Laboratory for cognitive neuroscience of language
Department of Psychology, University of Connecticut

Technical report 2006.01
Original version, 14 November 2006

"Hindering" effects of feedback in the TRACE model of spoken word recognition: Simulations of phonological gang and embedding effects

Ted Strauss
ted.strauss@uconn.edu

Abstract
Results of simulations are presented that illustrate lexical conditions that encourage phonological gang effects, and related inhibitory effects in the TRACE model of speech processing. The report is structured as follows. An introduction outlines the context within which the simulations were prepared. Details of simulation setup are given, results presented, and a short discussion. The report functions as an extended appendix to Magnuson, Strauss, Harris and Mirman (submitted).

Introduction

A debate over modularity wtihin the speech comprehension system was reinvigorated by the claim that “feedback never helps” (Norris et al., 2000). On one side, the **interactive** activation framework permits continuous flow of information bi-directionally between prelexical and lexical stages of processing. Lexical effects on sublexical processing are explained in terms of top-down excitation, for example from word to phoneme. On the other, the **autonomous** framework adopts a modular view in arguing optimal processing and veridical perception can only be achieved by enscapsulating successive processing layers from higher levels of representation (see Magnuson et al., submitted, for more background).

In support of autonomomy, Norris et al. (2000) cite a TRACE simulation done by Frauenfelder & Peeters (1998); TRACE is the representative interactive-activation model. In the simulation, a set of 21 words was run once with lexical-to-phoneme feedback on and again with feedback off. There was no general advantage of feedback since about half the words were recognized more quickly with feedback and half were recognized more slowly with feedback. The conclusion drawn by Norris et al. was that if feedback appears to not help speech processing in general then its sole purpose is in explaining lexical effects within the interactive activation framework; and this is an unparsimonious explanation of speech processing.

In response to Norris et al., Magnuson, Strauss, Harris & Mirman (submitted; hereafter MSHM) ran new TRACE simulations and argued that Frauenfelder and Peeter’s simulations do not present the full picture of feedback’s role in speech processing. There were two principle results. First, when every word in a 901-item lexicon was run once with feedback and once without, the majority of recognized items were responded to faster with feedback. Second, when noise was added to the input, the presence of feedback mitigated the deleterious effect of noise on speech processing. According to MSHM’s simulations, feedback is helpful in general as well as in the case of noisy input. Therefore, contrary to Norris et al.’s characterization, feedback does help speech processing in the interactive-activation model TRACE.

The purpose of this report is to give further details about some secondary results from MSHM’s simulations. While 73% of the 718 words were recognized faster with feedback, 9% were recognized *more slowly* with feedback. Understanding why feedback could ever be inhibitory is a key question. MSHM concisely summarize the simulations reported here that dig into this question, and lead to some new insights into the structure of lexical neighborhoods.

Honing in on the feedback disadvantage question

It is clear from MSHM’s results that lexical to phoneme feedback *can* speed recognition in TRACE. As a word unit receives bottom-up activation, top-down connections send excitation down to the phoneme units that compose the word, which in turn reinforce the word; this positive feedback loop continues as the activation of the target word reaches ceiling levels. Without feedback, the target word tends to reach ceiling activation somewhat more slowly. This pattern is observed in the majority of cases.

In contrast, 9% of words were recognized more slowly with feedback on. Without yet knowing why one word versus another should have a feedback disadvantage, the network dynamics involved in either case are understood. Schematically, (i) lexical feedback causes a word to boost its constituent phonemes, (ii) those phonemes boost words that contain them, and (iii) all words compete with one another, and since the inhibition word nodes send to other word nodes is proportional to their activation, the most active words exert the most competition. The boost in (ii) has the capacity to facilitate recognition, and the competition in (iii) has the capacity to inhibit target recognition. When feedback is turned off these dynamics are muted; phoneme to word activation still occurs, but there is no extra boost contributed from feedback.

The fact that feedback usually helps recognition indicates that the boost in (ii) is usually more potent than the competition in (iii). Therefore, in feedback disadvantage cases it must be that the competition overwhelms the boost. Our question then becomes: under what conditions does feedback lead to increased lexical competition such that it can overwhelm the facilitatory effect of feedback.

Several factors can be identified, all having to do with the relationship between the target word and its competitor set (a subset of the lexicon). The factors are: cohort size, target length, phonetic salience, initially embedded words and negative gang effects. No single factor is predictive of a feedback disadvantage. Rather, two or more factors must be present for a disadvantage to occur.

