total (Low)
Target-present	W	"Hemlock" (10)	"Rascal" (10)	"Copper" (60)		80	20
	NW	"Lemlock" (10)	"Jascal" (10)		"Kepper" (60)	20	80
Target-absent	W		"Salami" (20)	"Fashion" (60)		80	20
	NW		"Ralami" (20)		"Shafen" (60)	80	20
Total (high)		20	60	120	0	200	
Total (low)		20	60	0	120		200

exactly two syllables, and stress on the first syllable (see Appendix A for a full list of critical stimuli with average onset-vowel occurrence rates). The word–non-word pairs were divided in half and each participant heard half of the words and the other half of the non-words (i.e., half of the participants heard "hemlock" and the other half heard "lemlock"; likewise, half heard "logic" and half heard "hogic"); thus, there were 10 critical items per condition per participant.

2.1.1.2. Filler stimuli: The purposes of the filler stimuli were to vary phoneme target position, to vary the stress and syllabic structure in the overall set of materials (critical items were all target-final, two-syllable with primary stress on the first syllable), and to equate target-present and target-absent trials. To these ends, two- and three-syllable words and non-words with targets in initial and medial positions were included. In addition, target-absent words and non-words were chosen such that phoneme targets occurred in 50% of the words and 50% of the non-words.

2.1.1.3. Attention-shifting filler stimuli: The purpose of the attention-shifting filler stimuli was to modulate participants' lexical attention by manipulating the overall proportion of words in the experimental session. To this end, attention-shifting filler stimuli were either 120 words *or* 120 non-words, depending on the attention condition, with varied stress patterns, two or three syllables, and phoneme targets in either initial, medial, or final position or no phoneme target (60 were target present and 60 were target absent). Non-words were derived from the words by changing nontarget phonemes without violating general phonotactic principles (e.g., Vander Wyk & McClelland, 2004). Each participant completed a block of 200 trials, of which 80% were words and 20% were non-words (high lexical attention condition) or 20% were words and 80% were non-words (low lexical attention condition).

2.1.1.4. Stimulus construction: All stimulus materials were spoken by a male native speaker of American English in the context of the sentence "Say [item] again" and digitally recorded at a 22,050 Hz sampling rate (non-words were spoken in their non-word form; e.g., "lemlock"). All tokens were digitally excised from the sentence and filtered to remove background noise. To match the acoustic realization of the target phoneme across critical words and non-words the final consonant (i.e., the phoneme target) was spliced from one member of a

word–non-word pair to the other member of the pair. For half of the items the phoneme was spliced from the word to the non-word, for the other half it was the opposite. This method insured that the phoneme target was identical in each member of a word–non-word pair and that there was no systematic bias introduced by splicing.

2.1.2. Procedure

Participants were seated in sound-attenuating booths where they heard words and non-words presented through headphones at comfortable listening levels and made "yes"/"no" responses using an electronic button box. For each token, participants were asked to determine whether the spoken item contained the target phoneme (/t/ or /k/) or not. Half of the participants were assigned to the high lexical attention condition (80% words, 20% non-words) and half to the low lexical attention condition (20% words, 80% non-words). Target phoneme was counterbalanced across participants: half of the participants monitored for /t/ and half for /k/. Button label assignments were also counterbalanced across participants. The first 40 trials were constrained to be filler trials with feedback presented on the first 20 trials.

2.1.3. Participants

Participants were 98 students at Carnegie Mellon University who received course credit or a small payment for participation. All participants reported normal hearing and English as their native language.

2.2. Results and discussion

Eighteen participants were excluded from analyses because their critical item accuracy was below 80%: low accuracy may indicate a hearing problem or low motivation; further, since reaction time analyses were based on correct trials only and there was a maximum of 10 trials per condition per participant, low accuracy makes individual reaction time measures unstable. Analyses including these participants showed the same pattern as described below but the additional noise made the results less reliable. Critical item accuracy and response time (RT) for the remaining 80 participants are shown in Fig. 1. The top panel in Fig. 1 shows that phonemes were detected more accurately in words than nonwords ($F[1, 76] = 15.4$, $p < 0.001$), but there was no interaction with attention condition ($F < 1$) or any other main effects or interactions (all other $F < 1$).

