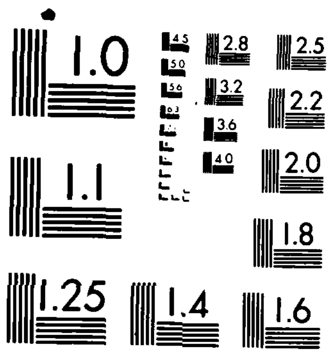
AD-A161 362 INTERACTIVE ACTIVATION MODELS OF PERCEPTION AND COMPREHENSION(U) CALIFORNIA UNIV SAN DIEGO LA JOLLA J L ELMAN ET AL 01 OCT 85 N00014-85-K-0076 1/1

UNCLASSIFIED F/G 5/10 NL

END
FILMED
DTIC


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

(12)

AD-A161 362

ONR YEARLY STATUS REPORT

INTERACTIVE ACTIVATION MODELS OF PERCEPTION AND COMPREHENSION

Drs. Jeffrey L. Elman & James L. McClelland

Date: 7 October 1985

Period Covered by this Report: 1 December, 1984 to 1 October, 1985

Contract Number: N00014-85-K-0076

Contract Authority Identification Number: NR 667-542

Effective Date of Contract: 1 December 1984

Expiration Date of Contract: 30 November 1987

Amount of Contract: $465,715

Scientific Officers: Dr. Harold L. Hawkins & Dr. Susan Chipman

Contractor: The Regents of University of California La Jolla, CA

Principal Investigators:

Jeffrey L. Elman
Phone: (619) 452-2537

James L. McClelland
Phone: (412) 578-2789

Sponsored by
Office of Naval Research

CONTENTS

Letter on file

A-1

1. Research Program Plan: Overview

The objective of this research is to construct a computationally sufficient, biologically plausible, and behaviorally adequate account of human information processing skills in visual and auditory language processing.

We have the following specific research goals for our contract:

(1) To implement a model of reading printed text through a series of fixations. The model is intended to account for the integration of visual information over successive fixations, and the interaction of visual and contextual information in reading.

(2) To implement a new version of our model of speech perception (TRACE), using programmable connections to allow the model to tune itself, in the course of processing, to changes in global parameters such as rate. This new model (which we will call the **Programmable TRACE**) is intended to account for human sensitivity to global as well as local contextual influences on the speech signal while retaining all the virtues of the present version of TRACE. It will allow us to make the crucial distinction between *types* of different items and *tokens* of those types. This distinction is not made in our current model.

(3) To begin work on the development of simulation models designed to capture aspects of the interactions between lexical, syntactic, and semantic constraints on the construction of syntactic and functional representations of sentences.

Our technical approach is to develop explicit conceptual models of information processing and to embody these models in computer simulations. Using the computer simulations, we are then able to determine how well our model fares, both in terms of its computational adequacy to actually carry out a specified information processing task, but also in terms of its behavioral adequacy, to carry out the task in a way that accords with what we know about how human subjects do the tasks. We also feel it is important to collect data of our own to constrain the model in key places or to provide evidence supporting our general approach and distinguishing it from the kinds of approaches represented in other models.

2. Description of Progress

During the first year of the contract, we have concentrated on the following specific efforts:

(1) Development of the **Programmable TRACE** version of the speech model;

(2) Experimental studies with human subjects, in order to testing predictions of the TRACE model with regard to the need for feedback between processing levels;

(3) Exploration and application of *Connection Information Distribution*, which allows networks of uncommitted units to be programmed by signals arising elsewhere in the network.

2.1. Programmable Trace

We have completed the conceptual development of the **Programmable TRACE** model at the feature and phoneme level. This version of the model contains (a) central structures which contain information about the canonical relationships between acoustic/phonetic features and phonemes; and (b) programmable structures which are capable of being tuned by the central structures in such a way that they can process the actual speech input received at any given instant. In this way, the model is able to maintain a distinction between different *types* of units (represented by the central structures), and different *tokens* of those units (represented by the programmable structures).