Simulations

Simulation software. TRACE simulations were prepared and run in jTRACE (Strauss et al., in press).

TRACE parameters. For simulations described in this report, the standard TRACE parameter set was used. Following Frauenfelder & Peeters (1998), the manipulation of feedback uses either the pair [0.03, 0.00] for smaller lexicons or [0.015, 0.00] for larger lexicons.

Decision rule. A decision rule is used to decide when activation within the word layer should be treated as an occurance of word recognition. The rule used is the Goodness-of-fit rule described in Frauenfelder and Peeters (1998). This rule is used to maintain consistency with the approach of the earlier simulations. It works as follows.

Since each word unit in TRACE is reduplicated for all temporal positions, the rule must choose which temporal alignment is to be used to represent the word. The strategy with this rule is to use the alignment known to be perfectly aligned with the input, in this case alignment 4. Response probabilities were then calculated for each word at each TRACE processing cycle using the Luce (1959) choice rule:

$$R_i = \frac{e^{ka_i}}{\sum e^{ka_j}} \qquad (1)$$

where R_i is the response probability for item i, a_i is that item's activation in TRACE, k is a constant, set to 20, that controls target-competitor separation, and the summed activations in the denominator include all target and competitor units. As in the Frauenfelder & Peeters (1998), an item was considered recognized when its response probability exceeded a threshold of 0.9.

Simulation paradigm To investigate the subtle effects of target-competitor interaction mediated by feedback, a simulation paradigm was developed as follows. A starting lexicon was created and a target word selected. New words were added to the lexicon one at a time. As each new item was added to the lexicon, recognition of the target word was simulated once with feedback and once without. This allowed me to examine the effect of feedback as the target's competitor set changed.

A simple example of this paradigm would be to create a starting lexicon that contains only one word with /s/ as onset, and choose this word as the target. Therefore, initially the target's starting cohort size equals one. Next, one increases the cohort size by adding /s/ onset words one at a time, and simulating the target once with feedback and once without each time a word is added. The result reveals the effect (and interaction effect) of cohort size and feedback on response time.

Though this simulation paradigm involves a changing lexicon, it is not intended to imply anything about language development. The changing lexicon is simply a method to observe subtle changes in processing as a result of changes to the target word's competitor set.

Variations of this *one-at-a-time* paradigm are used to study the feedback disadvantage factors listed in the introduction.

Simulation 1 – Phonetic salience, cohort size & target length

Phonetic salience implies that certain speech sounds are more readily perceived than others. In TRACE, variation in phonetic saliency affects recognition and has the potential to mediate feedback

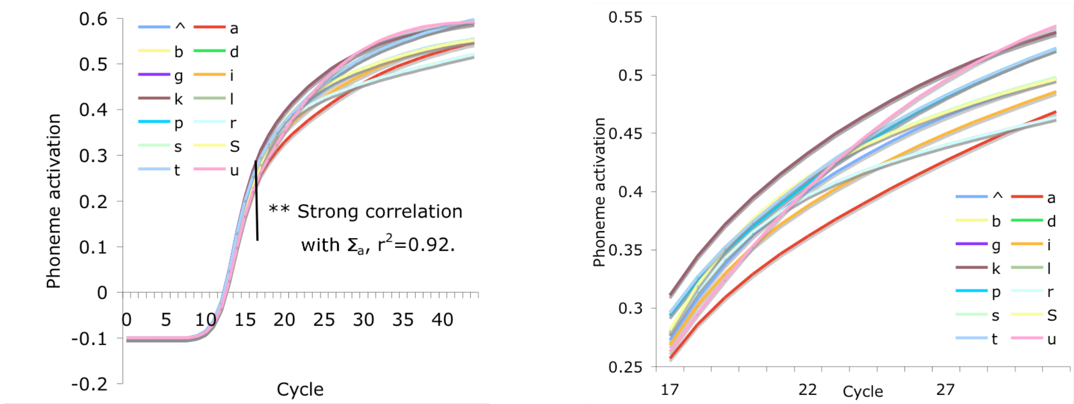

Figure 1 – Simulation of each TRACE phoneme in isolation. Note slight differences in time course of activation.

effects, in some cases causing feedback disadvantages. This section describes simulations that back up these claims.

Simulation 1.1 The first simulation very simply illustrates that TRACE's 14 phonemes are not processed identically. Each phoneme was simulated in isolation and the time-course of phoneme activation recorded. Figure 1 shows the rate at which each phoneme is activated.