Response times were measured from target offset and only trials on which the participant provided the correct response were included in analyses. Overall, the word advantage was greater in the high lexical attention (80% words) condition (119.4 ms faster phoneme detection in words relative to non-words) than in the low lexical attention (20% words) condition (36.5 ms faster phoneme detection in words relative to non-words). Full ANOVA results of RT showed a main effect of lexical status (i.e., phoneme detection was faster in words than non-words; $F[1,76] = 17.2$, $p < 0.001$) and a lexical status by attention condition interaction, indicating that the word advantage was bigger in the high lexical attention condition than in the low lexical attention condition ($F[1, 76] = 4.87$, $p = 0.03$). No other reliable effects were found (all other $F < 1$).
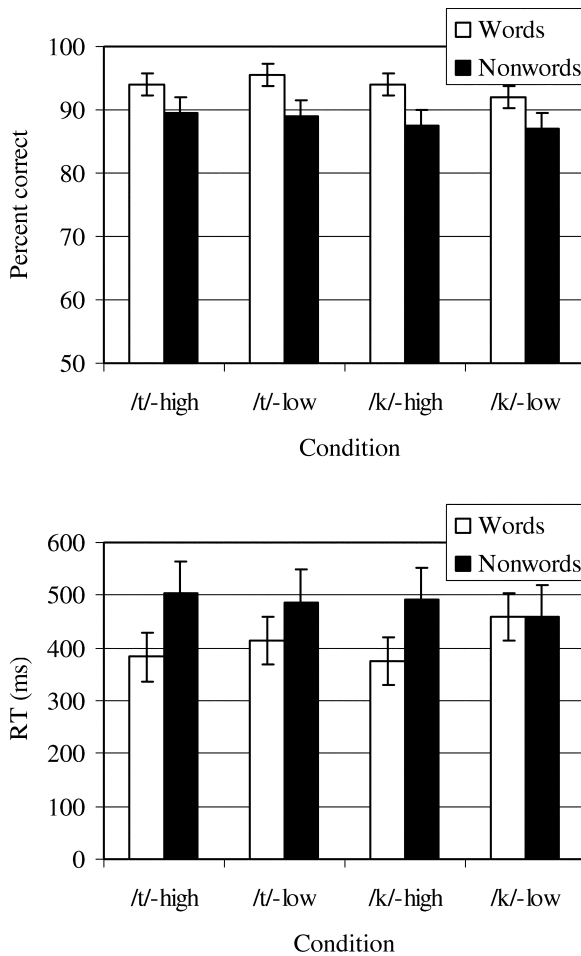
Fig. 1. Experiment 1: Mean accuracy (top panel) and response time (bottom panel) for recognition of /t/ and /k/ in words and non-words under high and low lexical attention. Error bars reflect ±1 standard error.

When proportion of words was low, lexical information was generally less helpful to task performance, which led to a reduction in lexical attention and consequently a reduction in the magnitude of the word advantage. Importantly, manipulating the proportion of words preserved the critical stimuli and task across attention conditions, allowing the conclusion that the decrease in the word advantage was due to decreased lexical attention. This is the first demonstration of attentional modulation of lexical effects in phoneme detection in which the critical stimuli and task are identical across attention conditions.

There is a possible alternative account of these results: high proportion of words provided more practice detecting phonemes in words and thus reduced word RT and increased the word advantage; low proportion of words provided more practice detecting phonemes in non-words and thus reduced non-word RT and decreased the word advantage. Although it cannot be ruled out as a possibility, this account requires that the mechanism of phoneme detection in words

be at least substantially distinct from that of phoneme detection in non-words so that practice with one stimulus type does not transfer to the other. This constraint seems unparsimonious given that the TRACE model and other related models can account for these effects without invoking such distinct mechanisms. Furthermore, the attention interpretation of reduced word advantage at low proportion of words is consistent with previously demonstrated effects of proportion of words manipulations, such as increased word reading errors on inconsistent words (Monsell et al., 1992), decreased word bias in speech errors (Hartsuiker et al., 2005), decreased lexical neighborhood effects (Vitevitch, 2003), and increased verbal short-term memory errors on words (Jeffries et al., 2006). It is unclear how these additional effects could be accounted for simply by invoking a differential practice effect for words vs. non-words. In sum, although we cannot rule out the practice effect account on empirical grounds, such an account of the present results and previous findings appears to require further motivation and development before it could be viable. As a result, we interpret our findings as demonstrating an effect of attentional modulation of lexical effects in speech perception.

In light of this demonstration of attentional modulation of lexical effects, it is important to consider the mechanisms by which attention might influence speech processing and whether attentional modulation requires a departure from the principle of bidirectional information flow between lexical and prelexical representations. In the following sections, insights from visual attention are developed into a conceptual framework for incorporating effects of attention on speech perception, two concrete implementations of this framework are proposed and tested in simulations, and conflicting predictions from these implementations are tested behaviorally.

3. Attention mechanism and implementations

A large body of empirical and theoretical research on visual attention has suggested that neural representations of stimuli that are attended are more active than representations of stimuli that are not attended. Single-unit recording studies of visual attention in monkeys have found that neurons are less responsive when their preferred stimulus is a distractor that must be ignored than when their preferred stimulus is the target (e.g., Moran & Desimone, 1995). Similarly, functional magnetic resonance imaging (fMRI) studies in humans have found that neural activity in motion-responsive area MT is reduced when participants are instructed to ignore moving stimuli (O'Craven, Rosen, Kwong, Treisman, & Savoy, 1997). Although less work has explored attention in language processing, a recent magnetoencephalography (MEG) study found that the N400m response to syllables was stronger (more word-like) when these syllables were presented among words and sentences compared to when they were presented among other syllables only (Bonte, Parviainen, Hytonen, & Salmelin, 2006; for a review and interpretation of word recognition studies using MEG see Pylkkanen & Marantz, 2003). This finding suggests that syllables activated word representations when the overall proportion of words was high but not when the proportion of words was low. This result could be due to damping of lexical representations under conditions that favor low lexical attention.

In models of speech processing, the principle of increased neural response to attended stimuli may be modeled by attentional modulation of the excitability of the lexical layer. That is, task or stimulus conditions that cause participants to direct attention to lexical information

may cause an increase in activation of mental representations for words and task or stimulus conditions that cause participants to direct attention away from lexical information may cause an overall decrease in activation of mental representations of words. In an interactive model such as TRACE (McClelland & Elman, 1986), lexical feedback to speech sound processing is proportional to lexical activation. Thus, modulating lexical layer activity would modulate lexical effects on speech processing. This approach does not alter the interactivity of the system—lexical information feeds back to earlier levels of processing at all degrees of attention; but due to attentional modulation, there is simply less lexical activation to feed back when attention is directed away from lexical information and consequently weaker lexical effects on phoneme processing.