The model also contains structures which are sensitive to rate of speech. These structures participate in the tuning of the of the programmable structures so that they are adjusted appropriately for the rate of speech.

We are now in the process of constructing a computer simulation of this model. We estimate the first version of this program will be ready in approximately two months. At that time we shall begin testing it with simulations of human experimental data.

2.2. Experimental findings

There were two issues which arose in the construction of the **Programmable TRACE** model, and in discussions with colleagues. First, we have suggested that there is feedback from lower to higher processing levels. It has seemed to us that such feedback is useful and necessary to account for findings in the human experimental literature. It is also true that the the interactive activation framework encourages us to postulate a high degree of exchange of information between levels. However, several people have claimed that this feedback is not necessary, and that the effects we try to explain by means of the feedback can be accomplished by simply having lower processing levels furnish higher levels with more detailed information. In recent years a number of scientists and philosophers have argued that much of perceptual processing is carried out by highly modular subsystems which are "informationally encapsulated" from each other. In this view, the flow of information is from "lower" to "higher" levels of processing.

A second issue concerned the question of how what is happening in one moment of time can affect processing that occurs either earlier or later. Speech is characterized by a high degree of variability in the acoustic form which different speech sounds have, depending on the adjacent sounds. This led us to modify **TRACE** so that processing units for phonemes in one time slice could modify the connections between acoustic feature nodes and phoneme nodes in neighboring time slices, in just such a say as to capture this interdependence. Whether or not these connections are actually used -- as reflected in perception -- remained an open question.

During the first year of the contract period we carried out a series of simulations and experiments which were designed to test these two issues. The tests grew out of the following two observations.

First, it is known that there is a lexical effect on phoneme perception. The nature of the effect is such that when subjects are presented with stimuli which contain ambiguous phonemes, perception of the sounds can be affected by the context in which they occur. Thus, if a sound mid-way between [g] and [k] (the mid-point is established earlier in a neutral context) is embedded in a context such as "_iss", subjects will tend to hear it as a [k]; whereas the same sound, embedded in a "_ift" context, will be heard as a "g".

Second, it is also known that the perception of an ambiguous sound can be affected by the context in a manner which suggests listeners are compensating for the affects of coarticulation. For example, the pronunciation of [t] and of [k] varies depending on whether the preceding sound is a [s] or [S] (the latter symbol denotes the "sh" sound). [s] causes both [t] and [k] to be pronounced as more [t]-like; [S] causes [t] and [k] to be pronounced as more [k]-like. Listeners appear to compensate for these pronunciation affects in the following way. A sound mid-way between [t] and [k] will be heard as [k] if it is preceded by [s] and as a [t] when preceded by [S]. In this way the effect of the [s] and [S] sounds is "undone" by the listener.

We carried out a series of simulations and experiments based on these effects in order to address the issues posed above. In our first simulation, we presented **TRACE** with sequences consisting of the words "abolish" and "progress", followed in each case by stimuli which were intermediate between [t] and [k]. The model responded to these stimuli by categorizing the [t]/[k] stimuli differently; the phoneme boundary was shifted toward the [t] end after "progress" (resulting in more of the stimuli being perceived as [k], and toward [k] after "abolish" (resulting in more [t]'s).

In the second simulation, we asked the question: Suppose the sound preceding the ambiguous [t]/[k] is not itself clearly either [s] or [S]. Suppose rather that the evidence for the identity of the preceding sound as [s] or [S] comes from the lexical item itself. Will there still be a shift in the phoneme boundary of the [t]/[k] stimuli? The answer was that there was indeed such a shift, similar to that obtained in the first simulation although smaller in magnitude.