Variations in rate of activation for isolated phonemes reflect slight differences in phoneme representations. Phonemes are specified by seven features, each of which is a 9 point vector, for a total of 63 values per phoneme. The extent to which an input pattern matches a phoneme template determines how much activation the phoneme will receive. Table 1 compares the phoneme templates for two phonemes, /t/ and /a/. The representations of the two phonemes do not overlap at all, therefore the same input will never activate both of them, i.e. they are not ambiguous with respect to one another. Other phonemes are mutually ambiguous, e.g. /t/ and /k/, and this ambiguity does affect processing.

	/t/ (Σ_v= 7.2)									/a/ (Σ_v= 6.4)								
POW	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	0	0	0	0
VOC	0	0	0	0	0	0	0	**1**	0	**1**	0	0	0	0	0	0	0	0
DIF	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0
ACU	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	**0.1**	**0.3**	**1**	0
GRD	0	0	0	0	0	0	0	**1**	0	0	**1**	0	0	0	0	0	0	0
VOI	0	0	0	0	0	0	0	**1**	0	**1**	0	0	0	0	0	0	0	0
BUR	0	0	**1**	**0.2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 1 – Phoneme representations for /t/ and /a/. Summing over all feature vectors, /t/ is more "salient" by providing more input. This is reflected in the isolated phoneme simulations, figure5.

The sum of the phoneme vectors may differ; /t/ totals 7.2 and /a/ totals 6.4. Therefore, more featural information is generated to create an instance of /t/ versus /a/. These global differences in phoneme representation do affect processing, as figure 1 shows that /a/ is activated more slowly than /t/. Indeed, in repeated correlation analyses, it was found that the summed phoneme representation score (Σ_v in table 1) is highly correlated with the phoneme activation plots, especially in the early part of the simulation. At cycle 18, correlation between Σ_v and phoneme activation values approaches linear at r^2=0.92. The correlations become weaker before and after cycle 18.

Simulation 1.2 Simulation 1.1 shows that the details of TRACE's phoneme representation do affect processing in predictable ways. The next simulation attempts to observe whether the small variations in phoneme processing translate into differences in word processing; and whether any insights into feedback disadvantages can be gleaned from phoneme saliency affects. Along the way, evidence is collected that implicates cohort size and target length in the occurance of feedback disadvantages.

The simulation was designed as follows. A lexicon of 560 items was created that distributed TRACEs 14 phonemes evenly. There were 140 items each of 3,4,5 and 6 segments in length. Each of TRACEs phonemes started 40 words (40x14=560). Second, third, fourth, fifth and sixth segments were randomly assigned, such that an equal number of each phoneme occurred at those positions. Therefore, each phoneme occurred the same number of times over all positions in the lexicon. After words were generated randomly, duplicate entries were filtered and replaced with words that maintained the designed distributions. The result was a lexicon that should not contain biases for any particular phonemes. Note that randomly generating words meant that there was no syllable structure.

The one-at-a-time simulation paradigm was used in the following way. For each phoneme, e.g. /t/, all members of the /t/ cohort were removed from the lexicon. One member of the /t/ cohort was selected, e.g. /tril/; it was inserted into the lexicon and treated as the target. A pair of simulations was run, once with feedback (0.015) and once without (0.00). A randomly chosen member of the /t/ cohort was inserted into the lexicon and the pair of simulations was run again. Then another member of the /t/ cohort was inserted, and so on for all members of the /t/ cohort (each cohort has 40 items). These 80 simulations were repeated for all members of the /t/ cohort; that is, each /t/ word got a turn as target. This procedure was repeated for all 14 phonemes. In total 44,800 (14 x 40 x 40 x 2) simulations were run. The results implicate cohort size, target length, and cohort phoneme as factors contributing to feedback disadvantages.

The dependant variable that we are concerned with in these results is the ratio of response time with feedback to response time without feedback; RT ratio equals $RT_{0.015}$ / $RT_{0.0}$. Each RT ratio is calculated from a single pair of simulations, where the only difference is the value of the feedback parameter. A RT ratio greater than 1.0 indicates a feedback disadvantage, a ratio less than 1.0 indicates a feedback advantage, and a value equal to 1.0 means that response times were identical with and without feedback. What factors in simulation 1.2 contribute to the occurance of feedback disadvantages?