The principle of selective attentional modulation of activation is quite general and can be implemented in many different models and in different ways within a model. To examine concrete implementations of this general mechanism, we extended the TRACE model of speech perception (McClelland & Elman, 1986) to include two different implementations of attentional modulation of lexical layer activity. The TRACE model consists of processing units grouped into an acoustic/articulatory feature level, a phonemic level, and a lexical level. Mutually consistent units on different levels (e.g., /k/ as the first phoneme in a spoken word, "kiss" as the identity of the word) activate each other via excitatory connections and mutually inconsistent units within the same level (e.g., /k/ vs. /g/ as the first phoneme) compete through mutually inhibitory connections.

One way to modulate the lexical layer's activation is to modulate its responsiveness to input. In the TRACE model, as in the original interactive activation model of visual word recognition, the change in activation of a unit is a function of the net input to that unit and the unit's current activation state relative to its maximum and minimum activation levels (see McClelland & Rumelhart, 1981, for details). Modulation of responsiveness to net input was implemented by adding an attentional scaling parameter (α) to the function specifying the net input to a lexical unit:

$$net\_input = \alpha \left(\left[\sum a_p {}^* W_{p \to l} \right] + \left[\sum a_l {}^* W_{l' \to l} \right] \right) \tag{1}$$

Here the portion in parentheses is the standard net input equation that is based on feed-forward input from the phoneme layer (first term) and inhibitory lateral interactions within the lexical layer (second term). When $\alpha = 1.0$, this is the standard TRACE model as implemented by McClelland and Elman (1986); when $\alpha < 1.0$, net input is scaled down and thus lexical responsiveness is damped and lexical effects should be reduced. The implementation of gain as a multiplicative scaling of the net input has been used by other researchers to model the effects of attention at the level of neuromodulation (e.g., Servan-Schreiber, Printz, & Cohen, 1990) and at the level of strategic control (e.g., Kello & Plaut, 2003).

Another way to modulate lexical layer activity is to manipulate a global external input to lexical units. This global input was represented as a constant input to each lexical unit on each processing cycle, just like an additional input source that has constant activity. We treat the standard TRACE model as representing a condition of high lexical attention (standard lexical attention should be relatively high since outside the laboratory most speech input is known words) and so simulate lower levels of lexical attention by adding a negative bias to the

net input of each unit. We do not wish to suggest that low lexical attention represents active inhibitory damping, only that the relative level of global excitation is reduced when lexical attention is low compared to the situation where lexical attention is high.

$$net\_input = \left(\left[\sum a_p{}^* W_{p \to l} \right] + \left[\sum a_l{}^* W_{l' \to l} \right] \right) - \beta \qquad (2)$$

Similar implementations have been used in other models of attentional modulation (e.g., Cohen, Dunbar, & McClelland, 1990), and this implementation accords well with the neurophysiologically based biased competition theory of attention (Desimone & Duncan, 1995). Each of these implementations of modulation of lexical attention was tested in the context of two classic lexical effects on speech perception: lexical bias on identification of ambiguous phonemes and the word advantage in phoneme detection (as in Experiment 1).

4. Simulation 1: Identification of ambiguous phonemes

The finding that ambiguous phonemes tend to be perceived such that they form a word (e.g., an ambiguous /g/-/k/ sound is heard as /g/ when followed by "ift" and as /k/ when followed by "iss"; Ganong, 1980) provides a simple test bed for examining the implementations of attentional modulation described above. Some studies suggest that this effect is modulated by lexical attention (see Pitt & Samuel, 1993, for review and meta-analysis of this effect). That is, under conditions favoring lexical attention there is a robust lexical influence, but under conditions disfavoring lexical attention, the influence is reduced or nonexistent. For the simulations, late-occurring phonemes in relatively long words (5–7 phonemes) were replaced with ambiguous phonemes to test the lexical influence. Two ambiguous phonemes were tested: a fricative (which could be interpreted as either /s/ or /ʃ/) and a stop (which could be interpreted as either /t/ or /d/). For each ambiguous phoneme an equal number of lexical contexts for each interpretation were tested (4 for /s/, 4 for /ʃ/; 5 for /t/, 5 for /d/; see Appendix B for the full list of simulation materials). The simulations were carried out with high and low levels of lexical attention in each implementation. For a baseline high attention simulation, α and β were set to 1.0 and 0.0, respectively, so that the TRACE model would exhibit the previously reported lexical bias effect. Two low attention simulations were carried out to test net input gain and negative bias manipulations independently. In these simulations, α or β was set to 0.1 to reduce lexical attention and the other was held at the baseline high attention level. Standard values for all other parameters were used (McClelland & Elman, 1986; Mirman et al., 2005).