Given this behavior of the model, we then preceded to verify whether human listeners would behave in the same way. In the first experiment with humans, subjects heard sequences such as "Christmas [t/k]ape", "foolish [t/k]ape", and "Spanish [d/g]ear", "ridiculous [d/g]ear" ([d] and [g] are similar to [t] and [k], except for voicing, and thus expected to behave similarly). Listeners judged the initial stop of the second word in each sequence, and as predicted by **TRACE**, heard the sounds as more [t]-like following [S] and more [k]-like following [s].

We then tested the second prediction, by taking the same sequences and substituting a sound intermediate between [s] and [S] for the last sound in the first word. Thus (representing this sound as a [?]), subjects heard sequences such as "Christma? [t/k]ape", "fooli[?] [t/k]ape", "Spani[?] [d/g]ear", and "ridiculou[?] [d/g]ear". Their task was the same, to identify the ambiguous sounds in the second word. Again, we found a shift in the phoneme boundary, similar to that found in the first experiment. The effect was not large, but was statistically significant with three different groups of subjects, each hearing a different set of stimuli.

These results, it seems to us, argue strongly for connections between adjacent processing units in time, in order to dynamically retune the mapping from acoustic input to phonemes in such a way as to compensate for coarticulatory influences. They also present a strong case for the need for feedback from the lexicon to those processes which determine phoneme identity. We see no way in which subjects could have performed as they did without allowing the lexicon to "reach down" and bias perception of the ambiguous [s/S] sounds, which then in turn were able to retune the [t/k] sounds which followed.

2.3. Using connection information distribution to program parallel processing structures.

A major part of our effort on this contract involves the development of the idea of programming parallel processing structures via *Connection Information Distribution*. Connection Information Distribution (CID) allows networks of uncommitted units to be programmed by signals arising elsewhere in the network. We consider this an important theoretical development, because it allows the creation of tokens, or 'local copies" of connection information. This kind of ability is crucial in a wide range of domains, such as visual perception, language processing, and comprehension. Heretofore, it has been possible to solve this "type-token problem" only with conventional serial computer mechanisms. Now, with connection information distribution, we can create copies of centrally stored knowledge, and yet still preserve the desirable properties of interactive activation models.

Three theoretical papers exploring this idea have been completed, and one experimental paper testing predictions derived from the approach is nearing completion.

2.3.1. The CID model

In McClelland (1985), completed prior to the present contract, a model called the CID model is presented, and it is applied to the task of reading one or two words four letters in length. The model consists of two programmable networks and a central network. Stimuli presented for processing in each programmable network are projected simultaneously onto the central network, where they give rise to activations of central units which in turn cause activation to flow to programmable connections in each local network. The input to each programmable network, in conjunction with the active connections, determines the output of the network. The network is capable of processing more than one word at a time, but exhibits crosstalk, so that when more than one word is shown, there is a tendency for letters in one of the two words to show up in attempts to report the other. This aspect of the model's behavior is in accord with some recent findings by Mozer (1983) on human performance in processing displays consisting of several words.

2.3.2. Resource requirements of standard and programmable nets.

McClelland (in press-a) analyzes the resource requirements, in terms of units and connections, of standard connectionist networks and of the mechanisms like the CID model which make use of programmable networks. The first part of the paper uses a signal detectability approach to re-derive in a particularly simple way the ability of a network to produce the correct outputs as a function of the number of units, connections between units, and input-output pairs stored. The second part of the paper applies the same ideas to CID mechanisms.

Two major results are developed: First, the programmable networks needed in a CID mechanism can be quite small relative to standard networks, primarily because the programmable nets need only be "loaded" with connection information sufficient to process a subset of the known patterns. Second, the central network needed in a CID mechanism may be quite large relative to a standard network for processing only a single pattern at a time, and the central network must grow in size approximately linearly with the number of patterns to be processed simultaneously. These results together suggest that it is better to program local networks one or only a few at a time.

