Human Neurobiol (1986) 5:1–22

Human Neurobiology

Neural dynamics of speech and language coding: developmental programs, perceptual grouping, and competition for short-term memory

M. Cohen and S. Grossberg*

Center for Adaptive Systems, Boston University, 111 Cummington Street, Boston, MA 02215, USA

Summary. A computational theory of how an observer parses a speech stream into context-sensitive language representations is described. It is shown how temporal lists of events can be chunked into unitized representations, how perceptual groupings of past item sublists can be reorganized due to information carried by newly occurring items, and how item information and temporal order information are bound together into context-sensitive codes. These language units are emergent properties due to intercellular interactions among large numbers of nerve cells. The controlling neural networks can arise through simple rules of neuronal development: random growth of connections along spatial gradients, activity-dependent self-similar cell growth, and competition for conserved synaptic sites. Within these networks, a spatial frequency analysis of temporally evolving activity patterns leads to competitive masking of unappropriate list encodings in short term memory. The neurons obey membrane equations undergoing shunting recurrent on-center off-surround interactions. Several design principles are embodied by the networks, such as the sequence masking principle, the long-term memory invariance principle, and the principle of self-similar growth.

Key words: Unitization – Temporal order – Speech – Competition – Neural networks

Introduction: Context-sensitivity of self-organizing language units

One of the fundamental problem areas in speech and language research concerns the characterization of the functional units into which speech sounds are integrated by a fluent speaker. A core issue concerns the *context-sensitivity* of these functional units, or the manner in which the perceptual grouping into functional units can depend upon the spatio-temporal patterning of the entire speech stream. Such context-sensitivity is evident on every level of speech and language organization. For example, a word such as Myself is used by a fluent speaker as a unitized verbal chunk. In different verbal contexts, however, the components My, Self, and Elf or Myself are all words in their own right. Moreover, although an utterance which ended at My would generate one grouping of the speech flow, an utterance which went on to include the entire word Myself could supplant this encoding with one appropriate to the longer word. Thus in order to understand how context-sensitive language units are perceived by a fluent speaker, one must analyse how all possible groupings of the speech flow are analysed through time, and how certain groupings can be chosen in one context without preventing other groupings from being chosen in a different context.

This problem has been stated in different ways by different authors. Darwin (1976) has, for example, asked how "our conscious awareness . . . is driven to the highest level present in the stimulus." Repp (1982) has noted "that the perception of phonetic distinctions relies on the integration of multiple acoustic cues and is sensitive to the surrounding context in very specific ways . . . listeners make continuous use of their tacit knowledge of speech patterns." Studdert-Kennedy (1980) has written that "The view of speech perception that seems to be emerging . . . is of an active continuous process . . . of perceptual integration across the syllable."

The functional units into which a fluent speaker groups a speech stream are dependent upon the observer's prior language experiences. For example, a unitized representation for the word Myself does not exist in the brain of a speaker who is unfamiliar with this word. Thus an adequate theory of how an observer parses a speech stream into context-sensitive language units needs to analyse how developmental and learning processes bias the observer to experience some perceptual groupings above others. Such developmental and learning processes are often called processes of *self-organization* in theoretical biology and physics (Basar et al. 1983). Lindblom et al. (1983) have recently suggested the importance of self-organizing processes in speech perception.

The present article contributes to a theory of speech and language perception which arose from an analysis of how a language system self-organizes in real-time in response to its complex input environment (Grossberg 1978a, 1982a). This approach emphasizes the moment-by-moment dynamical interactions that control language development, learning, and memory. Within this theory, properties of language performance emerge from an analysis of the system constraints that govern stable language learning. This analysis has led to the discovery of a small number of dynamical principles

---

* Correspondence: Professor Stephen Grossberg, address above

and mechanisms which have been used to unify and predict a large data base concerning speech and language. Data concerning lexical decisions, recognition and recall of previous occurrences, development of circular reactions, imitation and unitization of novel sounds, matching phonetic to articulatory requirements, serial and paired associate verbal learning, free recall, categorical perception, temporal order information in short term memory, selective adaptation, auditory contrast, and word superiority effects have been analysed and predicted using this theoretical framework (Grossberg 1969, 1978a, 1978b, 1982a, 1984a, 1985; Grossberg and Pepe 1971; Grossberg and Stone 1986a, 1986b). These articles should be consulted for analyses of relevant data and of alternative models.

We believe that the unifying power of the theory is due to the fact that principles of self-organization – such as the laws regulating development, learning, and unitization – are fundamental in determining the design of behavioral mechanisms. This perspective suggests that the lack of alternative unifying accounts of this data base is due to the use of models that do not sufficiently tap the principles of self-organization that govern behavioral designs.

Developmental rules imply cognitive rules as emergent properties of neural network interactions

The present article quantitatively analyses and further develops a core process within this theory: the process whereby internal language representations encode a speech stream in a context-sensitive fashion. This process can be stated in several equivalent ways: how temporal lists of events are chunked into unitized representations; how perceptual groupings of past item sublists are reorganized due to information carried by newly occurring items; how item information and temporal order information are bound together to generate maximally predictive encodings of temporally occurring lists. This article briefly reviews the principles which Grossberg (1978a, 1984a) proposed for this process, outlines real-time networks which we have further developed to instantiate the principles, and demonstrates the competence of the networks using massive computer simulations.

These networks can be interpreted as networks of neurons whose interconnections arise through simple rules of neuronal growth and development. The context-sensitive language representations within these networks are emergent properties due to intercellular interactions among large numbers of nerve cells. These properties are not built into the individual cells. Nor are there any serial algorithms or cognitive rule structures defined within the network. Instead, the networks illustrate how simple rules of neuronal development – on the cellular level – can give rise to a system which acts *as if* it obeys complex rules of context-sensitive encoding – on the cognitive level.

This is not the only way in which the theory relates different levels of behavioral organization. We also show how organizational principles which are critical in visual processing can be specialized for use in language processing. In other words, similar mechanisms can be used both for spatial processing and for temporal processing. The theory hereby illustrates how a small number of dynamical laws can unify data on several levels of organization, ranging from microscopic rules of neuronal development to macroscopic properties of cognitive coding, and across modalities such as vision and audition.

Macrocircuit for the self-organization of recognition and recall

The encoding, or chunking, process which is analysed herein takes place within the macrocircuit depicted in Figure 1. This macrocircuit governs self-organization of language recognition and recall processes via a combination of auditorily-mediated language processes (the levels A_i), visual recognition processes (level V^*), and motor control processes for language production (the levels M_j). These stages interact internally via conditionable pathways (black lines) and externally via environmentally-mediated auditory feedback of self-generated sounds (dotted lines).

All the stages A_i and M_j within the theory obey similar general network laws. These laws describe cooperative and competitive interactions among the cells, or nodes, that exist at each level. Such cooperative-competitive interactions endow the network levels with properties of cellular activation and short term memory (STM). Different levels exhibit specialized properties of STM due to two types of factors: differences in the interconnections and other parameters of the cells at each level; and the very fact that the different levels, by occurring within different locations of the total network hierarchy, receive different types of inputs. One task of the theory is to show how a wide variety of STM properties can be generated from a small number of STM laws by choosing specialized intercellular wiring diagrams.

All of the learning and long term memory (LTM) processes within the theory occur in its inter-level pathways. All of these learning processes also obey similar dynamical laws.

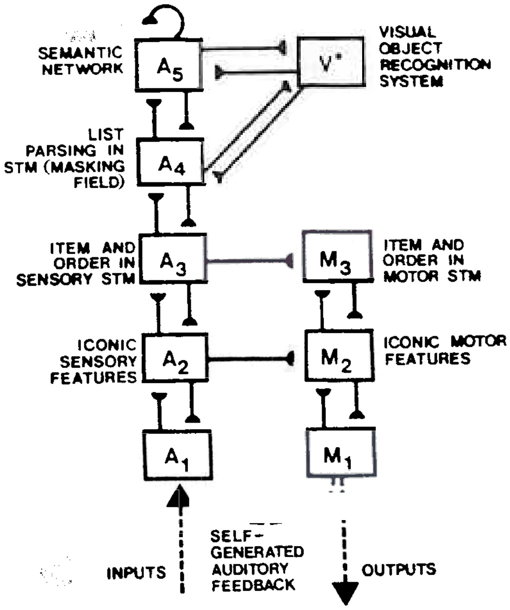


Fig. 1. A macrocircuit governing self-organization of recognition and recall processes: the text explains how auditorily mediated language processes (the A_i), visual recognition processes (V^*), and motor control processes (the M_j) interact internally via conditionable pathways (black lines) and externally via environmental feedback (dotted lines) to self-organize the various processes which occur at the different network stages

They encode different types of information due to their different parameter choices and their different locations within the total network hierarchy.

Masking fields

The present article focuses upon the design of level A_4 in this network hierarchy (Fig. 1). Level A_4, which is called a *masking field*, generates a context-sensitive encoding of the activation patterns that flicker across level A_3 through time. The activation patterns across A_3 influence A_4 via the conditionable pathways from A_3 to A_4. We will describe how developmental rules for growth of connections from A_3 to A_4 and for growth of connections within A_4 enable A_4 to achieve a context-sensitive parsing of A_3's activity patterns. In order to understand this masking field design, we review the design problems and principles which led to its discovery.

The temporal chunking problem: seeking the most predictive representation

The core problem leading to masking field design is called the *temporal chunking problem* (Grossberg 1978a, 1984a, 1985). Consider the problem of unitizing an internal representation for an unfamiliar list of familiar items; e.g. a novel word composed of familiar items, such as phonemes or syllables. The most familiar groupings of the list are the items themselves. In order to even know what the novel list is, all of its individual items must first be presented. All of these items are more familiar than the list itself. What mechanisms prevent item familiarity from forcing the list always to be processed as a sequence of individual items, rather than eventually as a whole? How does a not-yet-established word representation overcome the salience of well-established phoneme or syllable representations? How does unitization of unfamiliar lists of familiar items even get started? If the temporal chunking problem is not solved, then unitized internal representations of lists with more than one item can never be learned.

Another version of the temporal chunking problem becomes evident by noticing that every sublist of a list is a perfectly good list in its own right. Letters, syllables, and words are special sublists that have achieved a privileged status due to experience. In order to understand how this privileged status emerges, we need to analyse the processing substrate upon which all possible sublists are represented before learning occurs. This processing substrate exists within the theory at level A_3. Then we need to examine how prewired network processes interact with network learning processes to determine which of these sublists will succeed in activating a unitized representation at level A_4.

The subtlety of this unitization process is reflected by even the trivial fact that novel words composed of familiar items can be learned. This fact shows that not all sublists have equal prewired weights in the competitive struggle to be represented at A_4. Such prewired weights include the number of coding sites in a sublist representation and the strength of the competitive intercellular signals that are emitted from each sublist's representation. Somehow a word as a whole can use such prewired processing biases to competitively inhibit, or to *mask*, the learned potency of its constituent items. That is why the cooperative-competitive design of A_4 that solves the temporal chunking problem is called a *masking field*.

One property of the masking field design is that longer lists, up to some maximal length, can selectively activate cells that have a prewired competitive advantage over shorter sublists in the struggle for STM activation and storage. Such a competitive advantage enables a masking field to exploit the fact that longer sublists, other things being equal, are better predictors of subsequent events than are shorter sublists because they embody a more unique temporal context. Thus a masking field is designed to generate STM representations which have the best a priori chance to correctly predict the activation patterns across A_3. As an important side benefit, the a priori advantage of longer, but unfamiliar, sublists enables them to compete effectively for STM activity with shorter, but familiar, sublists, thereby providing a solution to the temporal chunking problem.

The word length effect

The postulate that longer sublists, up to some maximal length, have a competitive STM advantage led to the prediction of a word length effect in Grossberg (1978a, Section 41; reprinted in Grossberg 1982a). A word length effect was reported in the word superiority studies of Samuel et al. (1982, 1983). In these experiments, a letter was better recognized when it was embedded in longer words of lengths from 1 to 4. As Samuel et al. (1983, p 322) have noted, other "lexical theories had not previously included mechanisms that were explicitly length dependent." We believe this is because other lexical theories did not state, nor attempt to solve, the temporal chunking problem. Further discussion of the word length effect and related data is provided in Grossberg (1984a, 1985).

In the light of the word length effect, the conclusion that longer sublists have an a priori competitive advantage over shorter sublists, up to some maximal length, may seem to be self-contradictory. If prewired word biases can *inhibit* learned letter biases, then how is perception of letters *facilitated* by a word context, which is the main result of word superiority studies? This paradox can also be resolved through an analysis of how A_4 encodes activity patterns across A_3.

All letters are sublists: which computational units can self-organize?

A resolution of this paradox can be derived by further considering what it means to say that every sublist of a list is also a list. In order for sublists of a list to struggle for representational status, sets of individual items of the list need first to be simultaneously represented at some level of processing, which we identify with A_3. The theory shows how item representations that are simultaneously active in STM across A_3 can be grouped, or chunked, into representations of sublists at the next level of processing A_4. The sublist representations can then compete with each other for STM activation within A_4. Once the two levels A_3 and A_4 are clearly distinguished, it becomes obvious that individual list items, being sublists, can be represented at A_4 as well as at A_3. In the special case of letters and words, this means that letters are represented at the item level, as well as at the list level. Prewired word biases can inhibit learned letter biases at the level A_4, but not at the level A_3. Excitatory top-down priming from A_4 to A_3 and from A_4 to V^* can then support the enhanced letter recognition that obtains during word superiority experiments.

To clearly understand how the item representations at A_3 differ from the sublist representations at A_4, one must study the theory's processes in some detail. Even without such a study, one can conclude that "all letters are sublists." Indeed, all events capable of being represented at A_4 exist on a equal dynamical footing. This conclusion clarifies how changes in the context of a verbal item can significantly alter the processing of that item, and why the problem of identifying the functional units of language has proved to be so perplexing (Darwin 1976; Studdert-Kennedy 1980; Young 1968). In A_4, no simple verbal description of the functional unit, such as phoneme, syllable, or word, has a privileged status. Only the STM patterns that survive a context-sensitive interaction between associative and competitive rules have a concrete existence.

The conclusions that "all letters are sublists" implies the use of different computational units than one finds in many other models of language processing. In other models, levels such as A_3 and A_4 often represent letters and words, respectively. In the present theory, levels A_3 and A_4 represent items (more precisely, patterns of temporal order and item information in STM) and lists (more precisely, sublist parsings in STM), respectively. Thus in our theory, *all* familiar letters possess both item and list representations, not just letters such as A and I that are also words.