Activations for lexically consistent and inconsistent interpretations of ambiguous phonemes are shown in Fig. 2. For both implementations of attentional modulation, when lexical attention was high, the lexically consistent phoneme won quickly and clearly for ambiguous phoneme input. When lexical attention was low, the lexically consistent phoneme had a smaller advantage and this advantage was slower to build up. Initial phoneme activation was driven by bottom-up input from feature units, but as phoneme units became active and began to activate word units, feedback activation from word units provided additional excitatory input to their constituent phonemes. Thus, for ambiguous phonemes, initially both interpretations became active but as word activation ramped up, feedback began to drive the lexically
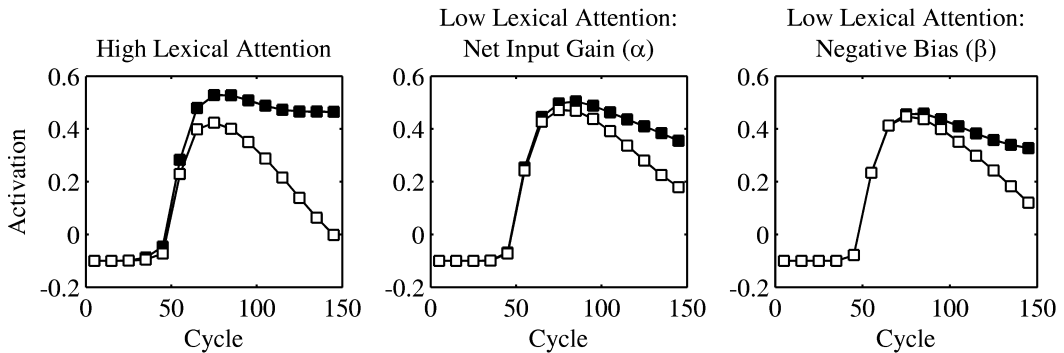
Fig. 2. Two implementations showed attentional modulation of the lexical influence on identification of ambiguous phonemes. Lexically consistent phonemes (black) were more active than lexically inconsistent phonemes (white), but this difference was smaller under low lexical attention (middle and right panels) compared to high lexical attention (left panel).

consistent phoneme ahead of the lexically inconsistent phoneme and lateral inhibition between phonemes enhanced this lexical advantage. When there was less lexical activation (due to attentional damping), there was less support for lexically consistent phonemes; consequently, the lexical advantage was smaller. The size of the lexical influence depended on the specific attention parameter values (Simulation 2 demonstrates this point explicitly).

These simulations showed that both implementations of attentional modulation can produce modulation in the lexical influence on interpretation of ambiguous phonemes.

5. Simulation 2: Word advantage

Simulation 2 was based on the word advantage in phoneme detection (i.e., faster phoneme detection in words than non-words). In Experiment 1 the word advantage was bigger under high lexical attention than low lexical attention (consistent with Cutler et al., 1987; Eimas et al., 1990). The inputs were 14 words and 14 non-words with the target phoneme (/t/) in the final position (see Appendix B for full list of simulation materials). Simulation non-words were created by swapping the onsets between words, as in Experiment 1, to control for phonotactic effects. Four different attention values were tested for each of the implementations. To capture the density of lexical neighborhoods, and thus the structural similarity between the non-words and words, an expanded 600-word lexicon was used. The lexicon was constructed by choosing all the words in the CMU Pronouncing Dictionary that were composed of the 14 phonemes defined in TRACE and cross-checking the words against an American English dictionary to eliminate items such as proper names and technical terms.

Simulated phoneme detection RT was computed as number of processing cycles from target phoneme onset required for the target phoneme unit to reach a 0.9 response probability threshold according to the R. D. Luce (1959) choice rule (as in previous studies using the TRACE model; McClelland & Elman, 1986; Mirman et al., 2005). Model RT for the words and non-words are shown in Fig. 3 for each of the tested attention values in each of the
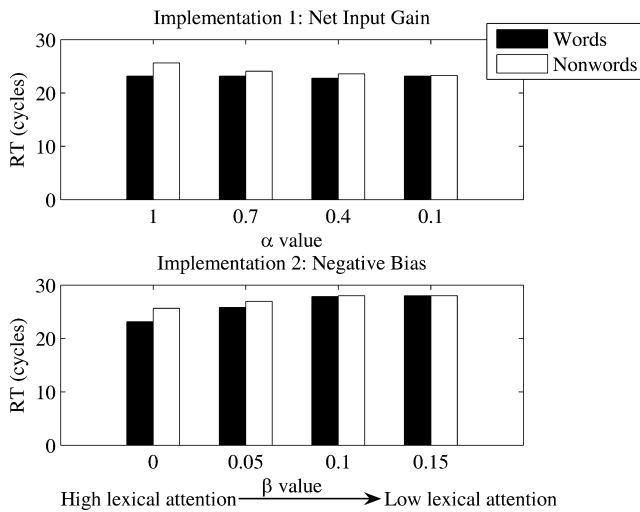
Fig. 3. Two implementations showed attentional modulation of the word advantage in phoneme recognition. The TRACE model recognized phonemes more quickly in words (black bars) than non-words (white bars), but this difference decreased as lexical attention decreased.

implementations. For each implementation, at high lexical attention, TRACE was faster to detect phonemes in words than in non-words. As lexical attention was reduced, lexical items were less active and so provided less support to their constituent phonemes; thus, the word advantage decreased at lower lexical attention values. This result is consistent with the behavioral data from Experiment 1.

Both implementations showed a graded decrease in word advantage with decrease in lexical attention, but there was a subtle difference between them. When net input gain (α) was reduced (Fig. 3, top panel), the word advantage was eliminated due to a RT decrease for non-words. In contrast, when negative bias (β) was increased (Fig. 3, bottom panel), RT for words and non-words increased and the word advantage was eliminated by virtue of the greater RT increase for words. This difference was due to the effect of the different implementations on the dynamics of activation and competition at the lexical layer. Modulation of net input gain made individual lexical units less responsive to all inputs: the lexical units were slower to become active in response to excitatory input and were less inhibited by lateral interactions. This decrease in competitiveness among lexical units allowed many lexical units to remain active at low activation levels rather than forcing a single unit to dominate activity. The population of active words then tends to reinforce activation at the phoneme level about equally for both words and non-words (recall that the words and non-words were matched in phonotactic probability).

