NETWORK REGIONS:
Alternatives to the Winner-Take-All Structure

Hon Wai Chun
Honeywell Bull
300 Concord Road, MA30-895A
Billerica, MA 01821-4186
U.S.A.

Lawrence A. Bookman
Computer Science Department
Brandeis University
Waltham, MA 02254
U.S.A.

Niki Afshartous
Honeywell Bull
300 Concord Road, MA30-895A
Billerica, MA 01821-4186
U.S.A.

ABSTRACT

Winner-take-all (WTA) structures are currently used in massively parallel (connectionist) networks to represent competitive behavior among sets of alternative hypotheses. However, this form of competition might be too rigid and not be appropriate for certain applications. For example, applications that involve noisy and erroneous inputs might mislead WTA structures into selecting a wrong outcome. In addition, for networks that continuously process input data, the outcome must dynamically change with changing inputs; WTA structures might "lock-in" on a previous outcome. This paper offers an alternative competition model for these applications. The model is based upon a meta-network representation scheme called *network regions* that are analogous to *net spaces* in partitioned semantic networks. *Network regions* can be used in many ways to clarify the representational structure in massively parallel networks. This paper focuses on how they are used to provide a flexible and adaptive competition model. *Regions* can be considered as representational units that represent the conceptual abstraction of a collection of nodes (or hypotheses). Through this higher-level abstraction, *regions* can better influence the collective behavior of nodes within the region. Several AI applications were used to test and evaluate this model.

I. INTRODUCTION

Winner-take-all (WTA) structures [1] represent competitive behavior among sets of alternative hypotheses in massively parallel networks [2, 3, 4]. These networks consist of large numbers of simple processing elements that give rise to emergent collective properties. The behavior of such networks has been shown to closely match human cognition in many tasks, such as natural language understanding and parsing, learning, speech perception and recognition, speech generation, physical skill modeling, vision and others.

In a WTA structure, whenever there is any activation, the structure forces only the node (which may represent a hypothesis) with the highest output level to remain activated, while the other nodes die out. This mechanism allows one to define "decision points" that can control inference and guide convergence. It can also reduce the possibility of uniform saturation ("heat death") and allow one to create a more stable system.

Figure 1* is a simple WTA structure that partially encodes phonological knowledge about the weak fricatives — /f/, /v/, /θ/, and /δ/**. This network selects a weak fricative given phonetic features as inputs. The nodes used are p-units [1] with decay. The functions used to compute potential p and output v are:

$$p \leftarrow p\,(1-d) + \Sigma(w_k \times i_k) \quad [0 \leq w_k \leq 1]$$
$$v \leftarrow \text{if } p > 0 \quad \text{then } p \text{ else } 0$$

where w_k is the link weight on the link from input i_k, d is the decay factor, and p and v are continuous values between 0 and 1.

The amount of competition, in terms of negative inhibition, to node j is given by:

$$competition_j = w_I\, \Sigma\, v_i \qquad [i \neq j]$$

where w_I is the inhibition link weight and v_i is the output levels of the other nodes.

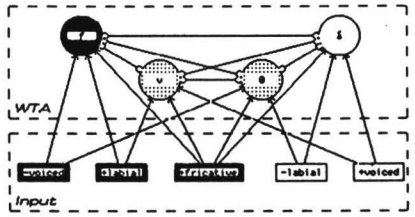

Figure 1. WTA structure that represents competing weak fricatives.

In the example, the link weights for inhibition and activation links are arbitrarily set to be −0.6 and 0.33 respectively and the decay factors are set to 0.3. Figure 1 is the final state after the inputs **−voiced**, **+labial**, and **+fricative** were activated and the network was relaxed. As the figure shows, the WTA structure enabled /f/ to compete and suppress the other candidates.

The main characteristic of WTA structures is that nodes mutually inhibit each other. Nodes with higher

* The dotted boxes are for illustration only and do not have any computational significance.

** /f/ is the sound of "f" in "fit", /v/ is the sound of "v" in "voice", /θ/ is the sound of "th" in "thin", and /δ/ is the sound of "th" in "that".

output suppress nodes with lower output. There are two parameters that influence this competitive behavior — the inhibition link weights and the nodes' decay factor. The link weights define the degree of competition. Lower link weights represent a milder, slower form of competition. Higher link weights represent stronger, quicker competition. The decay factor, on the other hand, controls how well a network retains the decision made by the WTA structure and allows the network to reset itself in the absence of inputs. In this paper, these two parameters are assumed to be uniform.