Figure 2 shows that cohort size and target length are direct contributors to feedback disadvantage effects. The graph shows mean RT ratio averaged over the four target lengths at increasing lexicon sizes, from successive insertion of cohort members. As cohort size increases, RT ratio increases. And as target length increases, RT ratio increases. This pattern is supported by a correlation analysis. Mean RT ratio is correlated with cohort size, r=+0.65, and mean RT ratio is correlates with target length, r=+0.65.

Figure 3 suggests that the effect of cohort size and target length on RT ratio is not equivalent for all phonemes. The two upper graphs show mean RT ratios grouped by length and focussing on targets from cohort groups /t/ and /k/, both of which evince a feedback disadvantage for longer targets and as the cohort increases in size. The two lower graphs – cohorts /l/ and /a/ – show cohorts that rarely undergo feedback disadvantages, regardless of target length and cohort size. What we see is that some phonemes are subject to phoneme disadvantages while others are unlikely to be.

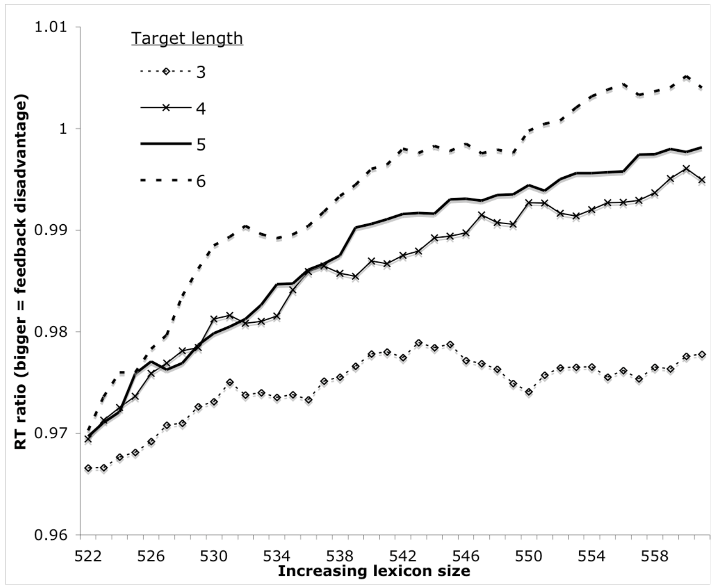

Figure 2 – Results of simulation 1.2 grouped by target word length, and presented as cohort members are iteratively inserted into the lexicon. Each point of this plot is the mean of 140 ratio values. A ratio equals $RT_{0.015}$ / $RT_{0.00}$ for a pair of simulations.