This property helps to explain the data of Wheeler (1970) showing that letters such as A and I, which are also words, are not recognized more easily than letters such as E and F, which are not also words (Grossberg 1984a, 1985). In a model which postulates a letter level and a word level, letters such as A and I are represented on both the letter level and the word level, whereas letters such as E and F are represented only on the letter level. In such a model, letters such as A and I might be expected to be better recognized than letters such as E and F. Choosing letter and word levels thus leads to serious data-related difficulties, including the inability to explain the Wheeler (1970) data and the Samuel et al. (1982, 1983) data without being forced into further paradoxes.

More generally, any model whose nodes represent letters and words, and only these units, faces the problem of describing what the model nodes represented before a particular letter or word enters the subject's lexicon, or what happens to these nodes when such a verbal unit is forgotten. This issue hints at the core problem that such a model cannot self-organize (Grossberg 1984a, 1985). The self-organization process which controls language processing hides the mechanistic substrate upon which it is built. Concepts from lay language, such as letters and words, provide a misleading tool for articulating the computational units which are manipulated by a self-organizing language system. Repp (1981, p 1462) has made the similar point "that linguistic categories are abstract and have no physical properties, and that, therefore, their physical correlates in the speech wave are appropriately described in acoustic terms only." This problem can be dealt with using a theory whose levels can learn to encode abstract item and list representations within a substrate of previously uncommitted nodes or cells.

Self-organization of auditory-motor features, items, and synergies

The conclusion that increasingly abstract computational units are activated at higher levels of a self-organizing language system does not deny the fact that more concete entities, such as auditory features and phonemes, are activated at earlier processing stages. Within our theory, however, even these entities are emergent properties due to intercellular interactions. Before describing A_4 in detail we briefly review properties of levels A_1, A_2, and A_3 to clarify the meaning of the activity patterns across A_3 that A_4 can encode. The mechanisms which rigorously instantiate this intuitive review are described in detail in Grossberg (1978a) and reviewed in Grossberg and Stone (1986a).

At an early stage of development, the environmentally activated auditory patterns at stage A_1 in Figure 1 start to tune the long-term memory (LTM) traces within the pathways from A_1 to A_2, and thus to alter the patterning of short-term memory (STM) auditory "feature detector" activation across A_2. After this LTM tuning process begins, it can be supplemented by a "babbling" phase during which endogenous activations of the motor command stage M_1 can elicit simple verbalizations. These verbalizations generate environmental feedback from M_1 to A_1 which can also tune the $A_1 \rightarrow A_2$ pathways. The learning within the feedback pathway $M_1 \rightarrow A_2 \rightarrow A_2$ helps to tune auditory sensitivities to articulatory requirements. This process clarifies aspects of the motor theory of speech perception (Cooper 1979; Liberman et al. 1967; Liberman and Studdert-Kennedy 1978; Mann and Repp 1981; Repp and Mann 1981; Studdert-Kennedy et al. 1970).

Just as the auditory patterns across A_1 tune the $A_1 \rightarrow A_2$ LTM traces, the endogenously activated motor command patterns across M_1 tune the $M_1 \rightarrow M_2$ LTM traces. The activation patterns across M_2 encode the endogenously activated motor commands across M_1 into "motor features" using the same mechanisms by which the activation patterns across A_2 encode the exogenously activated auditory patterns across A_1 into "auditory features."

The flow of adaptive signalling is not just bottom-up from A_1 to A_2 and from M_1 to M_2. Top-down conditionable signals from A_2 to A_1 and from M_2 to M_1 are also hypothesized to exist. These top-down signal patterns represent learned expectancies, or templates. Their most important role is to stabilize the learning that goes on within the adaptive pathways $A_1 \rightarrow A_2$ and $M_1 \rightarrow M_2$. In so doing, these top-down signal patterns also constitute the read-out of optimal templates in response to ambiguous or novel bottom-up signals. These optimal templates predict the patterns that the system expects to find at A_1 or M_1 based upon past experience. The predicted and actual patterns merge at A_1 and M_1 to form completed composite patterns which are a mixture of actual and expected information.

Auditory and motor features are linked via an associative map from A_2 to M_2. When M_1 is endogenously activated, it activates a motor representation at M_2 via the adaptive pathway $M_1 \rightarrow M_2$, as well as an auditory representation at A_2 via environmental feedback $M_1 \rightarrow A_1$ and the adaptive pathway $A_1 \rightarrow A_2$. Since A_2 and M_2 are then simultaneously active, the associative map $A_2 \rightarrow M_2$ can be learned. This map also links auditory and articulatory features.

The associative map $A_2 \rightarrow M_2$ enables the imitation of novel sounds – in particular, of non self-generated sounds – to get underway. It does so by analysing a novel sound via the bottom-up auditory pathway $A_1 \rightarrow A_2$, mapping the activation patterns of auditory feature detectors into activation patterns of motor feature detectors via the associative map $A_2 \rightarrow M_2$, and then synthesizing the motor feature pattern into a net motor command at M_1 via the top-down motor template $M_2 \rightarrow M_1$. The motor command, or synergy, that is synthesized in this way generates a sound that is closer to the novel sound than are any of the sounds currently coded by the system. The properties whereby the learned map $A_1 \rightarrow A_2 \rightarrow M_2 \rightarrow M_1$ enables imitation of novel sounds to occur clarifies with the analysis-by-synthesis approach to speech recognition (Halle and Stevens 1962; Stevens 1972; Stevens and Halle 1964).

The environmental feedback from M_1 to A_1 followed by the learned map $A_1 \rightarrow A_2 \rightarrow M_2 \rightarrow M_1$ defines a closed feedback loop, or "circular reaction" (Piaget 1963). The theory's explication of the developmental concept of circular reaction helps to clarify the speech performance concepts of motor theory and analysis-by-synthesis in the course of suggesting how an individual can begin to imitate non-self-generated speech sounds.

The stages A_2 and M_2 can each process just one spatial pattern of auditory or motor features at a time. Thus A_2 can process an auditory "feature code" that is derived from a narrow time slice of a speech spectrogram, and M_2 can control a simple motor synergy of synchronously coordinated muscle contractions. These properties are consequences of the fact that spatial patterns, or distributed patterns of activity across a field of network nodes, are the computational units in these real-time networks. This computational unit is a mathematical consequence of the associative learning laws that govern these networks (Grossberg 1982a). The later stages A_i and M_j in Figure 1 are all devoted to building up recognition and recall representations for temporal groupings, or lists, of spatial pattern building blocks.

A spatial pattern of activation across A_2 encodes the relative importance of all the "feature detectors" of A_2 which represent the auditory pattern that is momentarily activating A_1. In order to encode temporal lists of auditory patterns, one needs to simultaneously encode a sequence of spatial patterns across A_2's auditory feature detectors. The following way to accomplish this also addresses the fundamental problem that individual speech sounds, and thus their spatial patterns across A_2, can be altered by the temporal context of other speech sounds in which they are embedded.

In addition to activating the associative map from A_2 to M_2, each spatial pattern across A_2 also activates an adaptive pathway from A_2 to A_3. Although all the adaptive pathways of the theory obey the same laws, each pathway learns different information depending on its location in the network. Since the $A_2 \rightarrow A_3$ pathway is activated by feature patterns across A_2, it builds up learned representations, or chunks, of these feature patterns. Each such representation is called an *item representation*. The item representations include the representations of phonemes.

All new learning about item representations is encoded within the LTM traces of the $A_2 \rightarrow A_3$ adaptive pathway. Although each item representation is expressed as a pattern of activation across A_3, the learning of these item representations does not take place within A_3. This flexible relationship between learning and activation is needed to understand how temporal codes for lists can be learned and performed. For example, as a sequence of sound patterns activates A_1, the patterns of "auditory feature" activation across A_2 can build up and rapidly decay, via a type of iconic memory (Sperling 1960). These A_2 activation patterns, in turn, lead to activation of item representations across A_3. The item representations are stored in STM, as a type of "working memory" (Cermak and Craik 1979), due to the feedback interactions within A_3. As a succession of item representations across A_3 is stored in STM, the *spatial pattern* of STM activity across A_3 represents *temporal order information* across the item representations of A_3.

Temporal order information across item representations: the spatial recoding of temporal order

As more items are presented, the evolving spatial patterns of activity across A_3 include larger regions of the item field, up to some maximal length. Thus the *temporal* processing of items is converted into a succession of expanding *spatial* patterns within A_3. This is the main reason why spatial mechanisms that are used in visual processing can also be used to design the masking field A_4.

Each activity pattern across A_3 is a context-sensitive computational unit in its own right. In such a representation, changing any one activity changes the coded meaning of the entire list of items. The activity pattern "is" the code, and no further labels or algorithms are needed to define it. In order to understand how such a code works, it is necessary to specify laws for the unitized encoding and recognition of item sublists by A_4, and laws for the rehearsal and recall of items before and after they are unitized by A_4.

The LTM Invariance Principle

Before these tasks can be accomplished, it is first necessary to characterize the laws whereby items can reliably represent temporal order information via the spatial patterning of activation across A_3. These laws can be derived from an analysis of the self-organization process. In particular, an incorrect choice of STM laws within A_3 could cause an unstable breakdown of LTM within the conditionable pathways from A_3 to A_4.

Grossberg (1978a, 1978b) introduced the *LTM Invariance Principle* in order to derive STM laws for A_3 that are compatible with stable LTM encoding. This principle shows how to alter the STM activities of previous items in response to the presentation of new items so that the repatterning of STM activities that is caused by the new items does not inadvertently obliterate the LTM codes for old item groupings. For example, consider the word Myself from this perspective.

We would not wish the LTM codes for My, Self, and Elf to be obliterated just because we are learning the new word Myself. On the other hand, the predictive importance of the groupings My, Self, and Elf may be reduced by their temporal embedding within the list Myself. We therefore assume that A_3 is designed to satisfy the *LTM Invariance Principle:* The spatial patterns of temporal order information in STM are generated by a sequentially presented list in such a way as to leave the $A_3 \rightarrow A_4$ LTM codes of past event groupings invariant, even though the STM activations caused by these past groupings may change markedly across A_4 as new items activate A_3.

It turns out that a suitably designed cooperative-competitive interaction across A_3 can mechanistically realize this principle. This A_3 design has been used to analyse and predict data about free recall, serial verbal learning, intensity-time tradeoffs, backward coding effects, item grouping effects, and influences of presentation rate on recall order (Grossberg 1978a, 1978b, 1982b, 1985; Grossberg and Stone 1986b). For present purposes, we simply note that different STM activity patterns across the same set of item representations within A_3 can encode different temporal orderings of these items.

The emergence of complex speech and language units

The concept of temporal order information across item representations is necessary, but not sufficient, to explain how novel lists of items can be learned and performed. In addition, one needs to consider how the bottom-up conditionable pathway from M_2 to M_3 (Fig. 1) learns unitized representations of motor items (synergies); how the top-down conditionable pathway from M_3 to M_2 learns motor templates or expectancies that can read-out coarticulated performance of these synergies; and how the conditionable pathway from A_3 to M_3 learns an associative intermodality map. Using these mechanisms, one can analyse how novel item representations are formed. For example, suppose that an analysis-by-synthesis of a novel sound has been accomplished by the composite map $A_1 \rightarrow A_2 \rightarrow M_2 \rightarrow M_1$. Such a map generates a novel pattern of auditory features across A_2 and a novel pattern of motor features across M_2 (Section "Self-organization of auditory-motor features, items, and synergies"). These feature patterns can then trigger learning of unitized item representations at A_3 and M_3. These unitized representations can be learned even though the network never endogenously activated these feature patterns during its "babbling" phase. In this way, the network's learned item codes can continue to evolve into ever more complex configurations by a combination of imitation, self-generated vocalization, STM regrouping, and LTM unitization. An associative map $A_3 \rightarrow M_3$ between new unitized item representations also continues to be learned. Using this background, we can now summarize how a unitized representation of an entire list, such as a word, can be learned and performed.

List chunks, recognition, and recall

As the network processes a speech stream, it establishes an evolving STM pattern of temporal order information across the item representations of A_3. Since every sublist of a list is also a list, the conditionable pathway from A_3 to A_4 simultaneously "looks at", or filters, all the sublist groupings to which it is sensitive as the speech stream is presented through time. The masking field within A_4 then determines which of these sublist groupings will represent the list by being stored in STM at A_4.

These sublist representations contribute to the recognition of words (Grossberg and Stone 1986a) but cannot, by themselves, elicit recall. This raises the issue of how short novel lists of familiar items can be recalled even before they are unitized. The fact that a verbal unit can have both an item representation and a list representation (Section "All letters are sublists: which computational units can self-organize?") plays an important role.

Recall of a short novel list of familiar items is triggered by a nonspecific rehearsal wave to A_3 (Grossberg 1978a, 1978b). Such a wave opens an output gate that enables output signals of active items to be emitted from A_3 to M_3, with the most active item representations being read-out before less active item representations. As each item is read-out, it activates a negative feedback loop to itself that selectively inhibits its item representation, thereby enabling the next item representation to be read-out. Each item representation is recalled via the learned $A_3 \rightarrow M_3 \rightarrow M_2 \rightarrow M_1$ sensory-motor map.

This type of recall is immediate recall from STM, or working memory, of a list of unitized item representations. It is a type of "controlled" process, rather than being an "automatic" unitized recall out of LTM. In order for a unitized list chunk in A_4 to learn how to read-out its list of motor commands from LTM, the chunk must remain active long enough during the learning process to sample pathways to all of these motor commands. In the simplest realization of how temporal order information across item representations is encoded and read-out of LTM, the top-down template from A_4 to A_3 learns this information while the conditionable pathway from A_3 to A_4 is being tuned. Later activation of a list chunk in A_4 can read this LTM temporal order information into a pattern of STM temporal order information across the item representations of A_3. Activation of the rehearsal wave at this time enables the list to be read-out of STM. Unitized recall can hereby occur via the learned $A_4 \rightarrow A_3 \rightarrow M_3 \rightarrow M_2 \rightarrow M_1$ sensory-motor map.

The order of recall due to read-out of temporal order information from LTM is not always the order in which the items have been presented. Thus although the network is designed to stabilize learning and LTM insofar as possible, its interactions occasionally force a breakdown of temporal order information in LTM; for example, as occurs during serial verbal learning. See Grossberg (1982b, 1985) for recent analyses of how such breakdowns in temporal order information in LTM can occur.