II. ISSUES

Although WTA structures have been used successfully in many AI applications, this model of competition might not always be appropriate. For example, applications that involve noisy and erroneous inputs might mislead WTA structures into forming a wrong decision. In addition, for networks which continuously process input data, the network must be able to dynamically change decisions based upon changing inputs. WTA structures might "lock-in" on a particular decision. The following outlines the key problems when WTA structures are used in these applications.

Overly rigid competition — Since the mutual inhibition link weights are fixed, the degree of competition is also fixed. If the input data is not at a consistent activation level or if new hypotheses are encoded into the structure, the competition may be too weak or too strong. Competition that is too weak may result in multiple activations or decisions, some of which may be conflicting. Excessive competition may result in a structure that is very sensitive to the output levels of the nodes. This may mislead a WTA structure into prematurely forming an incorrect decision.

Decisions are not flexible — Once a WTA structure "locks-in" on one interpretation, it is difficult to shift to another interpretation as the input changes. In other words, newly activated nodes will have little chance of competing with the current "winner". Flexible decision making is essential for dynamic systems where the network continuously processes different inputs. The decay factor does alleviate the "lock-in" effect by allowing the winning node to gradually decay, but may not be in time to accommodate changes in the input.

No likelihood information — The output level of a node may sometimes be considered a "likelihood" or "certainty" measure [5]. WTA structures, by definition, allow only one "winner", thus the output levels of other nodes are suppressed and this "likelihood" information is lost. In certain applications, it is more desirable to have several outcomes with varying output levels to reflect the likelihood of these outcomes.

In general, WTA structures are useful in applications where only one outcome is desired and where the input data is not noisy and can be clearly categorized. In these well-defined applications, some of the above mentioned problems can be solved by using building blocks or combination rules in constructing the WTA structure [1]. However, for applications involving real world data with large variance, a more flexible and adaptive competition is needed. This paper presents an alternative model of competition for these applications.

III. NETWORK REGIONS

Network region is a meta-network structure that represents the conceptual abstraction of a collection of nodes or hypotheses. Intuitively, it represents a "chunk" of network where well-behaved properties can be defined. Graphically, a *region* is displayed as a solid-lined rectangular box around the collection of nodes it represents.

The following figure shows the previous WTA structure enclosed in a *network region*, labeled **WF**, which represents the category of weak fricatives.

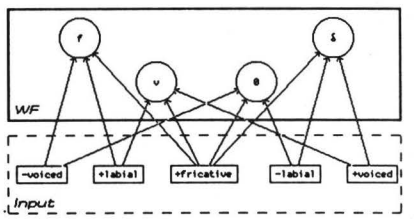

Figure 2. A region representing the category of weak fricatives.

Conceptually, there are similarities between *network regions* and partitioned semantic networks [6]. In semantic networks, nodes and arcs are partitioned into *net spaces*, while in our massively parallel networks, nodes and arcs are partitioned into *network regions*. The *net spaces* are used mainly to delimit the scopes of quantified variables. In this paper, *network regions* are used to delimit sets of competing hypotheses. In addition, besides being conceptual units, *network regions* are computational entities just as nodes and links are. *Network regions* can also activate or inhibit other nodes or regions, similar to *spaces* that can be linked to other parts of a semantic network.

Regions are in effect higher-order nodes with internal parameters, that have node counterparts (i.e. potential level, output level, and inputs) where:

S_r - sum of all the nodes' output within region.
P_r - average potential of activated nodes in region.
V_r - average output of activated nodes in region.
I_r - average input level activated nodes in region.

These parameters act as indicators of the collective state of the nodes within a *region* and are updated during each

relaxation cycle. This provides an abstract view of a set of nodes that can be used either to influence the network system outside the *region*, or to better control the collective behavior of nodes within the *region*.

In terms of influencing the network system outside a given *region*, *regions* can be used to encode *partition of concerns* [7]. In the above example, even though the network may still be determining the correct weak fricative, the state of the **WF** *region* (or more precisely, the value of V_r) can immediately provide information to other network structures which may only need to know that a weak fricative is present and not necessarily which particular weak fricative. This allows phonological rules, that apply to all weak fricatives, to trigger whenever there is any activity in the **WF** *region*.

Regions also simplify the task of knowledge encoding by permitting knowledge to be encoded at appropriate levels of abstraction. In speech recognition, phonological rules are used at various levels of abstraction. Through the use of *regions*, rules can be encoded into connectionist networks at the same level they are expressed. An example rule is: if a final voiced fricative is actually voiced, then the following sound is voiced. This may be encoded with a *region* that represents the set of all voiced fricatives. If this *region* is active and is followed by a silence, then the system should expect a **+voiced** sound to follow. Without the *regions* representation, this knowledge has to be encoded for each voiced fricative separately. With *regions*, knowledge is expressed at a more appropriate level of abstraction, that allows networks to be more modular and comprehensible.