An excitatory gain manipulation (i.e., gain on just the excitatory inputs) was implemented to test whether the net input gain results were due to gain effects on responsiveness to excitatory or inhibitory inputs. When the excitatory-only gain parameter was reduced (lower lexical attention), network performance matched the results of simulations manipulating negative bias (i.e., RT increased for phonemes in words and non-words). This result contrasts with the RT decrease found for the net input gain implementation, indicating that the effect of net input gain on RT must be due to lexical units' decrease in responsiveness to lateral inhibitory input.
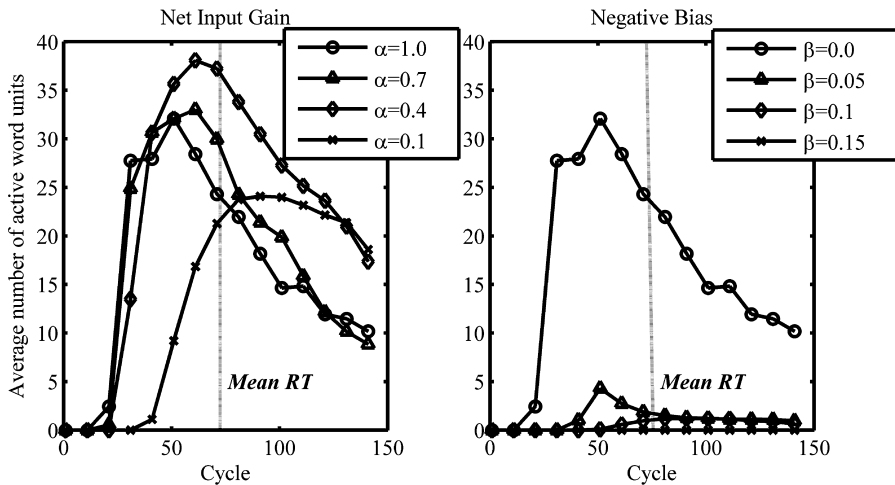
Fig. 4. Number of word units active (above 0.05) as a function of lexical attention manipulated by net input gain (left panel) and negative bias (right panel). Reduction of lexical attention by net input gain tended to increase active lexical neighborhood size (this pattern did not hold at the lowest level of α, though a relatively large lexical neighborhood did become active even at this low level of lexical attention). Reduction of lexical attention by negative bias rapidly decreased active lexical neighborhood size.

The reduction in sensitivity to lateral inhibition leads to lexical feedback that is dominated by the cumulative effects of many words rather than the activity of the single best matching word.

Figure 4 shows the effect of net input gain and negative bias manipulations on the number of word units that become active above a minimal threshold (0.05; a word unit must rise from a rest activation of -0.1 to above 0.0 activation to begin interacting with other units). As net input gain is reduced (lower lexical attention), more word units become active. At the lowest gain value ($\alpha = 0.1$), lexical units are slow to become active and the active lexical neighborhood is smaller, but a large number of lexical units are still able to reach the 0.05 level. In contrast, as negative bias is increased (lower lexical attention), the number of lexical units to pass the activation threshold very quickly drops to one or zero. In Fig. 4 the mean model RT is noted to emphasize that, at the point of phoneme recognition, reduction in lexical attention due to net input gain tends to increase the size of the active lexical neighborhood, but reduction in lexical attention due to negative bias severely decreases the active lexical neighborhood.

The excitatory-only gain test and the number of active words data (Fig. 4) demonstrate that as net input gain (α) is decreased, lexical feedback becomes dominated by the cumulative effects of many words rather than the activity of the single best matching word. At high lexical attention the high activation of a single matching word provided facilitative feedback to phonemes in that word, giving rise to a word advantage, but at low lexical attention, no single lexical item could reach high activation levels (due to decreased net input gain). As a result, feedback was equally supportive of phonemes in words and in non-words because the test non-words were specifically designed to be as similar to the overall population of

words as the test words were (i.e., the words and non-words were matched on phonotactic probabilities). One of the interesting properties of the TRACE model is that it is sensitive to the overall statistics of the lexicon as well as the effects of individual items (McClelland & Elman, 1986), but this dual sensitivity depends on the balance of excitatory and competitive dynamics, which is disrupted by manipulation of net input gain. In contrast, manipulation of bias appears to reduce group and individual item effects together.

These simulations demonstrate neurophysiologically plausible mechanisms that are consistent with interactive processing and show that different implementations offer somewhat different accounts. Under the net input gain implementation, reduced lexical attention produced faster recognition of phonemes in non-words and no change in RT for words. Under the excitatory gain and negative bias implementations, reduced lexical attention produced slower response times overall, particularly for words. The consequences of the two implementations make conflicting behavioral predictions regarding RT as lexical attention is decreased: net input gain predicts faster RT for non-words and no change in RT for words (or slightly faster RT) but negative bias predicts slower RT for both words and non-words (particularly for words). In Experiment 1, in the low lexical attention condition relative to the high lexical attention condition, there was a 57-ms increase in RT to phoneme targets in words ($t(78) = 1.27$, $p = 0.21$) and a 25-ms decrease in RT to phoneme targets in non-words ($t(78) = 0.43$, $p = 0.67$). The results for words and non-words point in different directions but since neither effect is significant it appears that they simply do not provide evidence that can be used to distinguish between the two model implementations. In Experiment 2, proportion of words was manipulated as in Experiment 1, but with increased power to test the conflicting predictions from the two implementations of attentional modulation.