2.3.3. The Programmable Blackboard Model of Reading

In McClelland (in press-b) the lessons of the just-described paper are used in the development of a model called "The Programmable Blackboard Model of Reading". This model postulates a structure called the Programmable Blackboard, consisting of programmable units. The units in the Programmable Blackboard can be programmed to process a word at a time, by way a sequential attentional mechanism which directs attention to successive words along a line of print in reading. The model is used to simulate a number of different experimental results from tasks involving the processing of single words or sequences of words. It also overcomes a limitation of the CID model, in that it is capable of processing words of different lengths (the present version goes only to four letter words, but words of greater length could be accommodated by an extension of the principles of design used in the present version of the model). Among the findings simulated are: a) Effects of word length on perceptual identification of letters in words. b) Within word letter transposition errors, occurring with greater frequency if i) the transformation makes a word and ii) if the original stimulus was not a word. Thus, transposition errors are particularly likely for stimuli like BCAK (error is BACK) than for GCAK (error would be nonword GACK) or for DRAT (error would be DART). c) Integration of information over successive fixations in reading. Here the simulation uses a parafoveal preview of a word in peripheral vision to facilitate processing of the word on the next fixation, simulating results of experiments by Rayner (1975) and Rayner, McKonkie, and Ehrlich (1978).

2.3.4. Perceptual interactions in multiword displays

McClelland and Mozer (ms. accepted pending minor revisions) have explored the letter migration errors reported by Mozer (1983) and others. In Mozer's task, subjects viewed two-word displays (e.g., SAND-LANE) and were post-cued to report either the left or right hand word. A migration error is defined as a response that intrudes a letter presented in the other display item. For example, in the example display, if the cue indicated that the left hand word was the target, a migration error response would be "LAND" or "SANE". Mozer established several further facts bout this phenomenon, including the "surround similarity effect", the finding that such errors are more likely if the two words have letters in common (as in SAND LANE) than if they do not (as in SAND LOVE). McClelland (1985) provided a simulation of these and other aspects of Mozer's data. The account is based on the programming of local networks for word recognition, based on simultaneous access to a word-processing network from the output of letter level processes in each of the two local networks. Based on the assumption that migration errors arise from this lexical access process, several predictions follow: 1) migrations should be more likely for letters in words than for letters imbedded in unrelated contexts such as digit strings. That is, subjects post-cued to report the contents of a particular letter position in a display should be more likely to misreport the first letter of the left hand string as an L in "SAND LAND" than in "S444 L444". 2) The surround similarity effect should not extend to letters in digit strings. 3) Migration errors should be much more frequent if the migration error response forms a word than if it forms a pseudoword. 4) migrations should be as likely between words in different cases (SAND lane) as between words in the same case (sand lane or SAND LANE). All of these predictions were confirmed. There are some migration errors for letters imbedded in digit strings, but there are far more for letters in words, and there is a large surround similarity effect for letters in words but not for letters in digits. And the probability of making a migration error was much greater if the migration error was a word than if it was a pseudoword. Finally, migration errors were in fact slightly more likely between words of different cases than between words typed both in the same case. A fourth experiment, using a variant of Reicher's forced choice procedure, ruled out a guessing interpretation of the letter migration effect.

Taken together the results confirm several of the basic principles of the CID model, and suggest strongly that migration errors are, at least in part, the result of simultaneous access to knowledge of words.

2.3.5. Steps toward a PDP model of Sentence Processing

One of our eventual goals for the contract is to implement a sentence processing mechanism that is capable of processing full sentences, possibly containing embedded clauses. As a first step, McClelland and Kawamoto (in preparation) have developed a model for processing single sentoids, consisting of a verb and a collection of noun-phrase and prepositional phrase arguments. The goal of the model is to take the syntactic structure of the sentence as input and to produce as output a representation of the case-structure of the sentence. The latter structure provides a description of which noun-phrase (if any) is the agent of the sentence, which is the patient, which is the instrument; it also indicates whether any of the pp's should be treated as constituents of complex noun-phrases, as in "the man with the hat".