The interaction of *regions* with other network structures is discussed in (Bookman and Chun, 1987). The current paper focuses on the use of *network regions* to influence the collective competitive behavior of nodes within a region.

A. N-REGION

The N-REGION (normalizing region) provides activation stability for competing hypotheses. In addition this mechanism allows "likelihood" information to be maintained and permits competition that is more tolerant to variations in input levels.

In certain applications, it is useful to have more than one outcome from a set of alternative hypotheses. In these cases, the output level of a node may be considered as a "likelihood" or "certainty" of a particular hypothesis [5]. Since WTA structures only allow one "winner", the output levels of the other nodes are suppressed and the "likelihood" information is lost. The N-REGION, without the mutual inhibition links, maintains this likelihood information within a set of alternative hypotheses. If we consider a hypothesis by itself, the output level would indicate the "likelihood" of this hypothesis as determined by the current low-level inputs. For example, if only half of the inputs to a particular hypothesis are activated, the hypothesis may have an output level of 0.5. This information is only useful if we are considering how well a particular hypothesis matches the current input. However, when hypotheses are in competition, it is more informative to have the output level indicate the likelihood of a particular hypothesis among all other competing hypotheses.

The N-REGION accomplishes this by a normalizing computation that limits the maximum total output within a region (the value of S_r of the N-REGION) to 1.0. In other words, the total "likelihood" within the N-REGION will not exceed 1.0. The general idea of normalization is of course not new, however with *network regions* the normalization is integrated into the relaxation process itself. Figure 3 and Table 1 shows the output levels of the four weak fricatives with and without the N-REGION, after the inputs **−voiced**, **+labial** and **+fricative** were activated for 1 cycle.

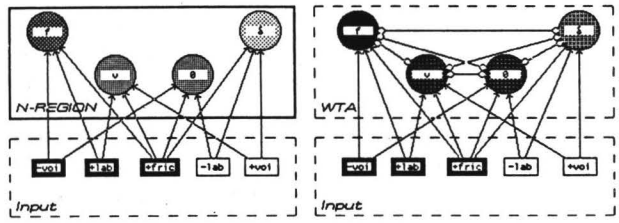

Figure 3. The activation with N-REGION and WTA.

Phonemes	/f/	/v/	/θ/	/δ/
with N-REGION	0.375	0.25	0.25	0.125
with WTA	0.99	0.66	0.66	0.33

Table 1. Table of output levels after 1 relaxation cycle.

The output level of /v/ with the WTA structure indicates that there is a 0.66 likelihood that /v/ is matched just by considering only the inputs (only two-thirds of the features for /v/ were activated)*. This output level is misleading, since /f/, /v/ and /θ/ are all highly activated. This would misinform other network structures that all three were closely matched. However, with the N-REGION, the output level of /v/ is adjusted to 0.25, indicating the likelihood of /v/ among the four alternatives is only 0.25. This mechanism puts a hypothesis in perspective among all the other competing hypotheses.

This approach is similar to the UB (upper-bound) parameter proposed in (Feldman and Ballard, 1982).

* In our example, the activation link weights are uniformly set to 0.33. In evidential reasoning systems, this weight would be E (hypothesis I input). However, the utility of the N-REGION is the same.

However, instead of inhibiting all the nodes equally, the N-REGION proportionally adjusts the nodes output level to maintain the "likelihood" information.

The virtual lateral inhibition proposed by (Reggia, 1985) also maintains a normalized activation in each layer when only a single input node is activated at a time. Since the effect of this competition model is similar to WTA structures, the network might also "lock-in" on a premature outcome.

Since there are no inhibition links between competing nodes in the N-REGION, it does not encode active competition or contour enhancement [10] (filter out noise by suppressing nodes with low output level). However, there is implicit competition through the conservation of activation within the N-REGION. In addition, by this lack of active competition, N-REGIONs will not "lock-in" on a particular decision. If active competition is required, the N-REGION can be used in conjunction with a WTA structure.

The amount of competitive inhibition received by a node in a WTA structure varies with the total output level in the structure. After the WTA structure is fine-tuned for a particular input activation level, the performance varies when the input level increases or decreases. When N-REGIONs are used on top of WTA structures, the N-REGIONs maintain the total output level to be constant. Hence, a more uniform competition results since the total amount of competition will not vary with input levels. This also eliminates the problem of too weak a competition where several hypotheses get fully activated. With N-REGIONs, the activation of these hypotheses will be normalized.