The design of a masking field: spatial frequency analysis of item-order information

With this background, we can now turn to the quantitative design of the masking field A_4. As a sequence of items is temporally processed, the masking field updates its choice of list representation, parsing the item sequence into a predictive grouping of unitized sublist choices based on a combination of a priori parameter choices and past learning. A spatial pattern of STM activity across the item representations of A_3 provides the inputs which are grouped by A_4. As more items are presented, new spatial patterns are registered that include larger regions of the A_3 item field, up to some maximum list length. Thus the temporal processing of items is converted by A_3 into a succession of *expanding* spatial patterns.

Given this property, the temporal chunking problem can be rephrased as follows. How do sublist chunks in A_4 that encode broader regions of the item field mask sublist chunks that encode narrower regions of the item field? This insight can be rephrased as a principle of masking field design (Grossberg 1984a, 1985):

Sequence Masking Principle. Broader regions of the item field A_3 are filtered by the $A_3 \rightarrow A_4$ pathway in such a way that they selectively excite nodes in A_4 with stronger masking parameters.

In other words, A_4 is sensitive to the *spatial frequency* of the input patterns that it receives from A_3. We will show how nodes in A_4 which are selectively sensitive to a prescribed spatial frequency range define a masking subfield. Each masking subfield is characterized by a different choice of numerical parameters, which are determined by simple growth rules. Subfields whose cell populations have broader spatial frequencies and/or more coding sites can competitively mask STM activities of subfields with narrower spatial frequencies and fewer coding sites (Fig. 2).

This on-line list parsing capability must reconcile several properties that could be in conflict in a poorly designed system. For example, how does a short sublist activate one representation, yet an updated list that includes the sublist activate a different representation? Why does not the representation of the shorter list always inhibit the representation of the longer list? Why does not the converse hold? In short, how does the masking field automatically rescale itself to *selectively* respond to *all* list lengths and orderings, up to some maximal length?

Several properties are implicit in these design requirements; namely:

1. *Sequence representation.* All realizable item sequences, up to a maximal sequence length, can initially generate some differential reaction, however weak, in the masking field.

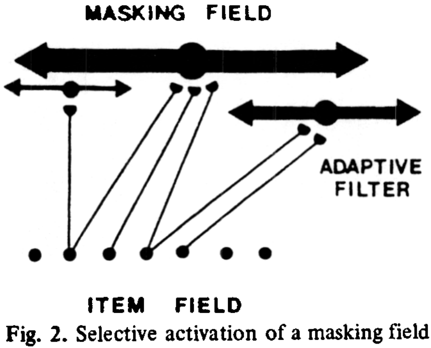


Fig. 2. Selective activation of a masking field. The nodes in a masking field are organized so that longer item sequences, up to some optimal length, activate nodes with more potent masking properies. Individual items, as well as item sequences, are represented in the masking field. The text describes how the desired relationship between item field, masking field, and the intervening adaptive filter can be self-organized using simple developmental rules

2. *Masking parameters increase with sequence length.* Critical masking parameters of masking field nodes increase with the length of the item sequences that activate them. This rule holds until an optimal sequence length is reached.

3. *Masking hierarchy.* Nodes that are activated by a given item sequence can mask nodes that are activated by subsequences of this sequence.

4. *Sequence selectivity.* If a node's trigger sequence has length n, it cannot be supraliminally activated by sequences of length significantly less than n.

Properties 1 and 2 suggest that the $A_3 \rightarrow A_4$ pathway contains a profusion of connections that are scattered broadly over the masking field. Property 3 suggests that closely related sequences activate nearby cells in the masking field. Postulate 4 says that, despite the profusion of connections, the tuning of long-sequence cells prevents them from responding to short subsequences.

The main problem is to resolve the design tension between profuse connections and sequence selectivity. This tension must be resolved both for short-sequence (e.g., letter) cells and long-sequence (e.g., word) cells: if connections are profuse, why are not short-sequence nodes unselective? In other words, what prevents many different item representations in A_3 from converging on every short-sequence cell in A_4 and thus being able to activate it? On the other hand, if many item representations from A_3 do converge on long-sequence cells in A_4, then aren't these long-sequence nodes activated by subsequences of the items? Somehow the growth rules that generate positional gradients in $A_3 \rightarrow A_4$ pathways and the competitive interactions within A_4 are properly balanced to achieve all of these properties.

Grossberg (1985) suggested how a combination of random growth rules in $A_3 \rightarrow A_4$ and activity-contingent self-similar growth rules within A_4 could achieve such a balance. These concepts led to several predictions concerning the developmental events that may regulate growth of lateral neural connections in response to afferent neural signals. In the present work, we have further developed these ideas to the point where the desired properties can be obtained even if critical numerical parameters in some of our networks are altered by a factor of 10. This numerical stability is all the more remarkable when one considers that *different* orderings of the *same* items, as well as the *same* orderings of items in *different* lists, can be selectively coded by such a masking field.

Development of a masking field: random growth and self-similar growth

The primary structure of a masking field can be understood in terms of two interacting growth rules: random growth of connections from A_3 to A_4, and self-similar growth of cells and connections within A_4. We now explain these concepts.

Suppose that each item node in A_3 sends out a large number of randomly distributed pathways towards the list nodes in A_4. Suppose that an item node randomly contacts a sequence node with a small probability p. This probability is small because there are many more list nodes than item nodes. Let λ be the mean number of such contacts across all of the sequence nodes. Then the probability that exactly

k pathways contact a given sequence node is given by the Poisson distribution

$$P_k = \frac{\lambda^k e^{-\lambda}}{k!} \quad . \qquad (1)$$

If K is chosen so that $K < \lambda < K + 1$, then P_k is an increasing function of k if $1 \leqslant k \leqslant K$ and a decreasing function of k if $k > K$. If λ is sufficiently small (approximately 4), then Eq. (1) implies that sequences of length $k \leqslant K$ will be represented within the masking field, thereby satisfying properties 1 and 2. Related random growth rules, such as the hypergeometric distribution, also have analogous properties.

Due to the broad and random distribution of pathways, list nodes will tend to be clustered near nodes corresponding to their sublists, thereby tending to satisfy property 3. A further property is also needed to satisfy property 3. Since a long-list node tends to mask all of its sublists, such a node must be able to send inhibitory signals to all the nodes which code these sublists. Thus the interaction range (viz., the axons) of an A_4 node should increase with the length of the list to which it is maximally sensitive (Fig. 2). This is called the *Principle of Self-Similar Growth* (Grossberg 1982a, 1985).

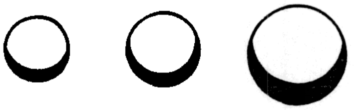
In order to realize property 4, an A_4 node that receives k pathways from A_3 somehow dilutes the input in each pathways so that (almost) all k pathways must be active to generate a suprathreshold response. As k increases, the amount of dilution also increases. This property suggests that long-list cells may have larger cellular volumes, since a larger volume can more effectively dilute a signal due to a single output pathway. Larger volumes also permit more pathways to reach the cell's surface, other things being equal. The constraint that long-list nodes are associated with larger parameters, such as number of sites and spatial frequencies, is hereby extended to include larger surface areas. This conclusion reaffirms the importance of the self-similarity principle in designing a masking field: a cell has longer interactions (viz., axons) because it has a larger cell body to support these interactions.

This discussion translates the formal properties 1–4 into two growth rules: random $A_3 \rightarrow A_4$ growth and self-similar $A_4 \rightarrow A_4$ growth. It remains to say how these two types of rules are joined together, as is required by the Sequence Masking Principle. In other words, how do larger cell surfaces attract more pathways, whereas smaller cell surfaces attract fewer pathways? Without further argument, a cell surface that is densely encrusted with axon terminals might easily be fired by a small subset of these axons. To avoid this possibility, the number of allowable pathways must be tuned so that the cell is never overloaded by excitation.

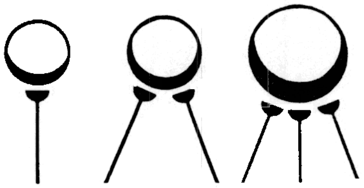
Activity-contingent self-similar cell growth

There exist two main ways to accomplish this property which have not yet been experimentally tested. A combination of the two ways is also possible:

Volume-dependent membrane receptors. At an early stage of development, a spectrum of cell sizes is endogenously generated across the masking field by a developmental program. Each cell of a given size contains a fixed number of membrane organelles that can migrate and differentiate into mature membrane receptors in response to developing input pathways (Patterson and Purves 1982). The number of membrane organelles covaries with cell size to prevent the internal level of cell excitation, say as measured by the maximum ratio of free internal Na^+ to K^+ ions, from becoming too large (Fig. 3).



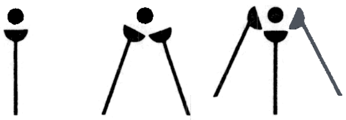
BEFORE
(a)



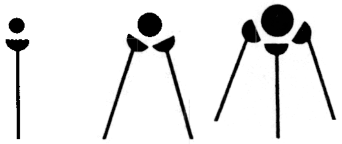
AFTER
(b)

Fig. 3a and b. Volume-dependent membrane receptors. **a** A spectrum of F_2 cell sizes is generated such that the number of membrane synaptic sites covaries with cell size. **b** More F_1 connections can arborize on the larger cells

Activity-dependent self-similar cell growth. Pathways from the item field grow to the list nodes via random growth rules. Before these pathways reach their target cells, these cells are of approximately the same size. As longer item lists



BEFORE
(a)



AFTER
(b)

Fig. 4a and b. Activity-dependent self-similar cell growth. **a** F_2 cells are initially all approximately the same size. **b** Variable numbers of F_1 cell connections across F_2 cells generate variable levels of average F_2 cell activation, which cause variable amounts of compensatory cell growth until a target average level of intracellular excitation is attained within all F_2 cells begin to be processed by A_3, these lists begin to activate their respective list nodes. The A_4 cells which receive many $A_3 \rightarrow A_4$ connections experience an abnormal internal cellular milieu (e.g., abnormally high internal Na^+/K^+ concentration ratios) due to the convergence of many active pathways on the small cell volumes. These large internal signals trigger self-similar cell growth that continues until the cell and its processes grow large enough to reduce the maximal internal signal to normal levels (Fig. 4).

The tuning of A_4 cell volume to the number of pathways from A_3 is thus predicted to be mediated by a self-similar use-and-disuse growth rule. The fact that internal cellular indices of membrane excitation trigger cell growth until these indices equilibrate to normal levels immediately shows why the mature cell needs simultaneous activation from most of its pathways before it can fire. A self-similar use-and-disuse growth rule has many appealing properties. Most notably, only item sequences that occur in the speaker's language during the critical growth period may be well-represented by the chunks of the speaker's masking field. This fact may be related to properties of second language learning.

In summary, the design of masking field can be realized by a simple developmental program: profuse random growth along spatial gradients from A_3 to A_4, which induces activity-contingent self-similar growth within A_4 that is constrained by competition for synaptic sites.

Sensitivity to multiple scales and intrascale variations

A masking field is sensitive to two different types of pattern changes.

Expanding patterns: temporal updating. As a word like Myself is processed, a subword such as My occurs before the entire word Myself is experienced. Figure 5a schematizes this type of informational change. As the word is presented, earlier STM activations are modified and supplemented by later STM activations. The STM pattern across A_3 expands as the word is presented. After Myself is fully presented, parts such as My, Self, and Elf are still (at least partially) represented within the whole. The masking field can nonetheless update its initial response to My as the remainder of Myself is presented. In this way, the masking field can react to the whole word rather than only its parts.

Internal pattern changes: temporal order information. The second type of masking field sensitivity is illustrated by the two words Left and Felt. This comparison is meant to be illustrative, rather than attempting to characterize the many subtle differences in context-sensitive alterations of sound patterns or reading patterns. The words Left and Felt illustrate the issue that the same *set* of items may be activated by different item *orderings*. To distinguish two such patterns, sensitivity to different spatial scales is insufficient because both lists may activate the same spatial scale. Instead, sensitivity to different STM patterns which excite the same set of items is required (Fig. 5b). The computer simulations summarized by Figures 6–13 below illustrate both types of masking field selectivity.

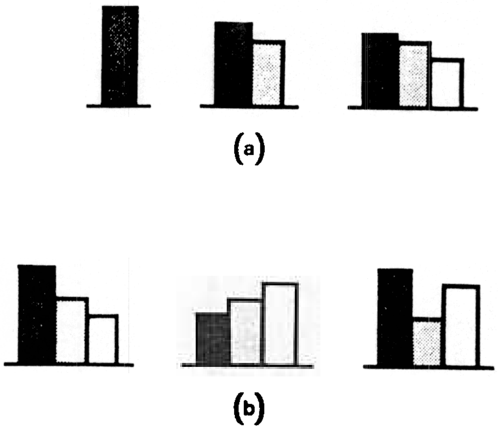


Fig. 5a and b. Two types of masking field sensitivity. **a** A masking field A_4 can automatically rescale its sensitivity to differentially react to activity patterns which activate variable numbers of A_3 cells. It hereby acts like a "multiple spatial frequency filter." **b** A masking field can differentially react to different A_3 activity patterns which activate the same set of A_3 cells. By **a** and **b**, it acts like a spatial pattern discriminator which can compensate for changes in overall spatial scale without losing its sensitivity to pattern changes at the finest spatial scale

Hypothesis formation, anticipation, evidence, and prediction

A third property of a masking field is of such importance that it deserves special mention. We will describe a masking field that is capable of simultaneously discriminating more than one grouping within a list. For example, such a masking field might respond to the A_3 representation of the word Myself by strongly activating an A_4 population that is sensitive to the whole word and weakly activating A_4 populations that are sensitive to the word's most salient parts. In such a representation, the total STM pattern across A_4 represents the A_3 STM pattern. The relative sizes of A_4's STM activities weight the relative importance of the groupings which are coded by the respective cell populations.

The suprathreshold STM activities across A_4 are approximately normalized, or conserved, due to its competitive feedback interactions. The STM activities across A_4 may thus be interpreted as a type of real-time probabilistic logic, or hypothesis-testing algorithm, or model of the evidence which A_4 has about the pattern across A_3.

Such a masking field also possesses a predictive, or anticipatory, capability. In response to a single item across A_3, the A_4 population which is most vigorously activated may code that item. In addition, less vigorous activations may arise at those A_4 populations which represent the most salient larger groupings of which the item forms a part. Such a masking field can anticipate, or predict, the larger groupings that *may occur* of which the item forms a part.

