

Lexical contact during speech perception: A connectionist model

Eric Forbell (eforbell@bowdoin.edu)

Department of Computer Science; 8650 College Station
Brunswick, ME 04011

Eric Chown (echown@bowdoin.edu)

Department of Computer Science; 8650 College Station
Brunswick, ME 04011

Abstract

A connectionist architecture comprised of cell assemblies was developed and applied to the problem of speech perception at the phonemic and lexical levels. The problem addressed involved a disagreement amongst theorists over the possible sources of lexical priming effects. Speech was encoded in the model as the temporal activity of phoneme units that are connected to higher-level word assemblies. The lexical layer was topographically organized based upon similarity of phonemic structure. Lateral inhibition at the lexical level was shown to be both necessary and sufficient to support results from phonological priming experiments involving human participants.

Introduction

Speech processing represents a rich source of constraints for the development of neural models of cognition. These constraints are particularly challenging since they often arise out of the temporal nature of the task, a weakness of most connectionist models. This paper presents a connectionist model that addresses a long-standing matter of controversy in the psycholinguist literature involving a task that is highly temporal in nature.

The model task involves lexical contact, which is defined as the phase of speech processing whereby the representations activated by the speech input make initial contact with the lexicon. The data comes from a series of studies to test a variety of priming conditions (Hamburger & Slowiaczek, 1996; Slowiaczek & Hamburger, 1992). The basic design of these studies involved prime words that are related to target words by sharing initial phonemes (i.e. "black" and "bleed" share two initial phonemes). Primes are presented 500 ms. before the targets and the subject's only task, called shadowing, is to repeat the target word aloud as quickly and accurately as possible. The major result is that when three initial phonemes are shared, response time actually slows by about 40 ms. whereas there is very little effect when one or two phonemes are shared (Hamburger & Slowiaczek, 1996).

The model for these data is based upon a variant of Hebb's cell assembly called TRACE (Kaplan, Sonntag, & Chown, 1991). The cell assembly is particularly suited for this type of modeling since it is well grounded biologically and was originally proposed to address issues of temporal processing in a neural framework. TRACE modeled a single cell assembly and has since been extended to model multiple cell

assemblies in a more general system called multiTRACE (Chown, 1994; Sonntag, 1991).

Lexical and phonological priming

The data from Slowiaczek and Hamburger suggest that phonologically similar words compete at the lexical level of speech recognition (1996). Specifically, this competition was observed in a priming paradigm whereby the primes were phonologically related to target words by the number of initial phonemes. The critical number of overlapping initial phonemes was three, for primes that were presented 500 ms before the targets. Additionally, this data also elucidated the cause of a facilitatory effect (a decrease in RT) in the low-similarity case (1- to 2- phoneme overlap) that was presenting some difficulty in interpretation.

It was suggested from other studies that low-similarity facilitation only occurs if the phonological relatedness proportion (PRP) is high (50 %) (Goldinger, Luce, Pisoni, & Marcario, 1992). That is, the number of trials containing phonological overlap was manipulated across subject groups, and it was only in groups containing a majority of trials with phonological overlapping (high PRP) that displayed facilitation effects (Hamburger & Slowiaczek, 1996). Goldinger et al. explained this prelexical facilitation effect by suggesting that subjects were strategically assuming that the initial phoneme of the target would be the same as that of the prime, because the majority of previous trials had been this way (1992). When the experiment was controlled for subject expectancy, however, this facilitatory effect was virtually eliminated and only the 3-phoneme overlap interference was observed (Hamburger & Slowiaczek, 1996). Nevertheless, there remains a debate in the literature over whether either effect truly comes from competition among lexical candidates or whether it is simply an artifact of the experimental design.

A different kind of priming has been studied with a smaller interstimulus interval (the time between the end of the prime word and the beginning of the target) of 50 ms. Goldinger et al. found that for uncommon targets (low-frequency words) preceded by phonologically related primes, response time was increased (1992). This phonological inhibition result is included to provide a bigger picture of the priming literature, but was not a focus of this work.