To summarize, the most significant advantages of N-REGIONs are that they maintain "likelihood" information, permit uniform competition, eliminate excessive competition, allow multiple outcomes, and will not "lock-in" on a single "winner".

B. C-REGION

The C-REGION (competitive region) provides an adaptive and flexible competition that permits graceful shifting of decisions made when inputs change.

The WTA competition is rigid because the weights on the inhibition links are fixed. When input levels are not consistent, as in the case of noisy data, competition may either be too weak or too strong. For example, in speech recognition, the clarity in which certain phonemes are spoken varies greatly within each word. The C-REGION avoids this problem by having the inhibition link weights be sensitive to the current inputs to the region (i.e. the value of I_r in the region). As the C-REGION relaxes, the inhibition link weights are adjusted to adapt to the current input level.

Competition in the C-REGION is defined as:

$$competition_j = k_I \, I_r \, w_I \, \Sigma \, v_i \qquad [i \neq j]$$

where k_I is a constant, indicating the sensitivity to the input. In contrast to the original WTA competition, the "effective inhibition weight" is now $(k_I \; I_r \; w_I)$. The parameter k_I is set so that the "effective inhibition weight" for the expected input levels is the same as the original WTA inhibition weight.

The following experiments show how the C-REGION adapts to varying input levels. In the examples, the activation link weights from the inputs are 0.33. The inhibition link weights in both the WTA structure and the C-REGION are −0.6. The proportionality constant k_I is set to 1.5 so that the "effective inhibition weight" of the C-REGION is approximately equal to the inhibition weight (i.e. −0.6) of the WTA structure when inputs nodes are fully activated.

In experiment 1, when the input nodes were fully activated, the outcome ("winner") will shift when inputs change for both the WTA and C-REGION. However, when the average input level, I_r, drops to 0.4, the WTA structure "locks in" on the previous outcome, while the C-REGION will shift to the correct decision. Table 2 shows that the WTA locked onto /θ/, while the C-REGION correctly shifted to /δ/. Figure 4 shows the final state of the network at the end of 45 cycles.

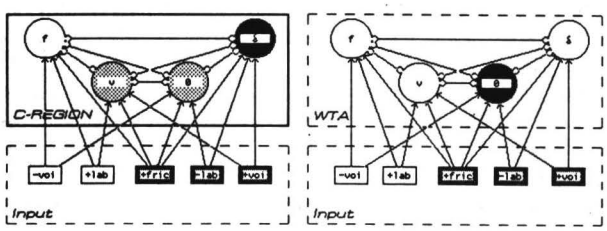

Figure 4. The state of the networks at the end of experiment 1.

	Input					WTA		C-REGION		
Cycle	+fric	+voi	−voi	+lab	−lab	w_I	result	$kI_r \, w_I$	I_r	result
15	1.0	0.0	1.0	1.0	0.0	−0.6	/f/	−0.59	0.66	/f/
30	1.0	0.0	1.0	0.0	1.0	−0.6	/θ/	−0.59	0.66	/θ/
45	0.6	0.6	0.0	0.0	0.6	−0.6	/θ/	−0.36	0.40	/δ/

Table 2. Tabulated results from experiment 1.

In experiment 2 (see Table 3), the average input level, I_r, for the first 15 cycles was lower than expected (only 0.4). However, both network structures still settled on the correct outcome of /f/. In the previous experiment, the WTA structure failed because it "locked in" on a previous outcome. Here, there is no previous outcome to lock onto. However, when the input was changed and the average input level dropped further to 0.26, the WTA structure failed while the C-REGION still

selected the correct outcome of /θ/.

Input						WTA		C-REGION		
Cycle	+fric	+voi	−voi	+lab	−lab	w_l	result	$kI_r w_l$	I_r	result
15	0.6	0.0	0.6	0.6	0.0	−0.6	/f/	−0.36	0.40	/f/
30	0.4	0.0	0.4	0.0	0.4	−0.6	/f/	−0.24	0.26	/θ/

Table 3. Tabulated results from experiment 2.

The main advantage of the C-REGION is that it provides a flexible competition which adapts to the current input level. The decision made by the structure is less dependent on previous decisions (i.e. will not "lock in" on previous results). This ability to graceful change decisions is crucial in networks which continuously process different inputs at varying levels. When the input level is higher than the expected value, the competition in both the WTA structure and C-REGION might be too strong. Consequently, more negative competition activation will be spread. If this activation is higher than the nodes' output level, nodes in both structures may toggle between active and inactive states. The CN-REGION eliminates this problem.