As more items are stored in STM across A_3, the set of possible groupings encoded by A_4 changes. In response to additional items, different groupings are preferred within A_4. Moreover, as more items are stored by A_3, A_4's uncertainty concerning the information represented at A_3 may decrease, much as the prediction of what follows ABC is less ambiguous than the prediction of what follows C alone. As A_4's uncertainty decreases, the spatial distribution of STM activity across A_4 becomes more focussed, or spatially localized. This type of spatial sharpening is not merely due

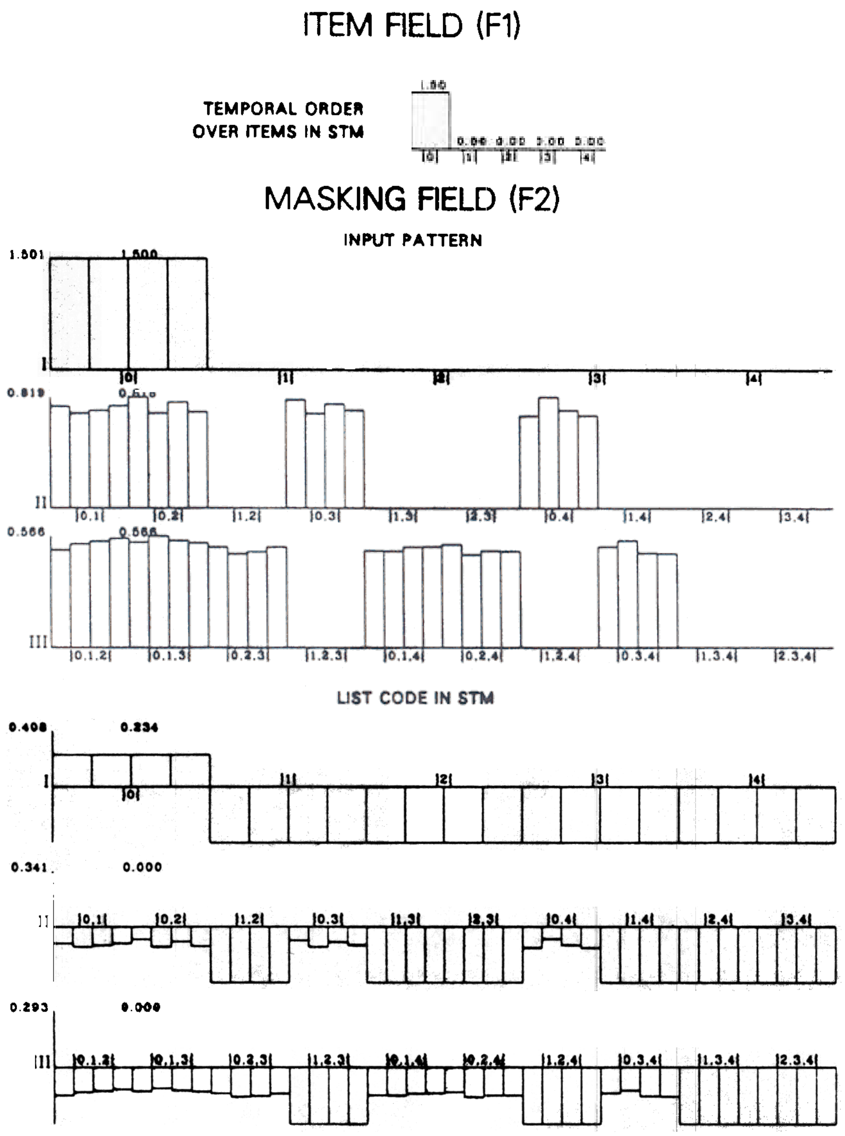


Fig. 6. List coding of a single item: network F_1 encodes in short term memory (STM) the pattern of temporal order information over item representations. In this figure, the single item $\{0\}$ is activated. Network F_2 encodes in STM the pattern of sublist chunks that are activated by F_1. The first three rows depict the inputs from F_1 to F_2. They are broadly distributed across F_2. The List Code in STM depicts the STM response to these inputs. Only the $\{0\}$ cells in F_2 are stored in STM, despite the broad distribution of inputs

to contrast enhancement. Rather it measures the degree of informational uncertainty within the A_4 code. These predictive, multiple-grouping properties of a masking field are illustrated by the computer simulations summarized in Figures 14–16.

Computer simulations

Before translating these developmental rules into mathematical equations, we describe some of our computer simulations of masking field properties. Figures 6–13 depict the simplest example of masking field dynamics. In this example, each distinct spatial pattern across A_3 chooses a unique nodal population within A_4. The same numerical parameters were used in all of these simulations. Only the input patterns varied. In Figures 14–16, a fixed but different set of parameters was chosen to illustrate how a masking field can generate spatially distributed and anticipatory sublist representations of spatial patterns across A_3. In these representations, the masking field is maximally sensitive to the entire list across A_3, but also generates partial activations to salient sublists and superlists of this list.

These figures can be combined in several ways to provide different insights into masking field properties. Consider Figure 6 to start. In this Figure, a single item in A_3 (which is denoted by F_1) is active. This item generates positive inputs to a large number of nodes in A_4 (which is denoted by F_2). The input sizes are depicted by the heights of the bars in the three rows labelled Input Pattern. Each row lists all F_2 nodes which receive the same number of pathways from F_1. The first row consists of F_2 nodes which receive one pathway, the second row consists of F_2 nodes which receive two pathways, and the third row consists of F_2 nodes which receive three pathways. In row 1, each F_2 node in the set labelled $\{i\}$ receives a pathway from the F_1 item node labelled $\{i\}$, $i = 0, 1, 2, \ldots, 4$. Note that four F_2 nodes receive inputs from the $\{0\}$ F_1 node. In row 2, all F_2 nodes labelled $\{0, 1\}$ receive pathways from the F_1 nodes $\{0\}$ and $\{1\}$. In row 3, all F_2 nodes labelled $\{0, 1, 2\}$ receive pathways from the F_1 nodes $\{0\}$, $\{1\}$, and $\{2\}$.

The input to *all* the F_2 nodes which receive pathways from the F_1 node $\{0\}$ are positive. There are 44 such nodes in Figure 6. Despite this fact, the only F_2 nodes capable of becoming active in STM are the nodes which receive path-

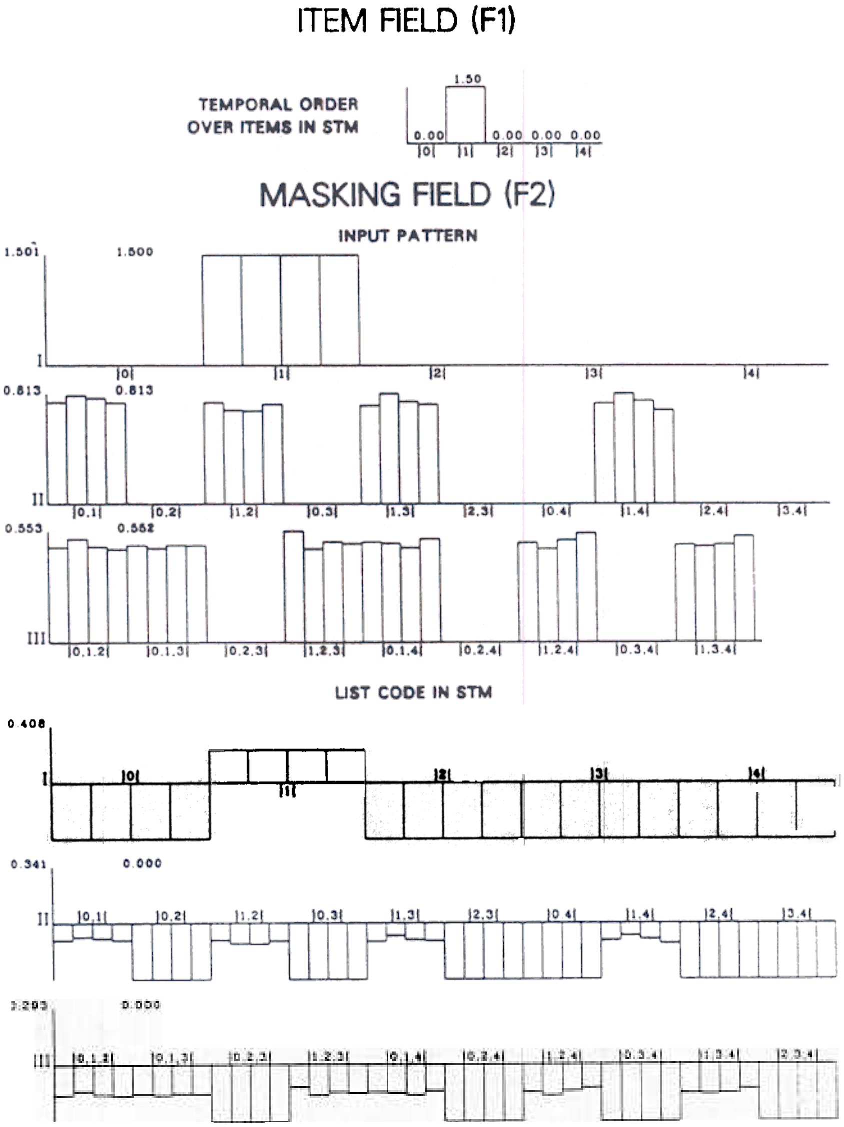


Fig. 7. List coding of a single item: in response to item $\{1\}$ in F_1, the masking field in F_2 chooses the $\{1\}$ cells in response to a broad distribution of inputs. Thus the level I of the List Code in STM responds selectively to individual items in F_1. The same thing is true in the next figure

ways *only* from the active item node $\{0\}$. These are the F_2 nodes labelled $\{0\}$. The STM activities of the F_2 nodes are listed in three rows under the heading List Code in STM. These are the activities which the nodes store in STM after the network equilibrates to the entire input pattern. Figure 6 illustrates how F_2 can transform a widespread input pattern into a focal, and appropriate, STM activation.

Figures 7 and 8 further illustrate this property. In each Figure, a different item at F_1 is activated. Each item generates a widespread input pattern to F_2. Each input pattern is contrast-enhanced into a focal STM activation. This STM activation is restricted to the F_2 nodes which receive pathways from only the active item node.

A comparison of Figures 6, 7, and 9 discloses a different property of masking field dynamics. Suppose that the temporally ordered list of items $\{0\}$, $\{1\}$ is received by F_1. The list as a whole generates a different spatial pattern across F_1 (Fig. 9) than does its first item (Fig. 6) or its second item (Fig. 7) taken in isolation. The list as a whole also activates even more nodes than does either item taken separately – 82 nodes in all. Despite this fact, only a single node become active in STM. This node is, moreover, an appropriate node because it is one of the F_2 $\{0, 1\}$ nodes that receive pathways only from the F_1 items $\{0\}$ and $\{1\}$. This comparison between Figures 6, 7, and 9 thus illustrates the following F_2 properties: sequence selectivity, and the ability of F_2 nodes which are activated by larger numbers of F_1 nodes to mask the activity of F_2 nodes which are activated by smaller subsets of F_1 nodes.

A comparison of Figures 9 and 10 reveals another important F_2 property. In both of these figures, the same set of F_1 items – $\{0\}$ and $\{1\}$ – is activated, but a different spatial pattern of activity exists across the items. The spatial pattern in Figure 9 may represent the temporally ordered list $\{0, 1\}$, whereas the spatial pattern in Figure 10 may represent the temporally ordered list $\{0, 1\}$, whereas the spatial pattern in Figure 11 may represent the temporally ordered list $\{1, 0\}$. The simulations show that F_2 is sensitive to the item pattern *as a whole*, because F_2 can generate different STM responses to these patterns even though they activate the same unordered set of F_1 nodes. In particular, in Figures 9 and 10, different F_2 nodes become active within the set of F_2 nodes which receives pathways only from items $\{0\}$ and $\{1\}$.

This comparison between Figures 9 and 10 clarifies what

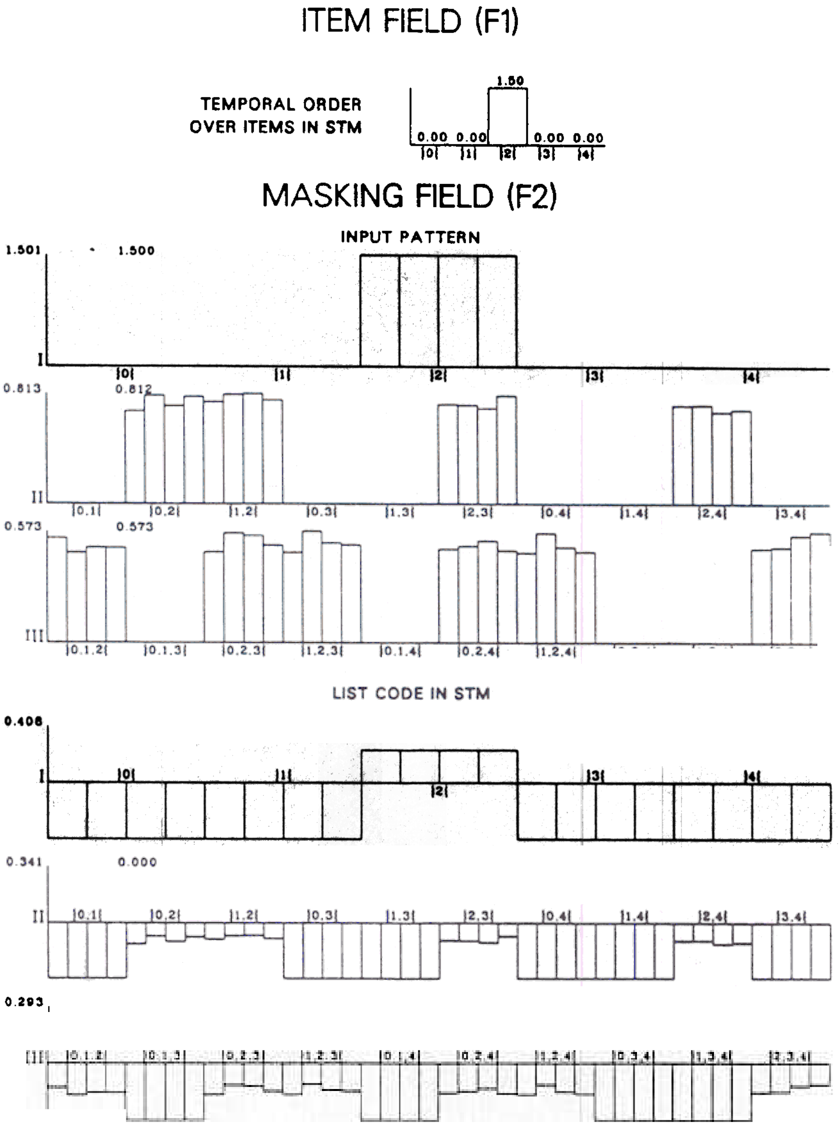


Fig. 8. See legend to Figure 7

we mean by the assertions that the spatial pattern across F_1 is the computational unit of the network, and that the differential STM responses of F_2 to these computational units embodies a context-sensitive list chunking process.