6. Experiment 2

Experiment 1 demonstrated that manipulation of proportion of words is an effective method of manipulating lexical attention—and consequently lexical feedback effects—while keeping critical stimuli and task constant. However, the results of Experiment 1 could not conclusively distinguish between the net input gain and negative bias computational implementations of attentional modulation. Simulation 2 showed that attentional modulation by net input gain predicts no change in RT to phonemes in words as a function of lexical attention, but attentional modulation by negative bias predicts slower phoneme detection, particularly for words, under lower lexical attention. Experiment 2 was designed to test this difference with a more powerful manipulation than Experiment 1. By focusing on just the words and excluding the critical non-word items, it was possible to strengthen the proportion of words manipulation (in Experiment 1 the manipulation was 80% vs. 20% words, in Experiment 2 it was 100% vs. 20% words) and to increase the number of critical items per participant (from 10 to 20). In Experiment 1, we avoided similarity priming effects by using a two-list design to ensure that individual participants would not hear both members of a word–non-word pair such as "hemlock" and "lemlock"; in Experiment 2, each participant could hear the complete set of 20 critical words because the derived non-words were not in the stimulus set.

This approach is preferable to the complementary focus on just the non-words because changes in RT to phonemes in non-words would be more difficult to interpret. A decrease in phoneme detection RT in non-words in a high non-word proportion block (relative to a low non-word proportion block) may be due to lexical attentional effects or may be due simply to familiarity with monitoring for phonemes in non-words (i.e., a practice effect). Some studies suggest that phoneme monitoring in non-words requires additional processing resources (e.g., Wurm & Samuel, 1997); possibly due to the fact that typical participants are unused to processing non-words. As a result, it is possible that this additional demand will be reduced with practice, that is, in blocks with a high number of non-words, thus masking an increase in RT. This non-word–specific practice effect is independent of a general task familiarity effect, which would of course affect both words and non-words. In addition to possible non-word–specific practice effects, researchers have found that phonological and lexical effects in non-words are less robust (e.g., Lipinski & Gupta, 2005) and less predictable (e.g., P. A. Luce & Large, 2001) than effects in words. In sum, Experiment 2 was designed to examine the effect of attentional modulation on phoneme detection in words in order to distinguish between two possible computational implementations of attention in speech perception.

6.1. Methods

The stimuli and procedure from Experiment 1 were used in Experiment 2, but the design of the blocks was changed to increase manipulation power by focusing on phoneme detection in words. All blocks contained 20 critical (target-present) words and 20 target-absent filler words. The high lexical attention block also contained the 120 attention-shifting words and the low lexical attention block contained the 120 attention-shifting non-words. The experiment began with a 40-trial practice session (during which feedback was provided) that consisted either entirely of words (high lexical attention condition) or entirely of non-words (low lexical attention condition). Thus, the high lexical attention condition was 100% words and the low lexical attention condition was 20% words with the initial 40 trials all non-words to induce a strong attention shift before any critical items were presented. As in Experiment 1, phoneme target (/t/ or /k/) and attention condition (high or low) were manipulated between participants.

6.2. Participants

Participants were 29 students at University of Connecticut who received course credit for participation. All participants reported normal hearing and English as their only language.

6.3. Results and discussion

One participant was excluded due to having mean response times more than two standard deviations above the mean. No participants had critical item accuracy lower than 80% (the
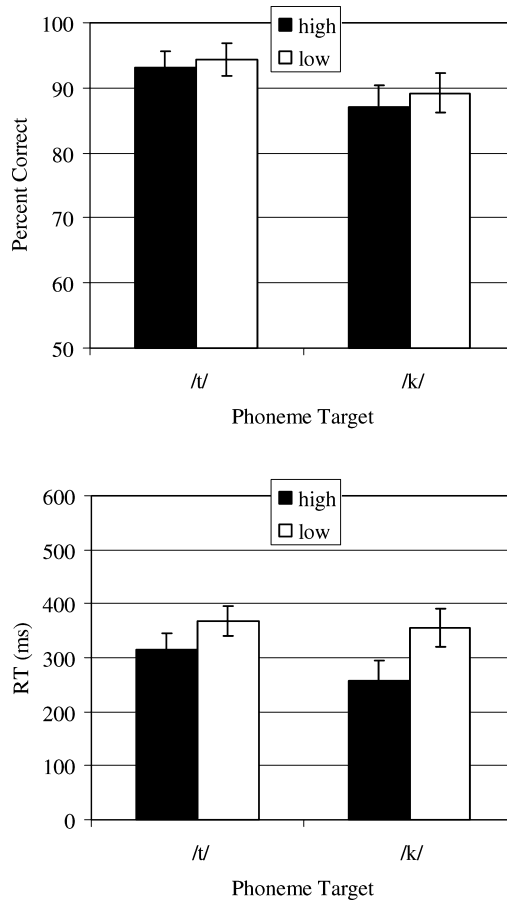
Fig. 5. Experiment 2: Mean accuracy (top panel) and response time (bottom panel) for recognition of /t/ and /k/ in words under high (black bars) and low (white bars) lexical attention. Error bars reflect ±1 standard error.

main exclusion criterion for Experiment 1), possibly due to more feedback during the practice session (in Experiment 1 feedback was provided only on the first 20 trials, in Experiment 2 feedback was provided on the first 40 trials). Fig. 5 shows critical item accuracy and response times for the remaining 28 participants. Participants were marginally more accurate at detecting /t/ than /k/ (F[1, 24] = 3.8, p = 0.06), but there was no reliable effect of attention condition on accuracy and no attention condition by phoneme target interaction for accuracy (both Fs < 1).