C. CN-REGION

CN-REGIONs, used on top of WTA structures, combine the advantages of both C-REGIONs and N-REGIONs. The N-REGION maintains uniform competition with possibly multiple outcomes, while the C-REGION adapts competition to varying input levels. When average input levels are lower than expected, the C-REGION can still permit the inputs to influence the competition. When average input levels are higher than expected, the N-REGION prevents excessive competition. The net effect of combining these two is that competition is highly flexible and uniform. The possibility of too strong or too weak competition is greatly reduced.

The effects of the CN-REGION are similar to the competition in Grossberg's on-center off-surround network [10]. Grossberg's network uses a quenching threshold to limit which nodes will be activated. In essence, the adaptive competition in the CN-REGION is similar to a self-adjusting quenching threshold that changes with the current input level.

IV. APPLICATIONS

The *network region* model of competition was tested in three AI applications — high-level vision, isolated word recognition, and microfeature-based natural language understanding. These examples illustrate the effectiveness of WTA structures and *network regions* in different classes of applications. The input to the labeling problem of high-level vision is well-defined without noise, whereas the input to the word recognition example is highly noisy and ambiguous. The natural language example shows the importance of flexible decision making. All examples were implemented using the AINET-2 system [11] (a massively parallel network simulator and development environment) which runs on Symbolics Lisp machines.

A. HIGH-LEVEL VISION

This section investigates the line drawing labeling problem [12] of high-level vision. In our line drawing labeling system, each junction in an object is represented by a WTA structure with nodes that represent the possible labels for that junction. The label of a junction is rigidly constrained; it must agree with the labels of adjacent junctions. These constraints are represented as activation links between adjacent junctions. This method of constraint propagation by spreading activation mimics the effects of the Waltz filtering algorithm.

An example object and its correct labeling is shown in Figure 5. The three types of junctions that occur in this object are the L's, forks, and arrows (see Figure 6).

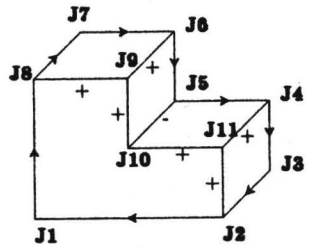

Figure 5. Line drawing of the example object with its labeling.

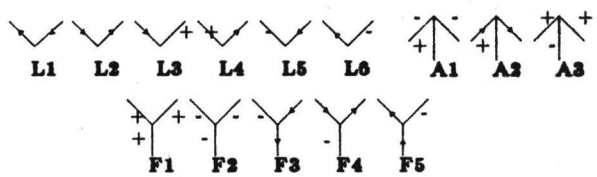

Figure 6. The various types of L's, forks (F), and arrows (A) used.

All nodes were initially active, to indicate all labels are equally possible. As the network relaxed, nodes which did not satisfy the defined constraints died out. Eventually, the network converges to the correct labeling as shown by the darkened nodes in Figure 7.

In our experiments, networks performed correctly as long as there was a balance between the inhibition and activation link weights. Both excessive inhibition and weak activation caused the network to converge incorrectly. Weak inhibition and excessive activation resulted in multiple labels being selected at some of the junctions.

This labeling system was also tested with N-REGIONs at each junction, in addition to the WTA structures. This resulted in a slightly greater range of possible inhibition and activation link weights in which the network would still converge to the correct labeling. In addition, the network tended to degrade gracefully when link weights varied beyond this range. Without N-REGIONs there was a sharp threshold after which the network performed improperly. This improvement in network behavior can be attributed to the uniform competition provided by N-REGION. Similar results were observed for the CN-REGION.

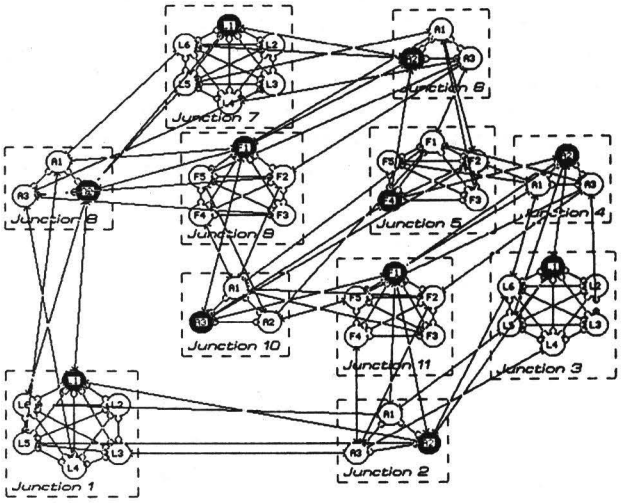

Figure 7. Network with WTA structures to perform labeling.