A comparison of Figures 6, 7, 8, 9, and 11 illustrates a more demanding variant of these F_2 properties. As a temporally ordered list of items {0} , {1} , {2} is processed by F_1, all the items become individually active at F_1 as the spatial patterns in Figures 6, 9, and 11 evolve through time. The final STM response in Figure 11 is, however, restricted to a single F_2 node, which is one of the nodes receiving pathways only from items {0}, {1} , and {2} .

A comparison of Figures 11–13 makes the same point as the comparison of Figures 9–10, but in a more demanding variation. In each of the Figures 11–13, all the same unordered set of items – {0},{1}, and {2}– is active across F_1. The different spatial patterns across F_1 represent different temporal orderings of these items: {0, 1, 2} ,{1, 2, 0} , and {2, 1, 0} , respectively. In each Figure, a different F_2 node is activated. The active F_2 node is, moreover, one of the nodes that receives pathways only from the item nodes {0} , {1} , and {2} .

Figures 14–16 illustrate how presentation of a list through time can update the sublist chunks in an F_2 field that is capable of simultaneously storing several sublist groupings in STM. In Figure 14, item {0} most strongly activates the {0} nodes of F_2, but also weakly activates other F_2 nodes in an appropriate fashion. The F_2 nodes which receive an item pathway only from {0} have a maximal activity of 0.163. The F_2 nodes which receive two item pathways, including a pathway from {0} , have a maximal activity of 0.07. The F_2 nodes which receive three item pathways, including a pathway from {0} , have a maximal activity of 0.007. These activity weights characterize the degree of "evidence" which the masking field possesses that each grouping is reflected in the input pattern.

In Figure 15, the {0, 1} spatial pattern across F_1 most strongly activates a node within the {0, 1} subfield of F_2, but also weakly activates other nodes of F_2 which receive inputs from {0} . The activity levels are 0.246 and 0.04, respectively. In Figure 16, the {0, 1, 2} spatial pattern across F_1 most strongly activates a node within the {0, 1, 2} subfield of F_2 (with activity 0.184) but also weakly activates the {0} subfield of F_2 (with activity 0.004). Note that the STM

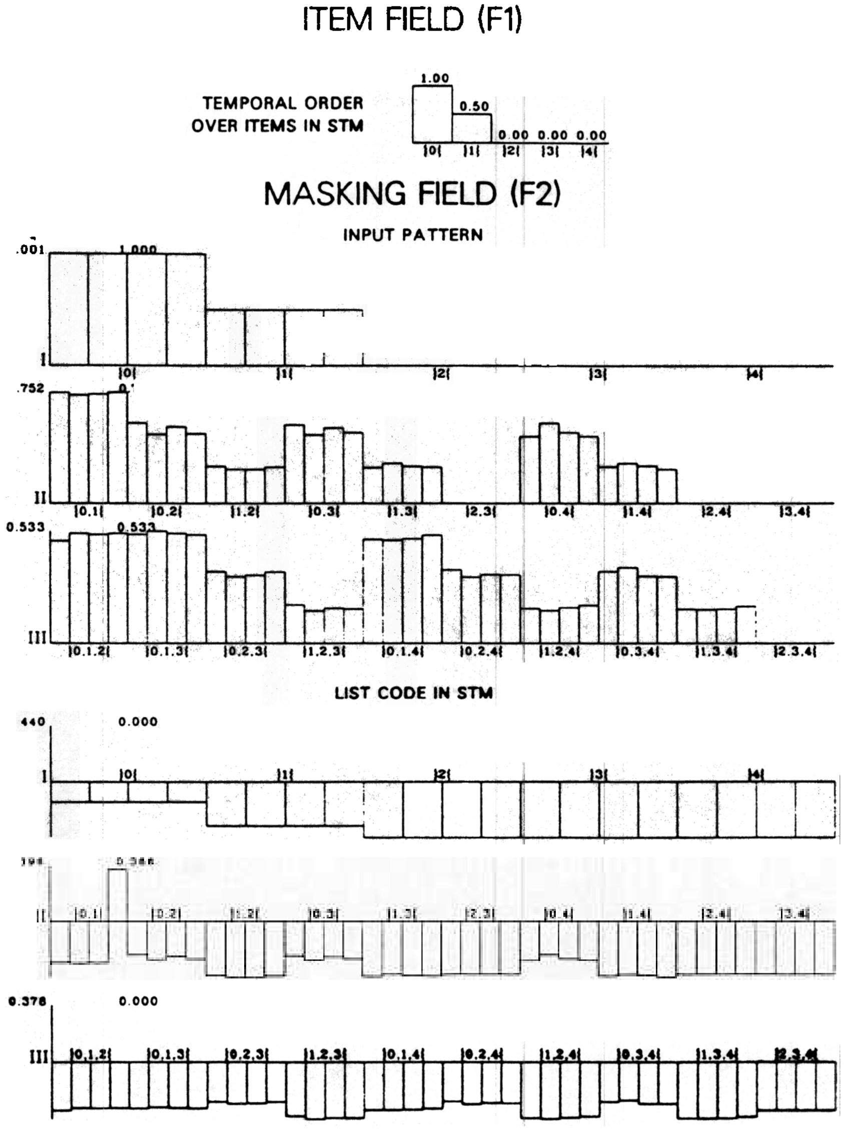


Fig. 9. List coding of an STM primacy gradient across two items: a primacy gradient in STM across two items of F_1 generates an even broader input pattern to F_2. The List Code in STM no longer responds at either the $\{0\}$ cells or the $\{1\}$ cells. Instead, a choice occurs among the set of possible $\{0, 1\}$ cells. Comparison with Figure 6 shows that F_2 can update its internal representation in a context-sensitive way

activity pattern across F_2 becomes more focussed from Figure 14 to 16, as increasing information reduces predictive uncertainty.

These simulations illustrate how simple growth rules can generate a masking field with context-sensitive list parsing properties. The results do not show how such an initial encoding can be refined and corrected by learning and memory search processes. Such simulations would need to develop other mechanisms of the *adaptive resonance theory* (Carpenter and Grossberg 1985; Grossberg 1978a, 1980a, 1982a, 1984a) of which the masking field forms a part.

We now provide a mathematical description of a masking field.

Shunting on-center off-surround networks

The cell populations v_i of a masking field have potentials $x_i(t)$, or STM activities, which obey the membrane equations of neurophysiology; namely,

$$C\frac{\partial V}{\partial t} = (V^+ - V)g^+ + (V^- - V)g^- + (V^p - V)g^p. \tag{2}$$

In Eq. (2), $V(t)$ is a variable voltage; C is a constant capacitance; the constants V^+, V^-, and V^p are excitatory, inhibitory, and passive saturation points, respectively; and the terms g^+, g^-, and g^p are conductances which can vary through time as a function of input signals. Due to the multiplicative relationship between conductances and voltages in Eq. (2), a membrane equation is also said to describe a *shunting* interaction.

In a masking field, the cells are linked together via recurrent, or feedback, on-center off-surround interactions. The properties of a masking field are thus part of the general theory of shunting recurrent on-center off-surround networks (Fig. 17). Grossberg (1981, 1983) reviews the most important functional properties of this class of networks. Masking field properties may be viewed as an evolutionary specialization of these general functional properties.

To emphasize the essential simplicity of the masking field equations, we will build them up in stages. We first rewrite Eq. (2) for the potential $x_i(t)$ in the form

$$\frac{d}{dt}x_i = -Ax_i + (B - x_i)P_i - (x_i + C)Q_i, \tag{3}$$

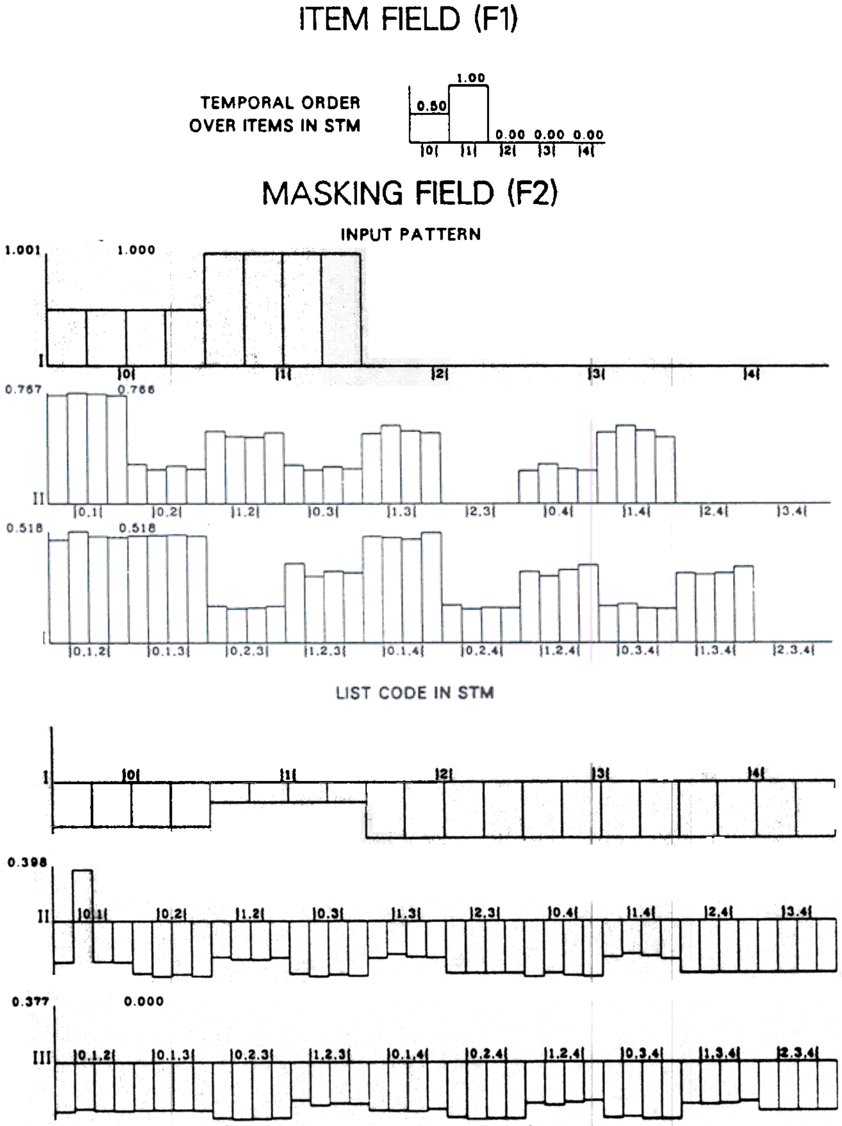


Fig. 10. List coding of an STM recency gradient across two items: a recency gradient in STM occurs across the same two items of F_1, rather than a primacy gradient. Again, the $\{0\}$ cells and the $\{1\}$ cells are suppressed. A *different* choice among the $\{0, 1\}$ cells occurs than in response to the primacy gradient of the preceding figure. Thus F_2 can distinguish different temporal orderings of the same items

where 0 is the passive equilibrium point, $B(>0)$ is the excitatory saturation point, and $-C(\leq 0)$ is the inhibitory saturation point. Term P_i is the total excitatory input and term Q_i is the total inhibitory input to v_i. Potential $x_i(t)$ can vary between B and $-C$ in Eq. (3) as the inputs P_i and Q_i fluctuate through time.

In a masking field, the excitatory input P_i is a sum of two components: the total input from the item field plus a positive feedback signal from v_i to itself (Fig. 17). Thus P_i can be written in the form

$$P_i = \sum_{j\in J} I_j p_{ji} z_{ji} + Df(x_i). \tag{4}$$

In Eq. (4), term I_j is the output from the item node j , p_{ji} is the connection strength of the pathway from v_j in F_1 to v_i in F_2, and z_{ji} is the LTM trace within this pathway. Each LTM trace was set equal to 1 in our simulations, since we did not investigate the effects of learning. The terms z_{ji} will thus be ignored in the subsequent discussion. Term $Df(x_i)$ describes the positive feedback signal from v_i to itself. Such a feedback signal is needed so that v_i can store activities in STM after the inputs I_j terminate.