Response times were measured from target offset and only trials on which the participant provided the correct response were included in analyses. Phonemes were detected more quickly under high lexical attention (293.5 ms) than under low lexical attention (362.3 ms) (F[1, 24] = 5.35, p = 0.03). There was no reliable difference in response times between the two phoneme targets (F[1, 24] = 1.22, p = 0.28) and no attention condition by phoneme target interaction (F < 1). This finding is consistent with the negative bias implementation of

attentional modulation, which predicted slower response times to detect phonemes in words under lower lexical attention, and conflicts with the net input gain implementation, which predicted no change in response times under lower lexical attention.

Experiment 2 provides behavioral evidence that distinguishes between two possible computational implementations of attention: net input gain and negative bias. The net input gain implementation is not consistent with these data. This evidence does not specify the nature of attentional modulation of speech perception, but it does provide constraining evidence toward understanding the effect of attention on speech perception. Any model or theory that predicts no change or faster phoneme detection in words under lower lexical attention is inconsistent with these behavioral data.

7. Summary and conclusions

This report described a test of attentional modulation of lexical feedback effects using a paradigm in which critical stimuli and task are identical across attention conditions. In this paradigm, attention is shifted between lexical and prelexical levels by the proportion of words relative to non-words among the filler items, thus allowing tight control of critical stimuli and task across attention conditions. Eliminating stimulus and task differences between attention conditions eliminates alternative explanations so the results can be more confidently attributed to attentional modulation. Experiment 1 demonstrated that manipulating the proportion of words in an experimental block produces attentional modulation of the word advantage effect in phoneme detection. Other researchers have attributed variability in lexical feedback effects to attentional modulation, but this is the first demonstration in which the critical stimuli and task were held constant across attention conditions.

To account for attentional modulation of lexical feedback effects, we proposed a general mechanism of attentional modulation based on selective modulation of lexical activation. This approach is consistent with neurophysiological studies of visual attention and with a recent MEG investigation of speech perception. The TRACE model of speech perception was extended to include two concrete implementations of this mechanism: net input gain and a global external input to all lexical units. Both of these implementations are consistent with the principle of interactive processing and simulations demonstrated that each of these implementations can account for attentional modulation of lexical effects on speech sound recognition. The approach and implementations are intended to generalize across specific lexical effects; the present work described tests in the context of two classic lexical effects: lexical bias in identification of ambiguous speech sounds (Simulation 1) and faster recognition of speech sounds in words than non-words (Simulation 2).

Critics (Norris et al., 2000) of the interactive view of speech perception have argued that interactive models cannot account for such attentional modulation effects without removing their interactivity. This criticism overlooks the possibility that, consistent with neurophysiological evidence, the effect of attention could take place at the lexical layer itself. The present simulations demonstrate that attentional modulation of lexical activation dynamics can produce modulation of lexical feedback effects while leaving the interactive architecture

intact (for a recent review of interactive processes in speech perception see McClelland et al., 2006).

The simulations also showed that different implementations offer somewhat different accounts: under the net input gain implementation, reduced lexical attention reduced sensitivity to lateral inhibition between lexical items, thus allowing large lexical neighborhoods to dominate lexical feedback. Since the words and non-words were matched in phonotactic probability, lexical neighborhood feedback led to faster recognition of phonemes in non-words and no change in RT for words. Under the excitatory gain and negative bias implementations, individual item support for phonemes in words was reduced along with a reduction in lexical neighborhood feedback to phonemes in both words and non-words. Consequently, response times increased overall, particularly for words. Experiment 2 tested these conflicting predictions of response time to detect phonemes in words using a more powerful version of the paradigm of Experiment 1. The behavioral results were consistent with the negative bias implementation of attentional modulation and conflicted with the net input gain implementation.

In the present work we have focused on a particular attention effect (modulation of lexical effects on speech perception) and a particular model (TRACE), but the key principle is very general. For example, applying our approach to the semantic layer in the triangle model of word reading (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996) would account for increased word reading errors on inconsistent words under low attention (Monsell et al., 1992). It is also possible to generalize our approach to models that learn contextual representations and have no explicit "word" level. For example, in the case of simple recurrent networks (SRNs; Elman, 1990), the attention effect would be applied to the context layer. This approach points toward a computational account of attention on verbal short-term memory (Jeffries et al., 2006) in the context of a recurrent model of serial order recall (Botvinick & Plaut, 2006). More generally, our approach is based on previous work on the Stroop effect (Cohen et al., 1990), which has become a critical paradigm for studies of selective attention (e.g., MacLeod & MacDonald, 2000), and our approach shares the central principle of the biased competition theory of attention (Desimone & Duncan, 1995).

In sum, the present computational investigations comprise an important step toward an understanding of the effects of attention on language processing, including a clear demonstration that attentional modulation is consistent with interactive processing. The behavioral experiments provide both a paradigm for testing attention modulation effects and some evidence elucidating the consequences of modulation of lexical attentional on speech perception and constraining further development of models of attention in speech perception.

Acknowledgments

help with model implementation, Joseph Stephens for recording the stimulus tokens for the experiments, Christi Gomez and Ann Kulikowski for their help in collecting the behavioral data, and Nicole Landi and three anonymous reviewers for their comments on an earlier draft of the manuscript.

References

Bonte, M., Parviainen, T., Hytonen, K., & Salmelin, R. (2006). Time course of top-down and bottom-up influences on syllable processing in the auditory cortex. *Cerebral Cortex, 16*, 115–123.

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review, 113*(2), 201–233.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*(3), 332–361.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology, 19*(2), 141–177.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193–222.