It should be noted that the task that the model must perform is a complex one, since it is not the case that there is a consistent mapping between surface position in the sentence and case role. Consider, for example, sentences involving the verb "break".

The boy broke the window.

The hammer broke the window.

The window broke.

The boy broke the window with the curtain.

The boy broke the window with the hammer.

In each of the first three sentences, the initial noun phrase (surface subject) of the sentence plays a different underlying role (agent, instrument, or patient). The patient can occur as subject or as object. The instrument can appear as subject or in the phrase "with the hammer". Finally, the phrase "with the hammer" might carry the instrument, or it might carry a modifier or the preceding noun phrase.

Even though word order information is not always reliable, it can and must sometimes be used. Thus, the only way we have of knowing who was the agent in the following two sentences is by a strategy that assigns the first noun as agent and the second as patient:

The boy hit the girl.

The girl hit the boy.

The model that has been developed by McClelland and Kawamoto provides a mechanism for dealing with these problems. It consists of a simple parallel-distributed processing system, in which the surface syntactic form of a sentence assignments of the arguments to case roles are both represented as patterns of activation over collections of units. Each collection of units represents a particular (surface or case) role, and the pattern of activation over the units in each collection of units provides a characterization of the basic semantic properties of the filler of the role. The model learns connections between the surface role units and the case role units that allow it to generate the appropriate case role patterns from given surface role patterns. It learns through experiences in which it sees surface patterns paired with the correct role patterns. What it learns can be generalized from the training sentences to novel sentences that it has never seen before, and to novel nouns and verbs that share semantic properties with familiar nouns and verbs. For example, if it has learned how to interpret "The boy walked" and "The man walked the dog", it can handle almost as well sentences such as "The boy ran" and "The man ran the puppy". The model exhibits a number of other properties that we think are characteristic of the human information processing mechanisms: 1) It is able to use context to select the correct meaning of an ambiguous word, so that it gets the correct meaning of bat in each of the following sentences: "The boy hit the ball with the bat" and "The bat ate the cheese". It exhibits an interesting tendency to interpret a sentence like "The bat broke the window" in two different ways; the pattern of activation in the model indicates that insofar as the bat is agent it is a flying bat and insofar as it is instrument it is a baseball bat. Both alternatives are simultaneously active in the network. 2) The model also has a tendency to modify the meaning patterns it assigns to words in contextually appropriate ways. Thus in it's response to "The girl broke the window with the ball" the ball is characterized as hard, even though the ball the model was actually trained on was described to it as soft. The reason is that instruments of breaking are generally hard, and the model picks up on this fact. 3) The model fills in default values for missing arguments. For example, in "the boy ate", the model will fill in a pattern of activation on the patient units that is characteristic of food, but which does not specify the particular variety of food in detail.

2.3.6. The next step

The next step our attempt to understand sentence processing will be an attempt to embed the mechanisms of case-role assignment described above in a mechanism for parsing sentences with embedded clauses. Exactly what the parser will look like remains to be seen. However, it will be designed with several key facts about human sentence processing in mind. 1) Humans are capable of indefinite "tail recursion" or "right-branching", but have great difficulty with center embedded structures (compare "The boy who the man who the girl like hit ran" to "the girl liked the man who hit the boy who ran.") 2) Humans exploit semantic constraints to help them assign arguments both to syntactic roles and to case roles (the syntactic structure of "The boy ate the spaghetti with the fork" is much different from the structure of "The boy ate the spaghetti with the cheese".) 3) Humans appear to adhere to what we call *the principle of immediate update*, in that they appear to update their understanding of sentences on line, as each new word comes in, taking both syntactic and semantic information into account. 4) More controversially, humans appear to be able to entertain two alternative structural possiblities, suspending a final choice between either one, if both seem equally likely, at least over a span of a few words. 5) There are certain syntactic principles (e.g., the principle of minimal attachment) but that these principles are easily over-ridden by semantic constraints (cf Ford, Bresnan, and Kaplan, 198?).