The multiTRACE model

Hebb developed the cell assembly construct to address questions concerning the temporal nature of neural processing. Essentially a cell assembly is a large collection of neurons which act in concert and which have temporal extent due to their recurrent connections and their corresponding ability to “reverberate.” Hebb’s theory lost favor initially in part because he omitted inhibition, a construct for which there was no evidence at the time. More recently, however, cell assemblies have undergone something of a revival as advances in neuroscience have been incorporated in the theory (Kaplan, et al., 1991) and experimental evidence for their existence has been found (Amit, 1995).

In the Kaplan model of cell assemblies, called TRACE (Tracing Recurrent Activity in Cognitive Elements), the emphasis was on simulating the internal dynamics of a population of neurons that would comprise a cell assembly. In the TRACE model various neural control mechanisms were postulated to play different functional roles in the cognitive system. For example, inhibition is useful as a selection mechanism when multiple cell assemblies are competing to become active. A major addition to cell assembly theory by the Kaplan group was to add fatigue to counterbalance the reverberation inherent in a highly recurrent system.

TRACE, which serves as the basis for multiTRACE, uses a set of difference equations that are updated at each time step to model the collective behavior of a large group of neurons. The equations model various biological functions such as activity, neural fatigue, short-term connection strength, long-term connection strength, sensitivity to firing, and network or external input (Table 1).

Kaplan et al. argued that units built with these basic properties have a number of advantages over the simple units used in many traditional connectionist models (1991). Different levels of activity in a cell assembly, for example, can serve different cognitive purposes, such as coding for conscious versus unconscious processing. The major questions left open by the original work on TRACE was how the notion of a single cell assembly could be extended to the cognitive system as a whole.

Table 1: The basic multiTRACE equations

Update Equations	Delta Equations
$A(t+1) = A(t) + \Delta A$	$\Delta A = (A + \bar{A}I)\bar{A}V\Delta A^{\bar{L}} + A\bar{A}^{\bar{L}c}\bar{V}$
$F(t+1) = F(t) + \Delta F$	$\Delta F = \Delta_g A\bar{F}\Delta_d F$
$S(t+1) = S(t) + \Delta S$	$\Delta S = \Delta_g AS\Delta_d S$
$L(t+1) = L(t) + \Delta L$	$\Delta L = 0.0$
	$V = \frac{1}{V}(S + L)F$
	$I = I^{exc} \Delta I^{inh}$ (expanded in text)

Δ : unit loss
 Δ : inh. competition
 v : normal factor
 Δ_g : fatigue growth
 Δ_d : fatigue decline
 Δ_g : STCS growth
 Δ_d : STCS decline

A: activity
 F: neural fatigue
 S: short-term connection strength
 L: long-term connection strength
 I: network input

* \bar{X} denotes quantity $(1 - X)$

The multiTRACE model extends the cell assembly idea by building models with collections of cell assemblies. Sonntag originally created multiTRACE to study sequence learning in the context of cell assemblies (1992). Chown later extended the model to deal with other forms of learning, for example modeling the effects of the arousal system on learning (1994). The development of multiTRACE has been increasingly less abstract, starting from the very general problem of modeling sequences, to the current work which addresses a very specific body of data.

Applicability to lexical priming

The multiTRACE framework provides a natural way to model the lexical priming data presented in the previous section. Each phoneme and lexical unit is represented by a cell assembly as part of a hierarchical structure (Fig. 1). Phonemes which are part of a word are strongly linked to the lexical units at the higher level (e.g. the phoneme “b” will be strongly linked to the cell assembly representing “black” but not the one representing “flack”). The activation of the lexical units at the higher level corresponds to perception and therefore the ability of the subject to repeat the word. The theory is that competition between these units accounts for the differences in timing.