WTA structures are effective for the line drawing labeling problem and other domains where constraints are rigidly defined. The problem of "locking in" on a decision is not relevant here since evidence for a particular hypothesis (label) cannot vary with time. However, as we shall see in the following sections, this problem becomes more significant in applications where evidence for a hypothesis can vary with time.

B. ISOLATED WORD RECOGNITION

An isolated word recognition system, called SECO [13, 14], was used as an application with highly noisy and inconsistent inputs. This system recognizes spoken letter-names and digits. It partially evolved from a massively parallel model of word perception called COHORT [15]. The structure of SECO is similar to COHORT with the addition of a temporal constraint layer.

The word recognition network consists of five layers — phonetic feature (e.g. vowels, stops, etc.) layer that is activated by an LPC-based front-end processor, phonetic segment (e.g. /f/, /v/, etc.) layer, phonetic segment token layer, temporal constraint layer, and a word layer which is the output of this system. Nodes in the word layer forms a WTA structure. Figure 8 shows an abstract overall view of the system.

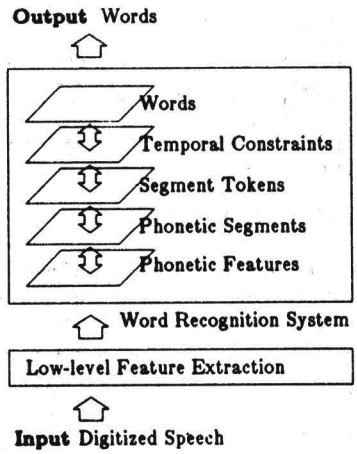

Figure 8. The overall structure of the word recognition model.

Word recognition involves evidence gathering (in the form of phonetic segments matched) over a period of time. An observation from the experiments is that a WTA structure might converge to a decision based on the initial data and "lock-in" on an incorrect word (very often the initial part of an utterance is noisy). In this case, the initial noise may gradually dominate the competition. In addition, during the recognition process, the correct phonetic segment may not always be the most highly activated segment. This would easily mislead a WTA structure, whereas a CN-REGION would be more capable of accommodating this type of noisy and erroneous data. The adaptive competition capability of CN-REGION would permit later inputs to be able to positively influence the network. Figure 10 illustrates how the WTA structure misinterpreted the utterance "six" to be "seven" based on initial data. Figure 11 is the result from using a CN-REGION on the same utterance.

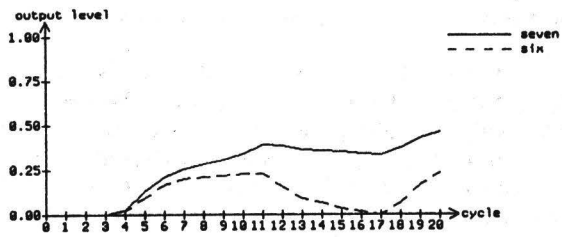

Figure 10. Using a WTA structure in recognizing the word "six".

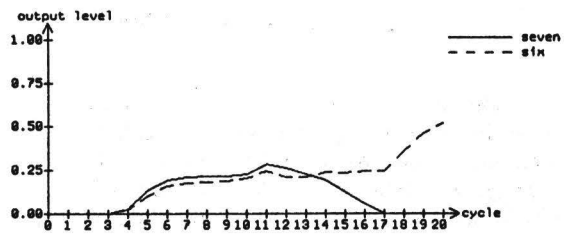

Figure 11. Using a CN-REGION in recognizing the word "six".

When a word is spoken, the duration of the speech varies with the speaker and the current context. This variation in the duration of input data causes problems for WTA structures. If an utterance is longer than expected, word level nodes will be highly activated before the utterance is complete and prematurely settle on a "winner." The normalizing effect of the CN-REGION prevents this "premature" recognition by adjusting the output levels to represent the current "likelihood" of a word at any point during the recognition process.

When competition is too weak, the WTA structure may fully activate more than one word node. The outcomes may also be activated before utterances is completed. With CN-REGION, normalization reduces the nodes' output to better indicate their "likelihoods." Figure 12 shows an example where the input utterance is "three." This particular utterance is somewhat noisy and the network confuses this with "four" while slightly favoring "three". The plot shows the WTA structure activates both "three" and "four." This would misinform other network structures that both words were recognized. In the case of CN-REGION (Figure 13), "three" and "four" are only partially activated.