The major challenge facing this work is not the actual construction of a parser using parallel-distributed processing mechanisms. For example, the CID scheme described above could probably be used to implement a parser similar to the parser proposed by Marcus (198?). The difficulty would be in preserving the beneficial aspects of the use of parallel-distributed processing mechanisms in doing so. Just how far we are from an adequate way of meeting this challenge remains to be seen.

3. Change in Key Personnel

There have been no changes in key personnel.

4. Summary of Substantive Information Derived from Special Events

None.

5. Problems Encountered and/or Anticipated

None.

6. Action Required by the Government

None.

7. Fiscal Status

1. Amount currently provided on contract: $200,251.
2. Expenditures and commitments to date: 190,798
3. Funds required to complete work: 274,917
4. Estimated date of completion of work: 30 November, 1987

8. Publications in 1985

Elman, J. L. and McClelland, J. L. Exploiting the lawful variability in the speech wave. In J. S. Perkell and D.H. Klatt (Eds.), *Invariance and variability of speech processes.* Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc. 1985.

Elman, J. L. and McClelland, J. L. An architecture for parallel processing in speech recognition: The TRACE model. In M. R. Schroeder (Ed.), *Speech recognition.* Basel: S. Krager AG. 1985.

McClelland, J. L. Putting knowledge in its place: A scheme for programming parallel processing structures on the fly. *Cognitive Science,* 1985, *9,* 113-146.

McClelland, J. L. and Rumelhart, D. E. Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General,* 1985, *114,* 159-188.

Rumelhart, D. E. and McClelland, J. L. Levels indeed! A response to Broadbent. *Journal of Experimental Psychology: General,* 1985, *114,* 193-197.

McClelland, J. L. (1985). Distributed models of normal and amnesic memory. In Olton, D., Gamzu, E., and Corkin, S. (Eds.) *Memory Dysfunctions: An integration of animal and human research.* New York: New York Academy of Sciences.

Stemberger, J.P., Elman, J.L., & Haden, P. Interference between phonemes during phoneme monitoring:

Evidence for an interactive activation model of speech perception. *Journal of Experimental Psychology: Human Perception and Performance,* 1985, *11,* 475-489.

In Press:

McClelland, J. L. & Elman, J. L. (in press). The TRACE model of speech perception. *Cognitive Psychology.* [This paper is essentially the same as the technical report of the same title].

Rumelhart, D. E., McClelland, J. L., and the PDP research group. (in press). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume I.* Cambridge, MA: Bradford Books.

McClelland, J. L., Rumelhart, D. E., and the PDP research group. (in press). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume II.* Cambridge, MA: Bradford Books.

Chapters of these two volumes specifically supported by the present contract include:

McClelland, J. L. The programmable blackboard model of reading.

McClelland, J. L. Resource requirements of standard and programmable nets.

McClelland, J. L., & Elman, J. L. Speech perception through interactive activation: The TRACE Model.

Accepted for publication pending minor revisions:

McClelland, J. L. & Mozer, M. C. (accepted pending minor revisions) Perceptual interactions in multi-word displays: Familiarity and similarity effects. Journal of Experimental Psychology: Human Perception and Performance.

Research completed, write-up in preparation:

McClelland, J. L. and Kawamoto, A. M. (in preparation). Mechanisms of Sentence Processing: Assigning roles to constituents. In J. L. McClelland, D. E. Rumelhart, and the PDP research group. *Parallel distributed processing: Explorations in the microstructure of cognition. Volume II.* Cambridge, MA: Bradford Books.

Elman, J.L. and McClelland, J.L. (in preparation). Testing the modularity perception: Evidence from Speech perception.

END

FILMED

1-86

DTIC