This sort of perceptual competition forms the basis of a number of connectionist models and stems from evidence that similar concepts tend to interfere with each other more than dissimilar ones as part of what Kinsbourne called “the functional cerebral distance principle” (Kinsbourne, 1982). This interference comes in the form of lateral inhibition between cells near each other in the cortex. The idea is simple; words that are similar (e.g. “black” and “blast”) will be stored in nearby locations in cortex, meaning that they will greatly inhibit each other. As one becomes highly active during perception it will naturally inhibit the other, making perception a kind of winner-take-all proposition. In terms of the brain, the cell assemblies underlying these representations will be close to each other. A given cell assembly will have a kind of inhibitory surround which will typically prevent its close neighbors from being simultaneously active (Fig. 1).

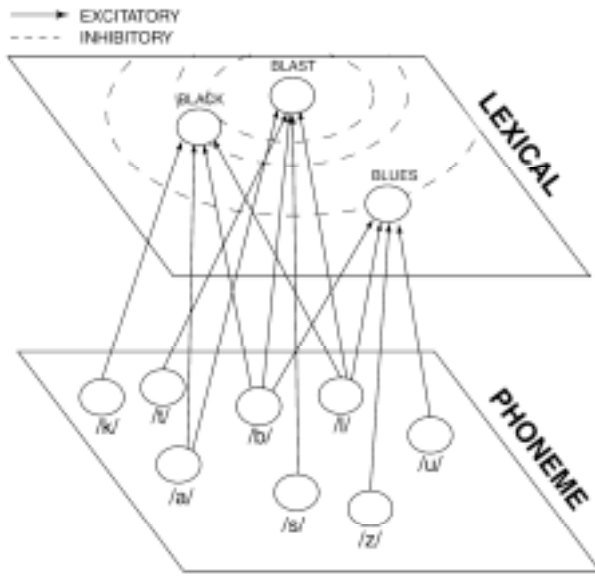


Figure 1: The two-tiered connectionist architecture. (Inhibitory output shown only for BLAST unit)

In the context of a cell assembly model the interference seen when the target word shares three phonemes with the prime must come from competition, and the competition must come from the prime word itself. Since the prime word and the target word share three initial phonemes in common they will be represented very near each other in the brain, and therefore they will have a great deal of lateral inhibitory connections. These data are very useful for extending cell assembly theory because they provide a number of useful constraints on the temporal dynamics of cell assembly activation, as well as information on the general layout of cell assemblies in the cortex.

The implementation

In the original implementations of multiTRACE it was not necessary to explicitly model lateral inhibition as the model was not applied to highly similar concepts as is the case here. In updating the model we devised a scheme, based upon Kinsbourne's functional distance principle, whereby the lateral inhibitory connections between cell assemblies was determined by the shared number of phonemes. This rule also reflects the spatial layout of inhibitory processes in association cortex, as well as areas like the early visual system, where lateral inhibition is widely known to exist and has important functional implications, such as enhancing contrast sensitivity.

The competition generated by lateral inhibition results from an active cell assembly. Since the priming effects vary according to the test conditions it is important to understand the factors that affect the time course of activity in a cell assembly. In Hebb's original formulation the only was a cell assembly's capacity for reverberation (in the multiTRACE model this comes in the form of internal long term connection strength or LTCS). Later through simulation it

was determined that inhibition was necessary as a control mechanism (Rochester, et al., 1956). More recently, Kaplan et al. (1991) did a series of simulations showing that with the addition of two more biologically and theoretically motivated mechanisms – fatigue and short term connection strength (STCS) – that it was possible to model different time courses of activity corresponding to different parts of the cognitive hierarchy. For example, cell assemblies near the perceptual interface would be expected to have a high refresh rate in order to be ready for the next input. On the other hand cell assemblies that participate in long term planning would be expected to stay active for longer periods of time. In our model we conceptually represent these differences as different hierarchical layers.