The inhibitory input Q_i in Eq. (3) is a sum of feedback signals $g(x_m)$ from other populations v_m in the masking field. Thus Q_i can be written in the form

$$Q_i = \sum_{m\in I} g(x_m) E_{mi} \tag{5}$$

Mass action interaction rules

We now refine the notation in Eq. (3–5) to express the fact that the cells in different subfields of a masking field possess different parameters as a result of random growth and activity-dependent self-similar growth. A notation is needed to express the fact that an F_2 population receives F_1 pathways only from a prescribed (unordered) set J of items. Let $x_i^{(J)}$ denote the STM activity of an F_2 population $v_i^{(J)}$ which receives input pathways only from the set J of F_1 items. There may, in principle, be any number of different populations $v_i^{(J)}$ in F_2 corresponding to each fixed set J of F_1 items. Eq. (3) is then replaced by the equation

$$\frac{d}{dt} x_i^{(J)} = -Ax_i^{(J)} + (B - x_i^{(J)}) P_i^{(J)} - (x_i^{(J)} + C) Q_i^{(J)}, \tag{6}$$

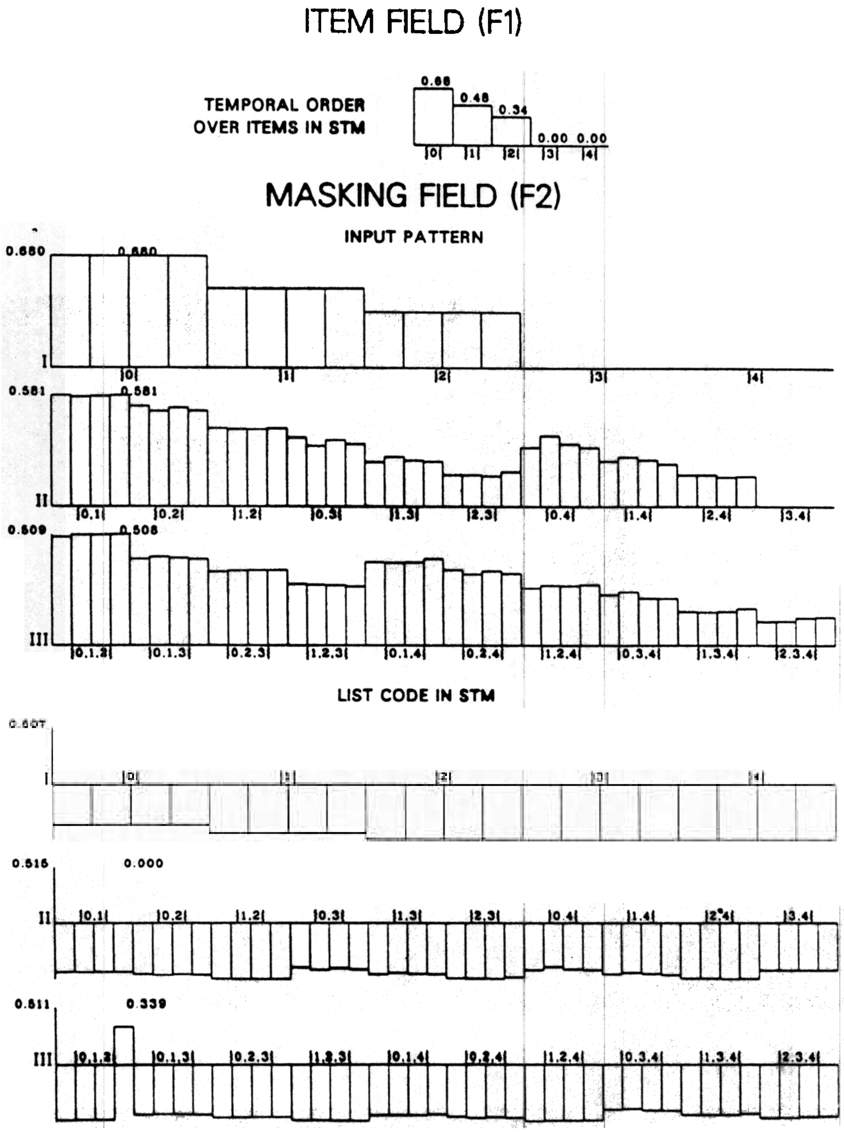


Fig. 11. List coding of an STM primacy gradient across three items: in this figure, a primacy gradient in STM occurs across three items of F_1. The input pattern to F_2 is even broader than before. However, the STM response of F_2 retains its selectivity. Network F_2 suppresses all $\{0\}$, $\{1\}$, $\{2\}$, $\{0, 1\}$, $\{0, 2\ \ ,\ldots\}$ cells and chooses for STM storage a population from among the $\{0, 1, 2\}$ cells

which holds for all unordered sets J of F_1 items that can selectively send pathways to nodes in F_2.

Eq. (4) for the excitatory input P_i is then replaced by

$$P_i^{(J)} = \sum_{j\in J} I_j p_{ji}^{(J)} + D_{|J|} f(x_i^{(J)}). \tag{7}$$

The only notable change in Eq. (7) is in term $D_{|J|}$. Notation $|J|$ denotes the size of set J. Thus $D_{|J|}$ depends upon the size of set J, but not upon the items in set J. This excitatory feedback coefficient is one of the self-similar parameters that is sensitive to the spatial scale of the population $v_i^{(J)}$.

Eq. (5) for the inhibitory input Q_i can be refined in several stages. First we note that $Q_i^{(J)}$ obeys an equation of the form

$$Q_i^{(J)} = \sum_{m,K} g(x_m^{(K)}) E_{KJ}. \tag{8}$$

Eq. (8) can be interpreted as follows. Coefficient E_{KJ} determines the strength of the inhibitory feedback pathway from $v_m^{(K)}$ to $v_i^{(J)}$. This path strength depends only upon the unordered sets K and J of items to which $v_m^{(K)}$ and $v_i^{(J)}$ respond. In particular, E_{KJ} can be written in a form which expresses the randomness of the self-similar growth process:

Mass action interactions

$$E_{KJ} = F_{|J|} G_{|K|} H_{|K\cap J|}. \tag{9}$$

By Eq. (9), E_{KJ} is a product of three factors. Each factor depends only upon the size of an unordered set of items. These unordered sets are set K, set J, and their intersection $K\cap J$. Eq. (9) says that the inhibitory interaction strength from $v_m^{(K)}$ to $v_i^{(J)}$ is the result of a random process. The net strength E_{KJ} is due to a statistically independent interaction between growth factors that depends on the sizes of K, J, and their overlap. By putting together all of these contraints, we find the following

Masking field equations

$$\frac{d}{dt} x_i^{(J)} = -Ax_i^{(J)} + (B - x_i^{(J)}) \left[\sum_{j\in J} I_j p_{ji}^{(J)} + D_{|J|} f(x_i^{(J)})\right] - (x_i^{(J)} + C) \sum_{m,K} g(x_m^{(K)}) F_{|J|} G_{|K|} H_{|K\cap J|}. \tag{10}$$

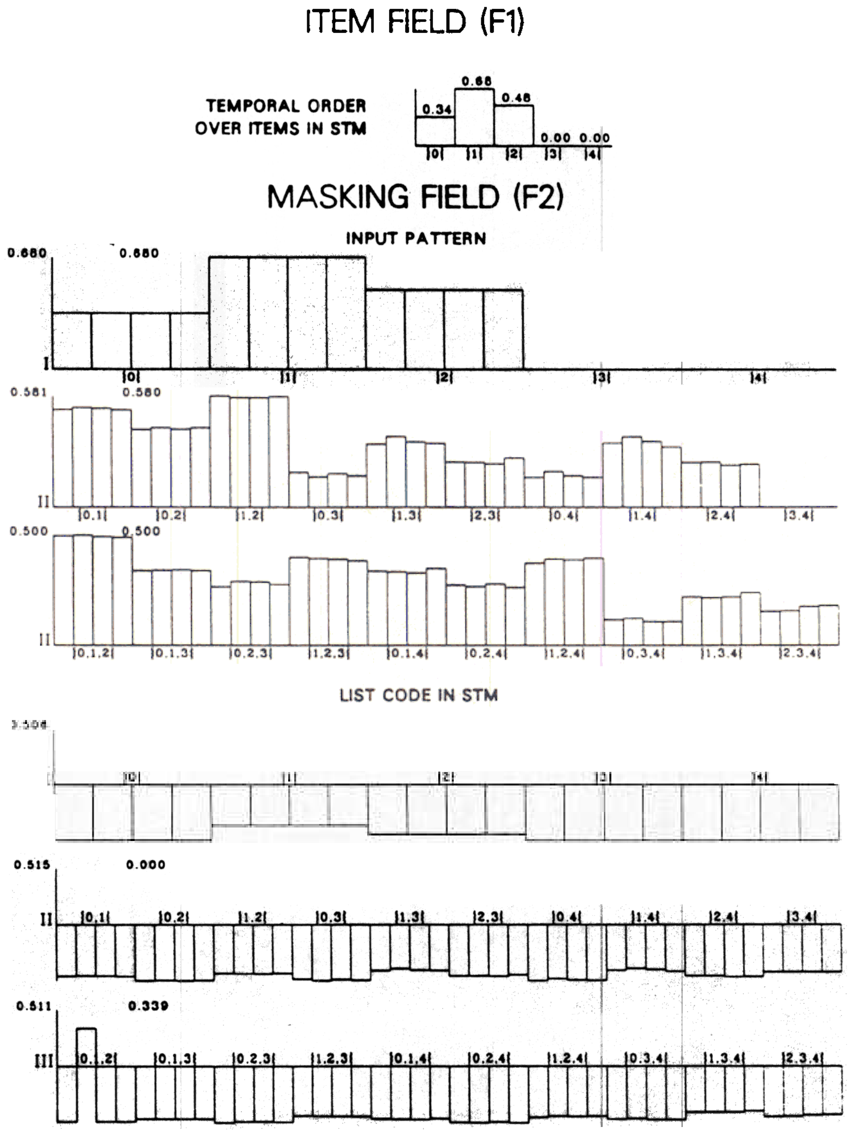


Fig. 12. List codings of different temporal orderings across three items: in this and the next figure, different temporal orderings of the same three items generate selective STM responses among the $\{0, 1, 2\}$ cells. Thus as future items activate an updated STM item code across F_1, the STM list coding within F_2 is also updated in a context-sensitive way

All of the "intelligence" of a masking field is embodied in the parallel interactions defined by such a network of equations. It remains to define how the coefficients $D_{|J|}$, $F_{|J|}$, $G_{|K|}$, and $H_{|K\cap J|}$ depend upon the unordered sets K and J; how the positive and negative feedback functions $f(w)$ and $g(w)$ depend upon their activities w; how the path strengths $p_{ji}^{(J)}$ from F_1 to F_2 express a random growth rule; and how numerical parameters were chosen.

Self-similar growth within list nodes

The coefficient $D_{|J|}$ determines how the positive feedback from a node to itself varies with the node's self-similar scale. We assume that $D_{|J|}$ increases with scale, thereby enabling nodes corresponding to longer sublists to gain a competitive advantage in STM, other things being equal. The simplest choice is made in our simulations, namely

$$D_{|J|} = D|J|, \tag{11}$$

where D is a positive constant. This rule is consistent with the possibility that, as an F_2 cell (population) grows in response to high levels of F_1 input, it also produces more excitatory synaptic sites for its own axon collaterals.

Conservation of synaptic sites

The dependence of the intermodal connection strengths $p_{ji}^{(J)}$, $F_{|J|}$, $G_{|K|}$, and $H_{|K\cap J|}$ on the sets K and J will now be described. The *total* connection strength to each population $v_i^{(J)}$ from all cells in F_1 and the *total* inhibitory connection strength to each population $v_i^{(J)}$ from all cells in F_2 are both chosen to be independent of K and J. This property is compatible with the interpretation that the size of each cell (population) is scaled to the total strength of its input pathways. If more pathways input to such a cell, then each input's effect is diluted more due to the larger size of the cell. This constraint may, in principle, be achieved by either of the mechanisms depicted in Figures 3 and 4. We call the net effect of matching cell (population) volume to its total number of afferents *conservation of synaptic sites.*

Conservation of synaptic sites enables the network to overcome the following possible problem. Due to the random-

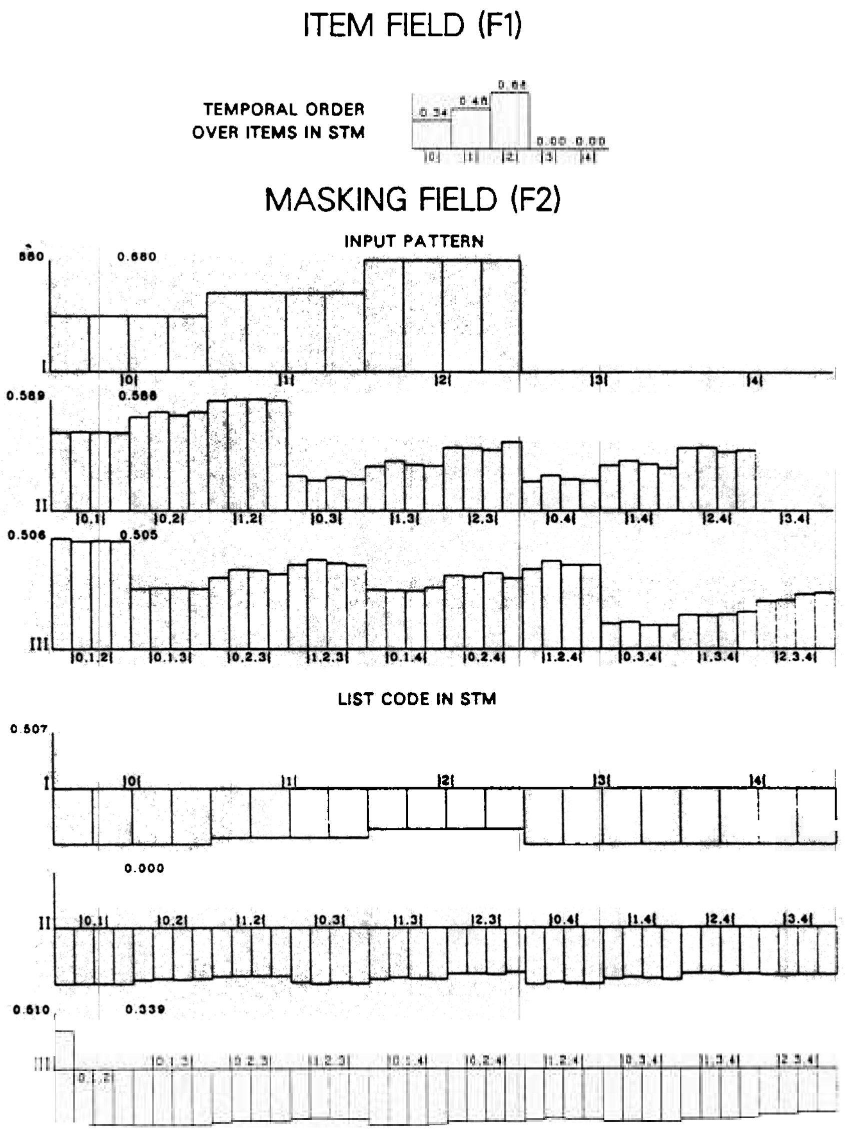


Fig. 13. See legend to Figure 12

ness of the growth rules, there may exist different numbers of cells in each of F_2's masking subfields. For example, 10^3 F_2 cells may receive inputs only from the F_1 node $\{0\}$, 10^4 cells may receive inputs only from the F_1 node $\{1\}$, 10^6 cells may receive inputs only from the F_1 nodes $\{0\}$ and $\{1\}$, and so on. As these F_2 cells compete for STM activity, the competitive balance could be seriously biased by accidents of random growth. Some mechanism is needed to compensate for the possible uncontrolled proliferation of random connections. Conservation of synaptic sites is one effective mechanism. The present results suggest a new functional role for such a growth rule. Thus we impose the following constraints:

Synaptic conservation rule

Let

$$\sum_{j \in J} p_{ji}^{(J)} = \text{constant} = 1 \tag{12}$$

and

$$\sum_{m,K} F_{|J|} G_{|K|} H_{|K \cap J|} = \text{constant}. \tag{13}$$

In particular, we choose

$$F_{|J|} = \frac{F}{\sum_{m,K} G_{|K|} H_{|K \cap J|}}$$

to satisfy Eq. (12), where F is a positive constant

Random growth from item nodes to list nodes

The connections $p_{ji}^{(J)}$ from F_1 to F_2 are chosen to satisfy the conservation law [Eq. (12)] as well as a random growth law. We therefore impose the following constraint:

Random normalized growth rule

Let

$$p_{ji}^{(J)} = \frac{1}{|J|}(1 - p_{|J|}) + r_{ji}^{(J)} p_{|J|}. \tag{15}$$

The *fluctuation coefficient* $p_{|J|}$ in Eq. (15) determines how random the growth is from F_1 to F_2. If $p_{|J|} = 0$, then growth is deterministic (but spatially distributed) because $p_{ji}^{(J)} = \frac{1}{|J|}$.