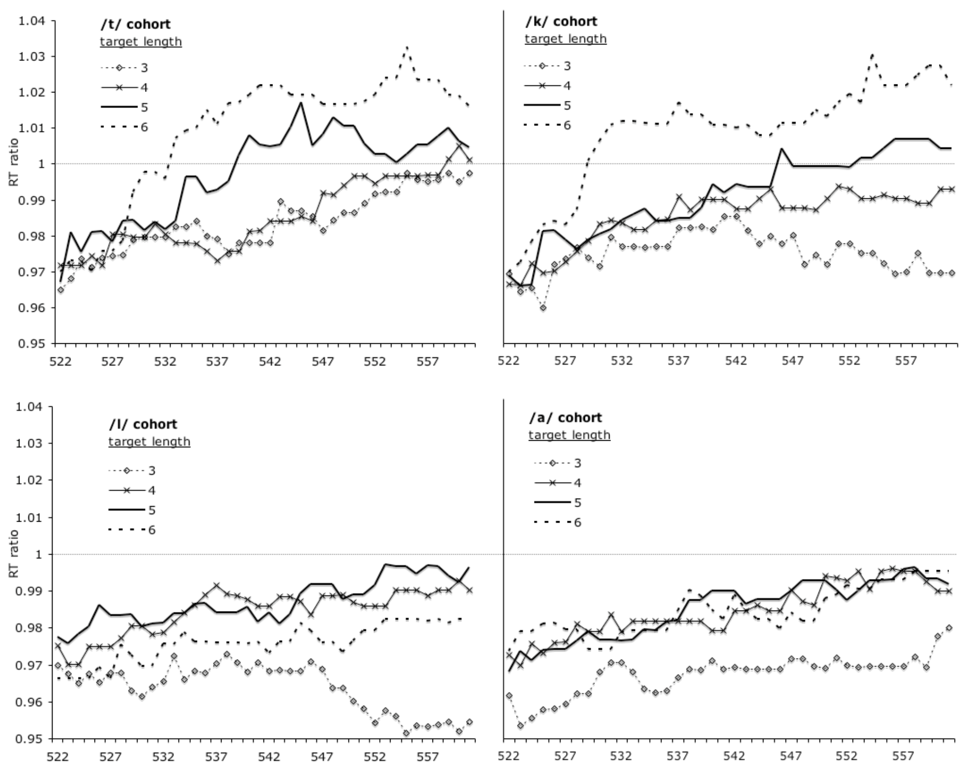

Figure 3 – Result of simulation 1.2 grouped by cohort and word length. Each graph point here is averaged over 10 RT ratios. Feedback affects different cohort groups inconsistently. In the two upper graphs – cohorts /t/ and /k/ - feedback in conjunction with longer targets and larger cohorts causes feedback disadvantages (i.e. RT ratio > 1). In the lower graphs – cohorts /l/ and /a/ - feedback disadvantages never obtain. In conjunction with a correlation analysis, these results suggest that more salient phonemes are more prone to feedback disadvantages.

To what extent can this pattern of feedback disadvantage be predicted from the saliency of the initial phoneme. Our earlier measure of saliency was Σ_v, the sum over phoneme vectors. I calculated the correlation of Σ_v to the proportion of feedback disadvantage occurances per cohort group (out of 1600 RT ratio scores per cohort group). The result was a moderate correlation, r=+0.64. This suggests that more salient phonemes, like /t/ and /k/, are more likely to undergo feedback disadvantages than less salient phonemes, like /l/ and /a/.

Simulation 1 offered evidence that larger cohort size, longer target length and greater phonetic saliency all contribute to the occurance of feedback disadvantages.

Simulation 2 - Initially embedded words

If the competitor set for a target contains a word that is embedded in the target, this contributes to the target having a feedback disadvantage. For example, ‘cow’ is initially embedded in ‘couch’. Because of competition between the carrier and embedded words, ‘couch’ –the carrier– may be recognized more slowly. The RT disadvantage affects the carrier due to a short word advantage in TRACE (longer words have more inhibition sites). In any event, the disadvantage incurred by having an embedded word is aggravated by feedback.

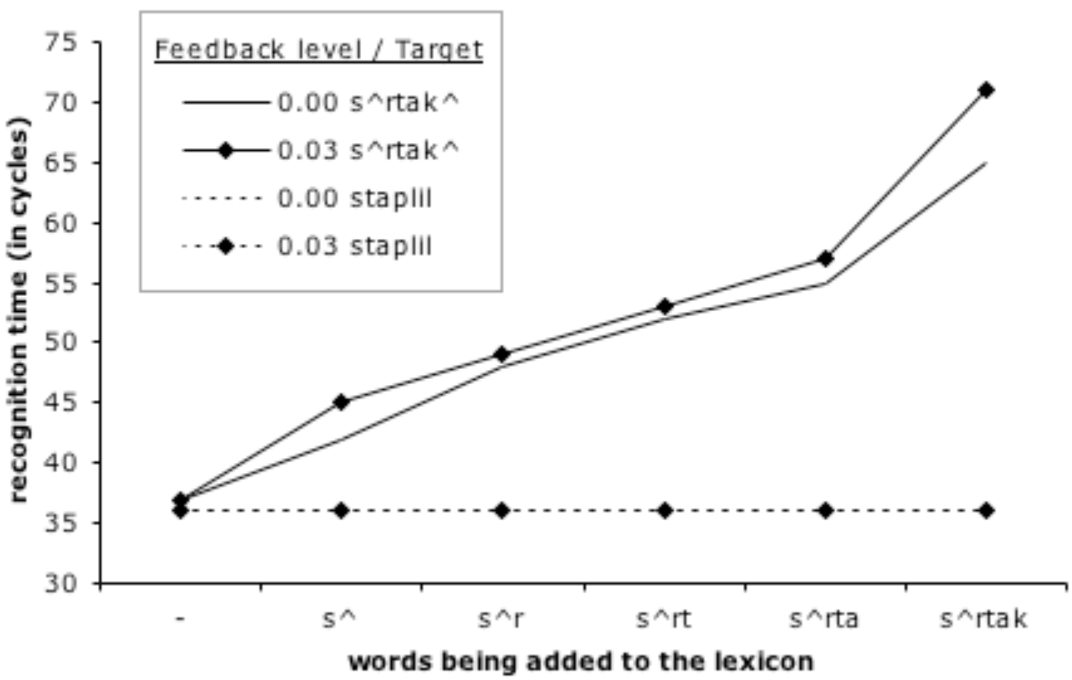

Figure 4: Illustration of initial embedding as an NFA factor.