Although it is possible that hierarchy can emerge naturally in a flat network structure, we felt that such a network design would quickly become confusing and therefore limiting. Additionally, our network structure reflects the layered organization of human cerebral cortex. Groups of multiTRACE units were combined into layers in the current system, and static properties of individual units were inherited from their "parent" layers. It is our conjecture that nearby layers will have similar parametric properties (e.g. in the perceptual layer cell assemblies will all tend to have high refresh rates). The lexical priming data provides an excellent test of such conditions and the potential usefulness of the cell assembly construct, since it can be applied to model widely different types of cognitive functioning.

The basic structure of the simulation was based upon a two-tiered network with each tier representing a different level of the cognitive hierarchy. In this case, because each layer is part of the perceptual interface, they have virtually identical parameter settings (Table 2). The primary layer contained units responsive to phonological stimuli theoretically produced by the primary and secondary auditory cortices. The secondary layer comprised of lexical units that respond to the phonological structure of a spoken word, not its meaning. For example, a lexical unit for "blast" received equal vertical connections from the phonemic /b/, /l/, /a/, /s/, and /t/ units. Because the data from Hamburger and Slowiczek was not concerned with the typicality effect, we built in the assumption that all of the cell assemblies had the same internal connection strength (1996). Differences in typicality could easily be modeled by introducing variability in connection strength within cell assembly units.

Table 2: Layer parameters and timings

Parameter	Phonemic layer	Lexical layer
Fatigue growth	0.15	0.15
Fatigue decline	0.04	0.04
STCS growth	1.0	1.0
STCS decline	0.2	0.2
Activity duration	300 ms	700 ms
Fatigue recovery	500 ms	950 ms

* STCS: Short-term connection strength

Despite both layers of units being similar in their static properties, their differentiation in cognitive speed emerged as a function of the hierarchical structure. The units in the word layer received input from several phonemic units over the course of time, depending on the length of the spoken word. Since the mapping between layers was not one-to-one, average unit durations were 300 ms and 700 ms for the phonological and lexical layers, respectively (Table 2). Fatigue recovery times, being dependent on activity, were similarly proportioned.

The topological organization of the lexical layer was also an important component of the current model. Interference, as suggested by the Hamburger and Slowiaczek data (1996), results from competition at the lexical level and is a function of word-form similarity. Therefore, the lexical map was built using the amount of initial phoneme overlap between lexical units as a distance approximation in cognitive space. The amount of inhibition imposed on some target unit k is a function of the distance to a neighboring unit j , as well as this neighbor's activity and fatigue level:

$$I_{jk}^{inh} = \frac{A_j(1-F_j)}{D_{jk}} \quad (1)$$

j: source unit
k: target unit
D: distance
t: Time

The net inhibitory input for a lexical unit then becomes the combination of local inhibition and regional inhibition imposed on the layer. This regional inhibition is a positive feedback mechanism that controls the spread of activity in a given region, in this case a layer, and is based upon the total activity in that layer:

$$I_k^{inh} = \frac{1}{L} \left(\sum_{j=1}^n I_{jk}^{inh} \right) + G \left(\sum_{i=1}^n A_i \right) \quad (2)$$

n: number of units in a layer
G: global inhibition factor (0.5)
L: lateral inhibition factor (2.0)

The excitatory input to a cell assembly in multiTRACE is computed in a conventional connectionist manner. However, the sum of a unit's long-term connection strength and short-term connection strength to another unit serves as the weight value (w_{ij}) typically seen in most connectionist models:

$$I_{jk}^{exc} = (LTCS_{jk} + STCS_{jk}) A_j \quad (3)$$

$$I_k^{exc} = \sum_{i=1}^n I_{jk}^{exc} \quad (4)$$

j: source unit
k: target unit
n: number of incoming connections for unit k

Simulation design and procedure

As in the original experiment, four prime conditions were created: *no relation* and three degrees of *phoneme overlap* (1-3). The representative words for each condition are presented in Table 3; the actual words were chosen arbitrarily for demonstrative purposes.