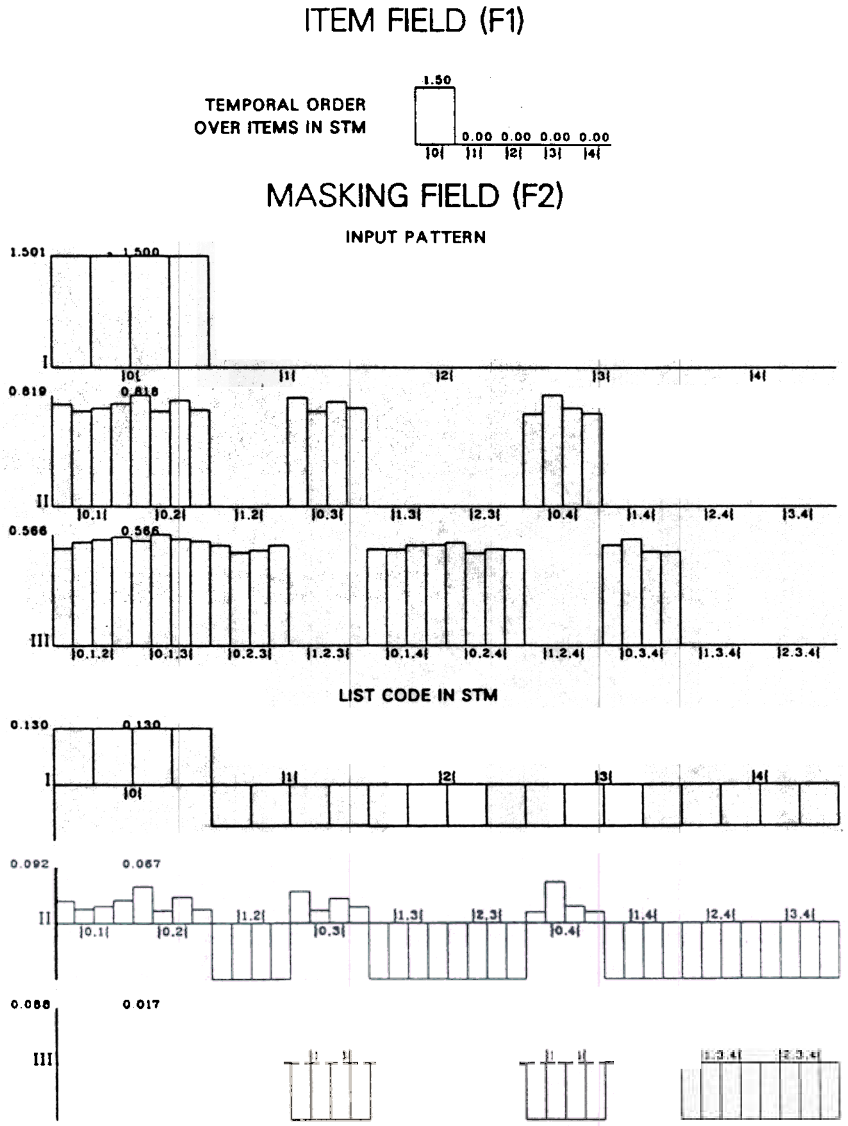


Fig. 14. Distributed sublist encodings of one, two, and three items: this and the subsequent two figures illustrate the STM responses of F_2 when numerical parameters are chosen outside of the STM choice range. Note that distributed STM reactions occur in every case, and that these STM reactions favor the populations that were chosen in the STM choice simulations

In this limiting case, all connection strengths from item nodes in F_1 to a *fixed* list node in F_2 are equal, and vary inversely with the number $|J|$ of item nodes that contact the list node. If $0 < p_{|J|} < 1$, then the coefficients $r_{ji}^{(J)}$ in Eq. (15) influence the connection strengths $p_{ji}^{(J)}$. The numbers $\{r_{ji}^{(J)} : j \in J\}$ are chosen pseudo-randomly: they are uniformly distributed between 0 and 1 such that

$$\sum_{j \in J} r_{ji}^{(J)} = 1 \tag{16}$$

(see Appendix). Eq. (15) and (16) together imply the conservation rule Eq. (12).

It remains to say how the fluctuation coefficients $p_{|J|}$ depend upon the set size $|J|$. We choose these coefficients to keep the statistical variability of the connection strengths independent of $|J|$. In other words, we choose $p_{|J|}$ so that the standard deviation of $\{r_{ji}^{(J)} : j \in J\}$ divided by the mean of $\{r_{ji}^{(J)} : j \in J\}$ is independent of $|J|$ (see Appendix).

Self-similar competitive growth between list nodes

Coefficient $F_{|J|}$ in Eq. (10) describes the total number of inhibitory synaptic sites within a population $v_i^{(J)}$. By Eq. (14), this quantity is chosen to keep the number of synaptic sites constant across all the cells. Small random variations could also be allowed, but we have absorbed all of the effects of randomness into the coefficients $p_{ji}^{(J)}$ in Eq. (15) for simplicity.

Coefficient $G_{|K|}$ in Eq. (10) measures the total number of inhibitory connections, or axons, emitted by each population $v_m^{(K)}$ to all other F_2 populations. Due to self-similar growth, $G_{|K|}$ increases with $|K|$. In our simulations, we make the simplest choice.

Self-similar axon generation

Let

$$G_{|K|} = |K|. \tag{17}$$

In particular, $G_{|K|} = 0$ if $|K| = 0$.

Coefficient $H_{|K \cap J|}$ in Eq. (10) describes how well growing axons from a population $v_m^{(K)}$ can compete for synaptic sites at a population $v_i^{(J)}$. In particular, coefficient $G_{|K|}$ describes the number of emitted axons. Coefficient $H_{|K \cap J|}$

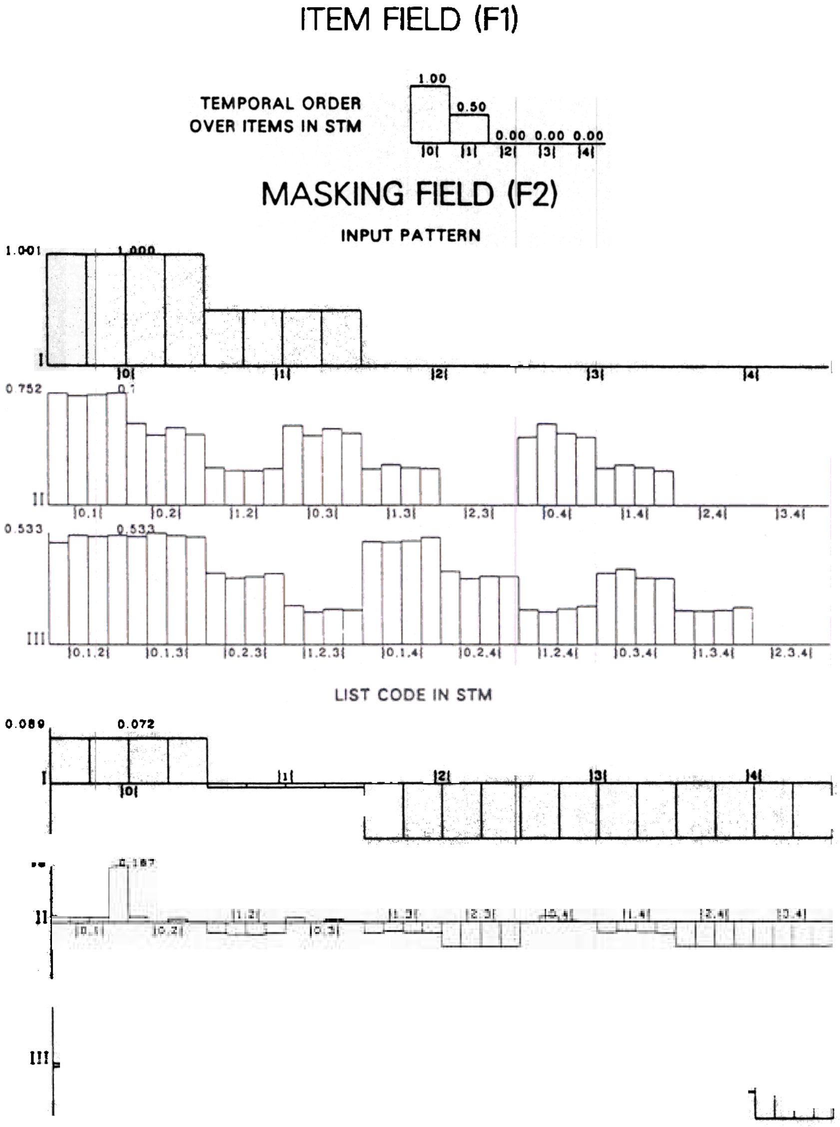


Fig. 15. See legend to Figure 14

measures the fraction of these axons that can reach $v_i^{(J)}$ and compete for synaptic space there. Due to self-similar growth, $H_{|K\cap J|}$ increases with $|K\cap J|$. Consequently, if either set K or J increases, then $H_{|K\cap J|}$ also increases, other things being equal. Given fixed sizes of K and J, then $H_{|K\cap J|}$ increases as the overlap, or intersection, of the sets increases. This last property reflects the fact that list nodes *become* list nodes due to random growth of connections from item nodes. Two list nodes therefore tend to be closer in F_2 if they receive more input pathways from the same item nodes in F_1. If a pair of list nodes in F_2 is closer, then their axons can more easily contact each other, other things being equal. In our simulations, we choose $H_{|K\cap J|}$ as follows. Let

$$H_{|K\cap J|} = 1 + |K\cap J| \tag{18}$$

By Eq. (18), $H_{|K\cap J|}$ increases linearly with $|K\cap J|$. We also assume, however, that $H_{|K\cap J|}$ is always positive. When $H_{|K\cap J|}$ multiplies $G_{|K|}$ in Eq. (10), this implies that every population $v_m^{(K)}$ can send weak long-range inhibitory pathways across the whole of F_2, but that these pathways tend to arborize with greater density at populations $v_i^{(J)}$ which receive inputs from the same F_1 nodes. In all, Eq. (14), (17), and (18) imply that

$$F_{|J|} = \frac{F}{\Sigma_{m,K}\ |K|(1 + |K\cap J|)} \tag{19}$$

Contrast enhancement by sigmoid signal functions

The positive and negative feedback signals $f(x_i^{(J)})$ and $g(x_m^{(K)})$ in Eq. (10) enable the network to contrast enhance its input patterns before storing them in STM. The mathematical theory of how to design shunting on-center off-surround feedback networks with this property was introduced in Grossberg (1973), further developed in Ellias and Grossberg (1975) and Grossberg and Levine (1975), and led to a rather general mathematical theory in Grossberg (1978c, 1978d, 1980b) and Cohen and Grossberg (1983). Salient properties of these networks are reviewed in Grossberg (1983).

Based on this analysis, we choose both $f(w)$ and $g(w)$ to be sigmoid, or S-shaped, functions of the activity level w. In particular, we let

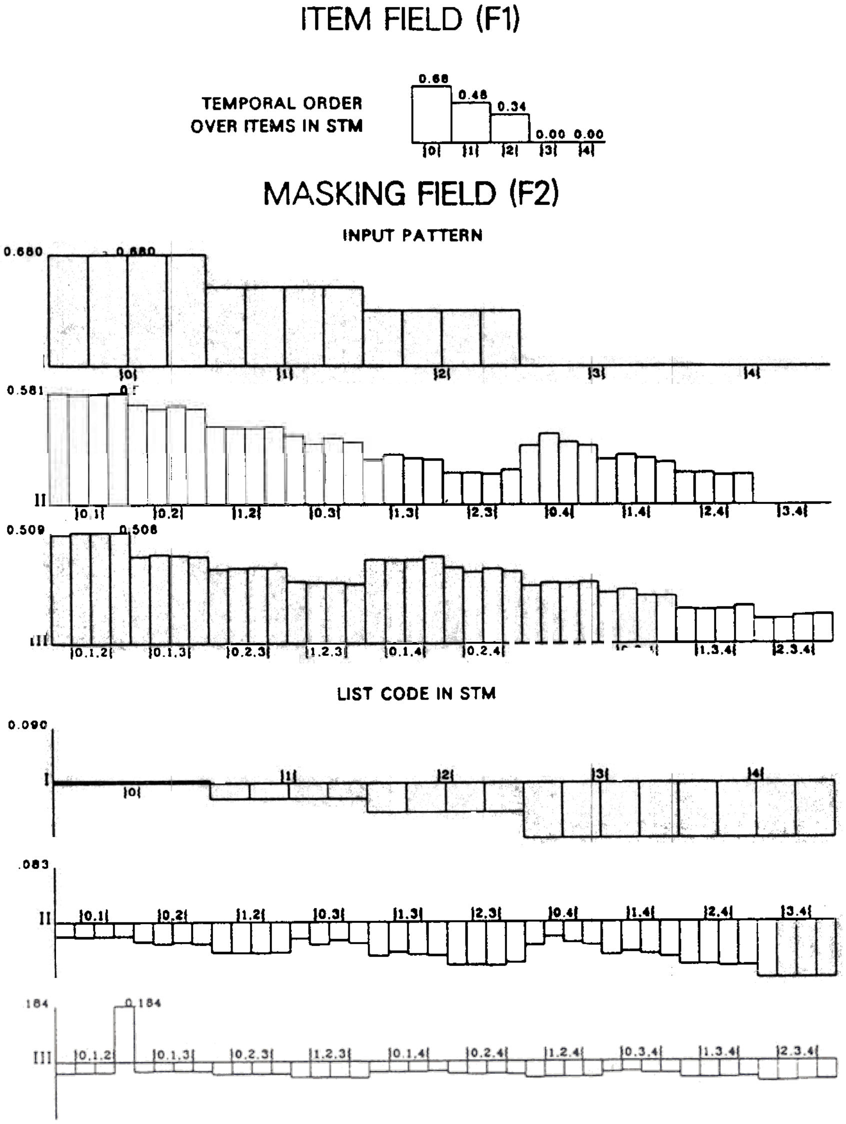


Fig. 16. See legend to Figure 14

$$f(w) = \frac{([w]^+)^2}{f_0 + ([w]^+)^2} \tag{20}$$

and

$$g(w) = \frac{([w]^+)^2}{g_0 + ([w]^+)^2} \tag{21}$$

The notation $[w]^+$ in Eq. (20) and (21) stands for $\max(w, 0)$. Thus $f(w)$ and $g(w)$ do not generate feedback signals if w is smaller than the signal threshold zero. As w increases above zero, both $f(w)$ and $g(w)$ grow quadratically with w until they begin to saturate at their maximum value 1. Sigmoid signal functions have been described in sensory neural processing regions (Freeman 1979, 1981).