Eimas, P. D., Hornstein, S. M., & Payton, P. (1990). Attention and the role of dual codes in phoneme monitoring. *Journal of Memory & Language, 29*(2), 160–180.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*(2), 179–211.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception & Performance, 6*(1), 110–125.

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related reply to Baars et al. (1975). *Journal of Memory & Language, 52*(1), 58–70.

Jefferies, E., Frankish, C. R., & Lambon Ralph, M. A. (2006). Lexical and semantic binding in verbal short-term memory. *Journal of Memory and Language, 54*(1), 81–98.

Kello, C. T., & Plaut, D. C. (2003). Strategic control over rate of processing in word reading: A computational investigation. *Journal of Memory & Language, 48*(1), 207–232.

Lipinski, J., & Gupta, P. (2005). Does neighborhood density influence repetition latency for nonwords? Separating the effects of density and duration. *Journal of Memory and Language, 52*(2), 171–192.

Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes, 16*(5/6), 565–581.

Luce, R. D. (1959). *Individual choice behavior*. Oxford, England: John Wiley.

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4*(10), 383–391.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences, 10*(8), 363–369.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I An account of basic findings. *Psychological Review, 88*(5), 375–407.

Mirman, D., McClelland, J. L., & Holt, L. L. (2005). Computational and behavioral investigations of lexically induced delays in phoneme recognition. *Journal of Memory & Language, 52*(3), 424–443.

Monsell, S., Patterson, K. E., Graham, A., Hughes, C. H., & Milroy, R. (1992). Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), 452–467.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*(4715), 782–784.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral & Brain Sciences, 23*(3), 299–370.

O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A., & Savoy, R. L. (1997). Voluntary attention modulates fMRI activity in human MT-MST. *Neuron, 18*(4), 591–598.

Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception & Performance, 19*(4), 699–725.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56–115.

Pylkkanen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences, 7*(5), 187–189.

Rubin, P., Turvey, M. T., & van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics, 19*(5), 394–398.

Servan-Schreiber, D., Printz, H., & Cohen, J. D. (1990). A network model of catecholamine effects: Gain, signal-to-noise ratio, and behavior. *Science, 249*(4971), 892–895.

Vander Wyk, B., & McClelland, J. L. (2004, August). *Toward a graded phonology*. Poster session presented at the 26th Annual Meeting of the Cognitive Science Society, Chicago, IL.

Vitevitch, M. S. (2003). The influence of sublexical and lexical representations on the the processing of spoken words in English. *Clinical Linguistics & Phonetics, 17*(6), 487–499.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory & Language, 40*(3), 374–408.

Wurm, L. H., & Samuel, A. G. (1997). Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring. *Journal of Memory & Language, 36*(2), 165–187.

Appendix A: Critical items for experiments

Table A1 contains word and non-word critical items used in Experiment 1 (word items from this list were critical items for Experiment 2). Non-words were designed by swapping onsets between words and matching average onset-vowel occurrence as measured by (a) number of words with the specific onset-vowel in the corpus, (b) the sum frequency of the onset-vowel occurrence, and (c) the sum natural log transformed frequency of onset-vowel occurrence. The CMU Pronouncing Dictionary was used for these analyses and the set was limited to two-syllable, initial stress words (to match the critical items); conditions were equally well matched according to analyses of the full corpus.

Table A1
Critical items for experiments. Onset-vowel occurrence means and standard errors (in parentheses) are included at the bottom

/t/		/k/	
Words	Non-words	Words	Non-words
bracelet	nacelet	cynic	mynic
debit	sebit	fabric	gabric
Words	Non-words	Words	Non-words
faucet	braucet	frolic	molic
forfeit	plorfeit	garlic	marlic
gadget	padget	gimmick	simmick

Table A1

Critical items for experiments. Onset-vowel occurrence means and standard errors (in parentheses) are included at the bottom (Continued).

	/t/		/k/	
	Words	Non-words	Words	Non-words
	habit	mabit	gothic	lothic
	limit	gimit	graphic	haphic
	magnet	lagnet	haddock	laddock
	nugget	sugget	havoc	mavoc
	pamphlet	hamphlet	hemlock	lemlock
	planet	vanet	lilac	gylac
	private	spivate	logic	hogic
	profit	pofit	lyric	syric
	pulpit	dulpit	magic	gragic
	rabbit	fabbit	mimic	fimic
	senate	venate	Mohawk	pohawk
	spirit	firit	music	pusic
	summit	rummit	panic	hanic
	velvet	prelvet	public	gublic
	visit	prisit	sumac	frumac
Count	105.50	98.25	112.30	113.50
	(14.73)	(15.91)	(14.18)	(13.86)
Frequency	989.65	858.10	657.20	715.95
	(276.34)	(276.11)	(156.46)	(171.86)
Log-freq	22.49	21.16	22.08	24.44
	(3.69)	(3.60)	(3.77)	(4.26)

Appendix B: Stimuli for simulations

Word contexts used for simulation 1

Fricative-final
 /s/-bias: decrease, produce, carcass, glorious
 /ʃ/-bias: abolish, brackish, publish, galosh

Stop-final
 /t/-bias: abrupt, carpet, secret, biscuit, product
 /d/-bias: crooked, regard, placid, solid, garbled

Stimuli used for simulation 2

 Words: ballast, basket, carpet, Charlotte, culprit, deduct, depart, dulcet, gasket, goblet, product, redact, sculpt, tablet
 Non-words: dallast, rasket, tarpet, barlotte, dulprit, geduct, Shepart, gulcet, prasket, skoblet, koduct, kedact, dulpt, bablet