As an illustration of this factor, consider the batch of simulations described in figure 4. A small lexicon was designed with 80 items. Two invented target words were considered, /s^rtak^/ and /staplil/. The initial lexicon was designed such that the two targets were balanced on cohort and neighborhood size. Recognition of the targets was simulated with (0.03) and without (0.00) feedback. Neither item showed an effect of feedback, in that recognition time was the same with and without feedback.

Next, a single item /s^/ was inserted into the lexicon, and the four simulations were repeated. Note that /s^/ is embedded in /s^rtak^/ but not in /staplil/. The result was a large overall increase in RT for /s^rtak^/, which was most pronounced with feedback on.

Recognition times for /staplil/ were not affected. By inserting a single item into the lexicon, the feedback disadvantage was triggered. Further simulations inserted /s^r/, /s^rt/, /s^rta/, and /s^rtak/ (all embedded in /s^rtak^/) one at a time into the TRACE lexicon, each time repeating the four simulations. The feedback disadvantage persisted, /s^rtak^/ was recognized more and more slowly, and the ultimate feedback disadvantage was 6 cycles. /staplil/ was (surprisingly) never affected by the addition of these words to the lexicon.

These simulations show that the presence of an initially embedded word can slow recognition of a target, especially when lexical feedback is on.

Simulation 3 - Negative gang effects

A negative gang effect occurs when a target word's competitor set has a phonological 'center of gravity' that is dissimilar from the target word. Consider English words that (in some dialects) begin /kæ/, like carrot, carry, Carolina, California, can, candle. Many of the words in this set have liquids or nasal alveolars as the third segment, so we'll say (as a simplification) that this set's center of phonological gravity includes a voiced alveolar in the third segment. If the input is /kæk^l/ (cackle) the third segment /k/ (a voiceless velar) diverges from this center. A positive gang effect facilitates performance when the input conforms to the pattern (center of gravity) of the competitor set, and a negative gang effect inhibits performance when the input diverges from the pattern. Therefore, if the input is /kæk^l/, a negative gang effect obtains during perception of the third segment.

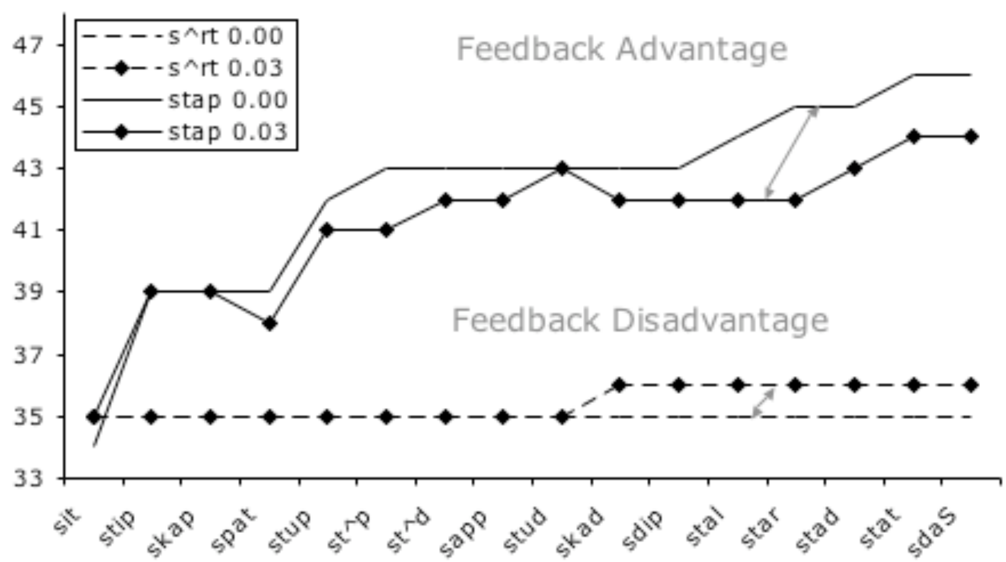

Figure 5: Illustration of Gang effects as an NFA factor.