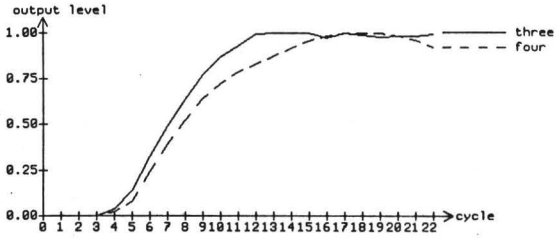

Figure 12. Using a WTA structure in recognizing the word "three".

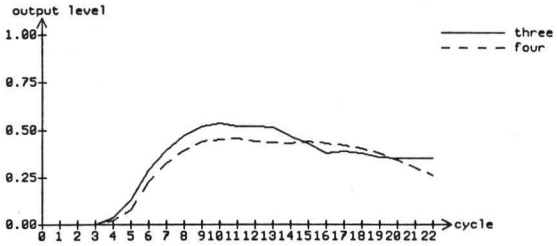

Figure 13. Using a CN-REGION in recognizing the word "three".

The domain of word recognition is typical of many applications that involve real world data. Input to the system is extremely noisy with a varying input level. In these applications, the CN-REGION is more appropriate.

C. NATURAL LANGUAGE UNDERSTANDING

A "locally-distributed" microfeature based natural language system [16] was used to perform word sense disambiguation [17, 18]. This model is built upon the work of (Waltz, 1985) and incorporates the strengths of both "local" and "distributed" massively parallel representations. The system is composed of two layers. The top layer is a "local" connectionist model, the bottom layer, a distributed layer of "microfeatures." The microfeatures are used as a basis for defining nodes (i.e. concepts, hypotheses) in the top layer, at least partially, and to associate the node with others that share its microfeatures. Each node in the top layer, is connected via bi-directional links to only those microfeatures that describe it. Conceptually related nodes have common microfeatures.

The following is one of the sentences used to contrast the competition found in WTA structures with CN-REGIONs.

"John went driving with five bucks in his truck."

The network that models this sentence is shown in Figure 14 (only the top layer is shown). The rectangular nodes in the center of the figure represent the input words. The structure above this is the syntactic parse tree for the sentence. The elliptical nodes below the input words represent the competing word senses. In addition, semantic constraints associated with the inputs, are encoded but not displayed. There is also a microfeature memory system that represents the currently active and inactive microfeatures (these are not shown in the figure).

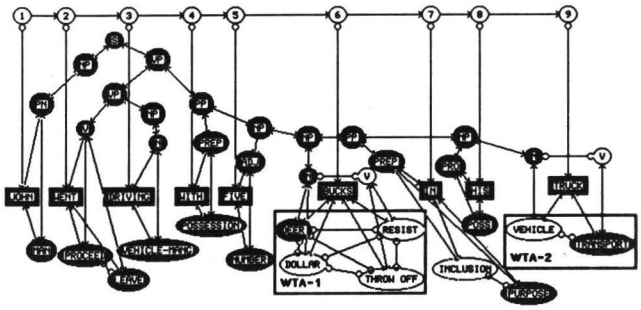

Figure 14. Network used to process the example sentence.

The example sentence is ambiguous, for it can either mean: he went driving with five dollars in his truck, or he went driving with five deer in his truck. In WTA-1 (Figure 14) two problems occur. Initially priming the *WEEKEND* and *GAMBLING* context, causes the system to settle on the *dollar* interpretation of *bucks*, while priming it with *HUNTING* causes it to settle on the *deer* interpretation of *bucks*. In both cases the competition is too strong and there really should not be a total winner in this case. Instead there should be some overlap, as the features shared by both senses of the sentence are shared by both priming contexts. Using CN-REGIONs, this is exactly what happens. Priming *WEEKEND* and *GAMBLING* causes the *dollar* sense to be higher than the *deer* sense, while priming *HUNTING* causes *the* deer sense to be higher than the *dollar* sense, yet neither totally win out, thus reflecting the ambiguity in the data.

In WTA-2, the correct sense of *truck* is not chosen. Here, initially the input *driving* activates the *transport* sense of *truck*, but later inputs activate the *vehicle* sense of the word. However, they are too late to enter competition. This was not the case with the CN-REGION, which correctly shifted to the *vehicle* sense of the word.