Table 3: Simulated experiment design

Condition	Prime	Target
No relation	"dream"	"black"
1-phoneme overlap	"bind"	"black"
2-phoneme overlap	"blues"	"black"
3-phoneme overlap	"blast"	"black"

In order to simulate an incoming stream of speech, the phoneme units comprising the prime and target words were activated in a serial manner, separated by an interval ranging from 20-40 ms, with a greater spacing reserved for vowel sounds. Using this approximation method, the experiment was easy to simulate. The set of phonemes corresponding to the prime word were activated first, followed by a the target phoneme string 500 ms after the prime sequence had concluded.

The simulations were expected to show that there is a fundamental difference in processing between the high similarity (3-phoneme overlap) and low similarity conditions (no relation and 1-phoneme overlap). That is, the response time of the target word unit represented at the lexical level should be increased due to the earlier presentation of the prime word stimulus.

Results

Our initial experiments show the relevant trends in the data (Table 4); interference resulting from lexical competition was observed in the high-similarity conditions (2- and 3-phoneme overlap) and not in the low-similarity conditions (0- and 2-phoneme overlap). To date we have not replicated the exact time-course for this interference that was found in the behavioral evidence, but we have found that the general trends are simple to generate in the model. Since the model presented here is considerably simpler than that of its human counterpart, and does not take into account the effects of semantic top-down influence, for example, which may also affect timing, we do not wish to spend too much of our effort trying for an exact match at this stage.

Table 4: RT differences (experimental – control) in simulated and actual experiment

Condition	Simulation (ms)	Actual (ms)
No relation	-	-
1-phoneme overlap	0	-4
2-phoneme overlap	40	-8
3-phoneme overlap	190	36

In the 3-phoneme overlap condition, the competition between "black" and "blast" is striking (Fig. 2). The time

course of the prime word's activity is sufficiently slowed in this condition as well as in that of the target's. That is, be

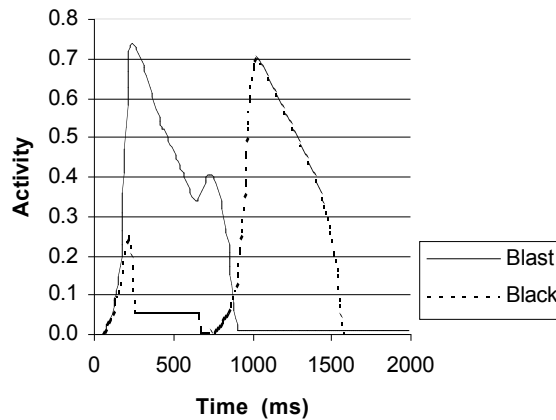


Figure 2: Competition between the “black” and “blast” lexical units

cause “blast” and “black” share three initial phonemes, each unit receives considerable input from the incoming speech stream. However, when the fourth segment of the speech stream (/s/) is presented to the phoneme layer, the net input for the “blast” unit begins to dominate over that of the “black” unit (Fig. 3). This initial competitive advantage is then reinforced by the increasing lateral inhibition “blast” is imposing on “black,” effectively increasing the competitive gap. However, the presentation of the target word 500 ms later in addition to the fatigue of the “blast” unit will allow “black” to win the second competition, albeit more slowly than in the control condition.

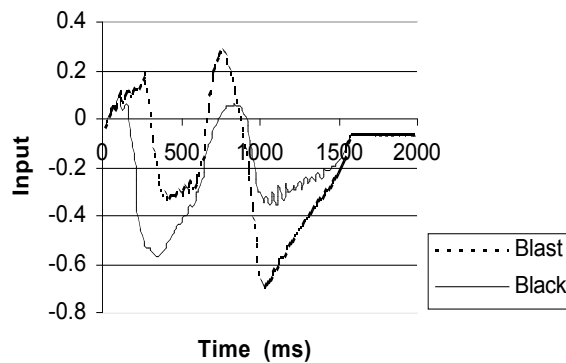


Figure 3: Net input for “blast” and “black” units

Conclusion

This work serves two purposes. First, we have presented a biologically grounded model that addresses a key controversy in the psycholinguistics literature. Our results support Hamburger and Slowiaczek’s theory that the lexical priming results can be explained in terms of competition between phonologically similar words. Second, their data

provides crucial constraints in exploring the temporal dynamics of neural processing in cognition.