Figure 5 describes a batch of simulations where I found negative gang effects to play a role in feedback disadvantages. An initial lexicon is prepared with two items balanced on frequency and cohort. Words are then inserted into the lexicon one at a time to cause a negative and positive gang effect, respectively. The target words are /s^rt/ and /stap/. The left edge of the plot shows that both are initially recognized quickly with or without feedback.

The items inserted into the lexicon (shown on the x-axis) were phonologically similar to /stap/. The criterion used for similarity requires that two items have the same initial phoneme, and that the feature matrices of the second phonemes correlate by at least 0.55, using the phoneme correlation matrix provided in McClelland and Elman (1986), reproduced here in figure 6. The third phoneme, similarly must correlate by at least 0.55. In the present simulation these conditions are satisfied for the relation between /stap/ and each inserted word, shown on the x-axis, therefore we predict a positive gang effect. These conditions are never satisfied for the relation between /s^rt/ and inserted words (although the third segments are sometimes correlated), therefore we predict a negative gang effect.

As a result of the growth of /stap/'s competitor set, its recognition time increases, but with feedback on a positive gang effect obtains, facilitating recognition. /s^rt/ is phonologically dissimilar to the gang of items being added incrementally, and it's recognition time is not generally affected, except for a small negative gang effect. Once half of the gang has been added /s^rt/ is recognized more slowly with feedback on.

The result indicates that the same mechanism in TRACE that brings about facilitatory gang effects for some items causes inhibitory gang effects for others. The next simulation describes factors that mediate the occurance of gang effects.

Phoneme	p	b	t	d	k	g	s	S	r	l	a	i	u	^
p	—	.76	.71	.56	.60	.46	.30							
b	.76	—	.56	.71	.46	.60								
t	.71	.56	—	.76	.56	.42	.35							
d	.56	.71	.76	—	.42	.56								
k	.60	.46	.56	.42	—	.77		.24						
g	.46	.60	.42	.56	.77	—								
s	.30		.35				—	.65						
S					.24		.65	—					.20	
r									—	.80	.29		.32	.37
l									.80	—	.32			.32
a									.29	.32	—	.65	.75	.67
i											.65	—	.65	.49
u								.20	.32		.75	.65	—	.59
^									.37	.32	.67	.49	.59	—

Note. Correlations of less than .20 have been replaced by blanks.

Figure 6 – Correlations of feature patterns of phonemes used in TRACE.

Simulation 4 – Gang effects, mediating factors

Further simulations implicate cohort size and target length as factors that mediate gang effects, such that large gang effects are triggered only under very specific circumstances. Cohort size and target length, though not factors that directly trigger feedback disadvantage cases, can create the necessary conditions for their occurrence. This section describes further simulations to this effect. Figure 7 describes a batch of simulations consisting of gang effects ranging in magnitude.

The largest and clearest gang effect (in the lower-left corner) occurs when the target word is short (3 segments) and initial /bl/ cohort is of moderate size (about 30 words). The y-axis in each figure describes RT ratio of RT with feedback (0.03) and RT without feedback (0.00). As before, the x-axis corresponds to the addition of words to the lexicon, one by one. The solid lines traces the effect of adding a consonant gang to the lexicon (i.e. one that is phonologically similar to the target word), and the dotted line traces the effect of adding a dissonant gang to the lexicon (i.e. one that is similar to a competitor that is in the target's cohort; in this case the dissonant gang is based on bli, blir, blirS). In all six figures, the same gangs are inserted for the solid and dotted lines, respectively.

The key pattern to look for in each figure is how widely the two lines diverge from one another and from the zero line. The principle of the gang effect dictates that the two lines should originate at zero (i.e. no gang effect when there are no gang items), the dotted line should monotonically increase and the solid line monotonically decrease. The fact that the ideal pattern only occurs in one of six figures shows that the gang effect is a complex convergence of factors. In this case, a moderate sized cohort (size 30 in a lexicon of size 430) and short target word (3 segments) is the sweet spot. The consonant gang yields a final advantage of 10 cycles and the dissonant gang yields a final disadvantage of 7 cycles.

Moving across the figure, as th e length of the target word increases the divergence of positive and negative gang effects is weakened. And, when the size of the initial cohort is reduced to 2 (only the target and competitor in the cohort), again the divergence of gang effects is weakened. In the upper-right figure, the two lines almost completely overlap. The results show that negative gang effects are mediated by target length and cohort size.