Results from our experiments indicate several problems that WTA structures pose for natural language processing. First, the semantic knowledge of the network and its microfeatures varies with one's experience. There is no one correct set. Thus, slight variations in the network's microfeature set will cause problems for WTA networks which are sensitive to transiently higher input. This can lead to premature stabilization of activation. On the other hand, the CN-REGIONs allow smooth competitions among the microfeatures, the syntactic structures, the semantic constraints, and the input words, thus enabling subtle differences in meaning to exist. Secondly, the processing of a sentence is sequential over time. As a result, one's understanding of the sentence and the sentences that follow are gradually refined as one processes the inputs. A WTA may prematurely "lock in" on a word sense and ignore later information. Whereas, CN-REGION permits a dynamic flexibility in shifting from one interpretation to another. In addition, WTA does not provide a good representation for ambiguous meanings, it always choose the one that is transiently higher. CN-REGION allows multiple outcomes each associated with a "likelihood" measure.

V. RESEARCH DIRECTIONS

The research documented here represents the first step in utilizing *network regions* to provide higher-level representation in massively parallel networks. As *regions* can provide a means of creating hierarchies, as well as shared structures, it may be possible to use current learning techniques to create networks that can generalize from existing concepts to form new *regions* that represent these generalized concepts. This is a topic for future research. Currently we are investigating the effects of interaction among *regions* [8], that is, the computational mechanisms needed to represent the interaction of higher-order concepts.

ACKNOWLEDGEMENTS

We would like to thank Dave Waltz and the members of the Brandeis AI Group for their comments. We would also like to thank the members of the Knowledge Engineering Center at Honeywell Bull.

REFERENCES

[1] Feldman, J.A. and D.H. Ballard, "Connectionist Models and Their Properties," *Cognitive Science*, 6, 1982, pp.205-254.

[2] Hinton, G.E. and J.A. Anderson (eds.), *Parallel Models of Associative Memory*, Hillsdale: Lawrence Erlbaum Associates, 1981.

[3] Rumelhart, D.E. & McClelland, J.L., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition Volume 1*, MIT Press, 1986.

[4] McClelland, J.L. & D.E. Rumelhart, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition Volume 2*, MIT Press, 1986.

[5] Shastri, L. and J.A. Feldman, "Semantic Networks and Neural Nets," Computer Science Department, The University of Rochester, TR131, June, 1984.

[6] Hendrix, G.G., "Expanding the Utility of Semantic Networks Through Partitioning," In *Proceedings of IJCAI-75*, 1975, pp.115-121

[7] Brachman, R.J. and H.J. Levesque, "On the Partitioning of Concerns in Knowledge Representation," In *Proceedings of AAAI-82*, 1982.

[8] Bookman, L.A. and H.W. Chun, "A Model of Higher-Order Concept Interaction Using Network Regions," Computer Science Department, Brandeis University, Technical Report, CS-87-130, Massachusetts, 1987.

[9] Reggia, J.A., "Virtual Lateral Inhibition in Parallel Activation Models of Associative Memory," In *Proceedings of IJCAI-85*, 1987, pp.244-248.

[10] Grossberg, S., "Contour Enhancement, Short Term Memory, and Constancies in Reverberating Neural Networks," *Studies in Applied Mathematics*, Vol. LII, No. 3, September, 1973, pp.213-257.

[11] Chun, H.W., "AINET-2 User's Manual," Computer Science Department, Brandeis University, Technical Report, CS-86-126, Massachusetts, 1986.

[12] Waltz, D. L., "Understanding Line Drawings of Scenes with Shadows," in *The Psychology of Computer Vision*, P. Winston (ed), McGraw-Hill, 1975.

[13] Chun, H.W., "A Representation for Temporal Sequence and Duration in Massively Parallel Networks: Exploiting Link Interactions," In *Proceedings of AAAI-86*, August 1986.

[14] Wong, M.K. and H.W. Chun, "Toward a Massively Parallel System for Word Recognition," *1986 IEEE - International Conference on Acoustics, Speech, and Signal Processing*, Tokyo, Japan, April 1986.

[15] Elman, J.L. and J.L. McClelland, "Speech Perception as a Cognitive Process: The Interactive Activation Model," In N. Lass (ed.), *Speech and Language: Vol. X.*, Orlando, Florida, Academic, 1984.

[16] Bookman, L.A., "A Microfeature Based Scheme For Modelling Semantics," In *Proceedings of IJCAI-87*, 1987.

[17] Cottrell, G.W. "A connectionist approach to word sense disambiguation," (PhD thesis) TR 154, U. of Rochester CS Dept., 1985.

[18] Waltz, D.L. and J.B. Pollack, "Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation," *Cognitive Science*, Volume 9, Number 1, January-March, 1985.