For better or worse the dominant connectionist modeling paradigm has long been back-propagation. In recent years, however, interest has grown in recurrent models such as the one presented in this paper. The development of such models will be predicated upon their ability to account for psychological data with a temporal component. This work represents an important step in that direction. The data being modeled required incremental changes in an existing model. As the goal of our continuing research is not to alter previous work in order to support future data, this current work was successful in that components were identified—lateral inhibition and layering—that when inserted into the existing model were able to support the new data. Also, while these additions extend the modeling capabilities of current cell assembly implementations, they do not accomplish this at the cost of simplicity. That is, lateral inhibition and hierarchy fit very naturally into the multiTRACE model and are well supported theoretically.

With regards to the specific modeling task discussed here, future work will involve observing the competition dynamics as the scale of the system is increased. The high-similarity interference phenomenon was observed in a system constructed from roughly ten words, but now that the architecture is in place it will be possible to see how robust the effect will be as the number of words is increased. Another goal of this continued effort will be to reduce the amount of manual network design, because that has resulted in a more discrete representation of the lexical space under study than may be desired to obtain truly generalized conclusions. At the time of publication, however, a sampling of systems of roughly 160 to 200 words generated from a more normal distribution have produced statistically significant effects similar to those discussed here (including low-similarity facilitation), and it has been shown that the amount of inhibition at the lexical level is crucial to mimicking the behavioral evidence.

In addition to the artificial nature of the network construction, this discreteness in representation is also a byproduct of the necessary simplicity of the current multiTRACE model, in that individual network units represent populations of neurons, thereby limiting our knowledge of how the simulated assemblies can relate to one another in a neurobiological sense. For example, it is not clear to what degree cortical representations for words are distinct or if they overlap. However, this simplicity in the model does not damage its biological credibility. That is, because the internal representations of these simulated cell assemblies remains unspecified, theoretically this allows neurons to be redundantly represented across several units in the model.

Acknowledgments

The authors would like to thank Louisa Slowiaczek for her invaluable assistance in this project. The first author was partially funded by a Hughes grant during the summer.

References

- Amit, D.J. (1995). The Hebbian paradigm reintegrated: Local reverberations as internal representations. *Behavioral and Brain Sciences*, 18(4): 617-657.
- Chown, E. (1994). Consolidation and learning: A connectionist model of human credit assignment. Doctoral dissertation. The University of Michigan.
- Hamburger, M. & Slowiaczek, L.M. (1996). Phonological priming reflects lexical competition. *Psychonomic Bulletin & Review*, 3(4): 520-525.
- Hebb, D.O. (1949). *The Organization of Behavior*. John Wiley.
- Goldinger, S. D., Luce, P.A., Pisoni, D.B., & Marcario, J.K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18, 1211-1238.
- Kaplan, S., Sonntag, M. & Chown, E. (1991). Tracing recurrent activity in cognitive elements (TRACE): A model of temporal dynamics in a cell assembly. *Connection Science*, 3, 179-206.
- Kinsbourne, M. (1982). Hemispheric specialization and the growth of human understanding. *American Psychologist*, 37(4), 411-420.
- Rochester, N., Holland, J.H., Haibt, L.H., & Duda, W.L. (1956). Tests on a cell assembly theory of the action of the brain, using a large digital computer. *IRE Transactions on Information Theory*, IT-2:80—93.
- Slowiaczek, L.M. & Hamburger, M. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18(6): 1239-1350.
- Sonntag, M.. (1991). Learning sequences in an associative network: A step towards cognitive structure. Doctoral Dissertation. The University of Michigan.