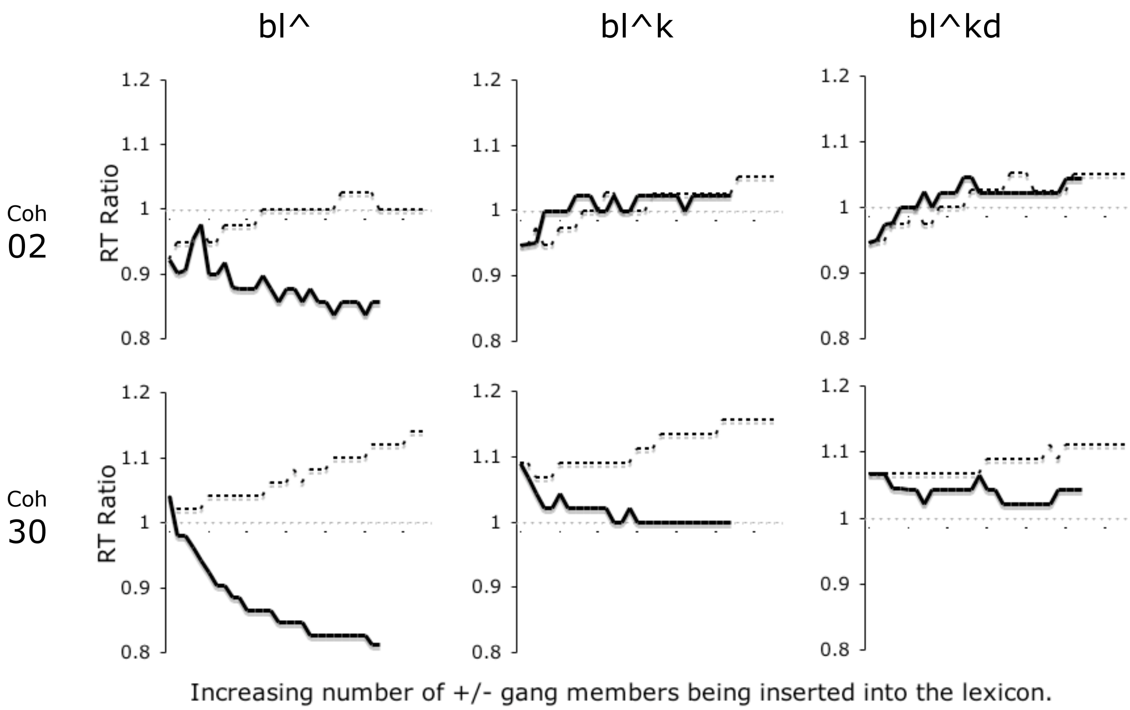

Figure 7: Gang effects mediated by cohort size and target length.

Discussion

The simulations presented demonstrate circumstances under which feedback disadvantages can occur in TRACE. For any language with a relatively constrained phonetic space, densely packed phonological neighborhoods, and mostly short words, these types of feedback disadvantages will affect a minority of words. Negative gang effects occur when the target's onset differs from the norm, embedding effects occur in longer words with at least one embedded word, and long words are less commmon. These circumstances are, by design, rare in the languages satisfying the above conditions.

Future research will attempt to pin down other factors that cause feedback disadvantages, and the precise combination of factors sufficient to elicit a feedback disadvantage. A formal understanding of feedback effects on recognition allows prediction of processing latency in comprehenders of different languages. Length, cohort, embedding and gang effects have all been attested in the literature (e.g. Marslen-Wilson, 1979, Grosjean and Gee, 1987, Pitt and Samuel, 2006), though there is not consensus on their origins in comprehension system. Precise prediction of facilitatory and inhibitory feedback effects is needed to confirm or disconfirm the interactive explanation of such effects.

References

Frauenfelder, U. H., & Peeters, G. (1998). Simulating the time course of spoken word recognition: An analysis of lexical competition in TRACE.; In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 101-146). Mahwah, NJ: Erlbaum.

Grosjean, F., & Gee, J. P. (1987). Prosodic structure and spoken word recognition. *Cognition, 25,* 135–156.

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

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

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

Pitt, M.A., & Samuel, A.G. (2006). Word Length and Lexical Activation: Longer is Better. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 1120-1135.

Strauss, T. J., Harris, H. D., & Magnuson, J. S. (in press). jTRACE: A reimplementation and extension of the TRACE model of speech perception and spoken word recognition. *Behavioral Research Methods*.