Concluding remarks: Grouping and recognition without algorithms or search

This article shows how to design a masking field capable of encoding temporally occurring lists of items in a context-sensitive fashion. The STM activations within such a masking field are sensitive to different temporal orderings of the same list items as well as to the same temporal orderings of different item sublists. Our results have focused upon two different levels of language processing, both of which are context-sensitive, but in different ways. The spatial patterning of activity across an item field defines computational units which blend item information and temporal order information in a context-sensitive, but non-unitized, code. Such item representations can be recalled even before unitization can take place. The STM activations of a masking field perceptually group different portions of a spatial pattern across an item field into unitized sublist representations.

We have illustrated how such properties of perceptual grouping, unitization, and recognition can be analysed as emergent properties of interactions within a large nonlinear network of recurrently-interacting neurons. These networks do not incorporate any serial programs or cognitive rule structures. Moreover, the appropriate sublist representations are directly accessed without any prior search.

We have furthermore shown how neural networks with these emergent properties can arise from simple developmental programs governing the manner in which the neurons

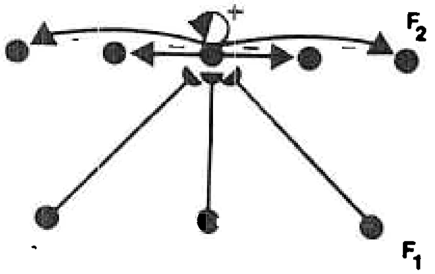


Fig. 17. Connections grow randomly from F_1 to F_2 along positionally defined gradients. Cells within F_2 interact via a shunting on-center off-surround feedback network

grow and interconnect. In our present simulations, we have made one choice of these developmental rules. This choice suggests that F_2 cells obey different synaptic rules for recognizing F_2 excitatory signals, F_2 inhibitory signals, and F_1 excitatory signals. However, the principles of network design – such as the principle of activitydependent self-similar growth – are much more general than this choice. This combination of general principles and rigourous examples provides a firm foundation for testing the theory experimentally.

In the present work, we have shown how a prewired developmental program can generate a network with the desired functional properties. These results place harsher demands upon the network than are, we believe, required in vivo. In the full Adaptive Resonance Theory of cognitive self-organization to which our results contribute, it is not necessary for the initial list encodings to be accurate. One only needs a sufficiently good processing substrate for top-down learned template-matching signals (from A_4 to A_3 in Fig. 1) to drive an automatic memory search that can provide the occasion for learning a better encoding (Carpenter and Grossberg 1985; Grossberg 1980a, 1984b). With a quantitative understanding of how prior development can set the stage for later matching, memory search, and code learning events, we can now frontally attack the full problem of list code self-organization.

Appendix

This section describes some technical details of our simulations. First we list the input values that are used in the simulations. The inputs are listed by Figure number. Only positive inputs are listed. All other inputs equal zero.

Inputs:

Fig. 6 (I_0 = 1.5); Fig. 6 (I_1 = 1.5); Fig. 8 (I_2 = 1.5); Fig. 9 (I_0 = 1.0, I_1 = 0.5); Fig. 10 (I_0 = 0.5, I_1 = 1.0); Fig. 11 (I_0 = 0.68, I_1 = 0.48, I_2 = 0.34); Fig. 12 (I_0 = 0.34, I_1 = 0.68, I_2 = 0.48); Fig. 13 (I_0 = 0.34, I_1 = 0.48, I_2 = 0.68); Figures 14–16 are the same as Figures 6, 9, and 11, respectively.

Connections from F_1 to F_2:

To produce a pseudorandom sequence of numbers $\{r_{ji}^{(J)} : j \in J\}$ distributed uniformly over the simplex

$$S_n = \{(y_1, y_2, \ldots, y_{n+1}) : y_i \geq 0, \sum_{j=1}^{n+1} y_j = 1\}, \tag{A1}$$

we proceed as follows. By a standard algorithm (Knuth 1981), we obtain a vector of numbers $w = (w_1, w_2, \ldots, w_n)$ uniformly distributed over the n-cube $I_n = \times_{j=1}^{n} [0, 1]$. Rearrange the numbers in w in order of increasing size to produce a new vector $w' = (w'_1, w'_2, \ldots, w'_n)$ such that $w'_1 \leq w'_2 \leq \ldots \leq w'_n$. The map $w \rightarrow w'$ from I_n into itself is determined by a permutation σ of the indices $\{1, 2, \ldots, n\}$ such that $w'_i = w_{\sigma(i)}$. Each permutation σ can transform a different subset of I_n into vectors with increasing entries. Thus I_n can be decomposed into sets D_σ such that a single permutation σ can map *all* $w \in D_\sigma$ into $w' \in I_n$. Hence the map $w \rightarrow w'$ transforms uniformly distributed vectors in I_n onto uniformly distributed vectors in I_n with elements in increasing order.

We next map vectors w' in I_n with elements in increasing order onto vectors y in S_{n+1} via the one-to-one linear transformation $y_1 = w'_1, y_2 = w'_2 - w'_1, \ldots, y_n = w'_n - w'_{n-1}$, and $y_{n+1} = 1 - w_n$. Since this linear transformation maps equal volumes onto equal surface areas, the vectors y are uniformly distributed on the simplex S_{n+1}.

The coefficient of variation of $\{p_{ji}^{(J)} : j \in J\}$ is made independent of $|J|$ (> 1) as follows. By the above construction, the marginal distribution $r_{ji}^{(J)}$ in Eq. (14) is distributed with density function $(|J| - 1)(1 - x)^{|J|-2}$. The mean of this distribution is $\frac{1}{|J|}$, and its standard deviation is $\frac{1}{|J|}\sqrt{\frac{|J|-1}{|J|+1}}$. Thus the mean of $p_{ji}^{(J)}$ is also $\frac{1}{|J|}$ and its standard deviation is

$$p_{|J|} \frac{1}{|J|} \sqrt{\frac{|J|-1}{|J|+1}} \; . \tag{A2}$$

The coefficient of variation of $p_{ji}^{(J)}$ is its standard deviation divided by its mean, which we set equal to a constant p independent of $|J|$. Thus we chose

$$p_{|J|} = p \sqrt{\frac{|J|+1}{|J|-1}} \; . \tag{A3}$$

In our simulations, $p = \frac{1}{10\sqrt{3}}$.

Interaction constants:

The following parameter choices were made: A = 1, B = 1, D = 4, f_0 = 1, and g_0 = 0.16. In the total choice runs (Figs. 6–13), we let C = 1 and F = 1088. In the partial choice runs (Figs. 14–16), we let C = 0.125 and F = 8704. Note that CF = 1088 in both cases. The behavior of a masking field has also been characterized over a wide range of other parameter choices.

Acknowledgements: We wish to thank Cynthia Suchta for her valuable assistance in the preparation of the manuscript and illustrations. M. Cohen was supported in part by the National Science Foundation (NSF IST-8417756) and the Office of Naval Research (ONR N00014-83-K0337). S. Grossberg was supported in part by the Air Force Office of Scientific Research (AFOSR 82-0148) and the Office of Naval Research (ONR N00014-83-K0337).

References

Basar E, Flohr H, Haken H, Mandell AJ (eds) (1983) Synergetics of the brain. Springer, Berlin Heidelberg New York

Carpenter GA, Grossberg S (1985) Neural dynamics of category learning and recognition: attention, memory consolidation, and amnesia. In: Davis J, Newburgh R, Wegman E (eds) Brain structure, learning, and memory. Washington DC, AAAS Symposium Series

Cermak LS, Craik FIM (eds) (1979) Levels of processing in human memory. Erlbaum, Hillsdale, NJ

Cohen MA, Grossberg S (1983) Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. IEEE Transactions on Systems, Man, and Cybernetics SMC-13:815–826

Cooper WE (1979) Speech perception and production: studies in selective adaptation. Ablex, Norwood, NJ

Darwin CJ (1976) The perception of speech. In: Carterette EC, Friedman MP (eds) Handbook of perception, vol VII: language and speech. Academic Press, New York

Ellias SA, Grossberg S (1975) Pattern formation, contrast control, and oscillations in the short term memory of shunting on-center off-surround networks. Biol Cybern 20:69–98

Freeman WJ (1979) Nonlinear dynamics of paleocortex manifested in the olfactory EEG. Biol Cybern 35:21–37

Freeman WJ (1981) A neural mechanism for generalization over equivalent stimuli in the olfactory system. In: Grossberg S (ed) Mathematical psychology and psychophysiology. American Mathematical Society, Providence, RI

Grossberg S (1969) On the serial learning of lists. Math Biosci 4: 201–253

Grossberg S (1973) Contour enhancement, short-term memory, and constancies in reverberating neural networks. Stud Appl Math 52:217–257

Grossberg S (1978a) A theory of human memory: self-organization and performance of sensory-motor codes, maps, and plans. In: Rosen R, Snell F (eds) Progress in theoretical biology, vol 5. Academic Press, New York

Grossberg S (1978b) Behavioral contrast in short-term memory: serial binary memory models or parallel continuous memory models. J Math Psychol 17:199–219

Grossberg S (1978c) Competition, decision, and consensus. J Math Anal Appl 66:470–493

Grossberg S (1978d) Decisions, patterns, and oscillations in the dynamics of competitive systems with applications to Volterra-Lotka systems. J Theoret Biol 73:101–130

Grossberg S (1980a) How does a brain build a cognitive code? Psychol Rev 87:1–51

Grossberg S (1980b) Biological competition: decision rules, pattern formation, and oscillations. Proc Nat Acad Sci 77:2338–2342

Grossberg S (1981) Adaptive resonance in development, perception, and cognition. In: Grossberg S (ed) Mathematical psychology and psychophysiology. American Mathematical Society, Providence, RI

Grossberg S (1982a) Studies of mind and brain: neural principles of learning, perception, development, cognition, and motor control. Reidel Press, Boston

Grossberg S (1982b) Associative and competitive principles of learning and development: the temporal unfolding and stability of STM and LTM patterns. In: Amari SI, Arbib M (eds) Competition and cooperation in neural networks. Springer, Berlin Heidelberg New York

Grossberg S (1983) The quantized geometry of visual space: the coherent computation of depth, form, and lightness. Behav Brain Sci 6:625–692

Grossberg S (1984a) Unitization, automaticity, temporal order, and word recognition. Cogn Brain Theor 7:263–283

Grossberg S (1984b) Some psychophysiological and pharmacological correlates of a developmental, cognitive, and motivational theory. In: Karrer R, Cohen J, Tueting P (eds) Brain and information: event related potentials. New York Academy of Sciences, New York

Grossberg S (1985) The adaptive self-organization of serial order in behavior: speech, language, and motor control. In: Schwab EC, Nusbaum HC (eds) Pattern recognition by humans and machines, Vol 1. Academic Press, New York

Grossberg S, Levine DS (1975) Some developmental and attentional biases in the contrast enhancement and short term memory of recurrent neural networks. J Theoret Biol 53:341–380

Grossberg S, Pepe J (1971) Spiking threshold and overarousal effects in serial learning. J Statist Phys 3:95–125

Grossberg S, Stone GO (1986a) Neural dynamics of word recognition and recall: attentional priming, learning, and resonance. Psychol Rev (in press)

Grossberg S, Stone GO (1986b) Neural dynamics of attention switching and temporal order information in short term memory. (Submitted for publication)

Halle M, Stevens KN (1962) Speech recognition: a model and a program for research. IRE Trans Inf Theor IT-8:155–159

Knuth DE (1981) Seminumerical algorithms: the art of computer programming, vol 2. Addison-Wesley, Reading, MA

Liberman AM, Cooper FS, Shankeweiler DS, Studdert-Kennedy M (1967) Perception of the speech code. Psychol Rev 74:431–461

Liberman AM, Studdert-Kennedy M (1978) Phonetic perception. In: Held R, Leibowitz H, Teuber H-L (eds) Handbook of sensory physiology, vol VIII. Springer, Berlin Heidelberg New York

Lindblom B, MacNeilage P, Studdert-Kennedy M (1983) Self-organizing processes and the explanation of phonological universals. In: Butterworth, Comrie, Dahl (eds) Explanations of linguistic universals. Mouton, The Hague

Mann VA, Repp BH (1981) Influence of preceding fricative on stop consonant perception. J Acoust Soc Am 69:548–558

Patterson PH, Purves D (eds) (1982) Readings in developmental neurobiology. Cold Spring Harbor Lab, Cold Spring Harbor, NY

Piaget J (1981) The origins of intelligence in children. Norton, New York

Repp BH (1981) On levels of description in speech research. J Acoust Soc Am 69:1462–1464

Repp BH (1982) Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. Psychol Bull 92:81–110

Repp BH, Mann VA (1981) Perceptual assessment of fricative-stop coarticulation. J Acoust Soc Am 69:1154–1163

Samuel AG, Santen JPH van, Johnston JC (1982) Length effects in word perception: we is better than I but worse than you or them. J Exp Psychol [Hum Percept] 8:91–105

Samuel AG, Santen JPH van, Johnston JC (1983) Reply to Matthei: we really is worse than you or them, and so are ma and pa. J Exp Psychol [Hum Percept] 9:321–322

Sperling G (1960) The information available in brief visual presentations. Psychol Monogr 74:1–29

Stevens KN (1972) Segments, features, and analysis by synthesis. In: Cavanaugh JV, Mattingly IG (eds) Language by eye and by ear. MIT Press, Cambridge, MA

Stevens KN, Halle M (1964) Remarks on analysis by synthesis and distinctive features. In: Wathen-Dunn W (ed) Proceedings of the AFCRL symposium on models for the perception of speech and visual form. MIT Press, Cambridge, MA

Studdert-Kennedy M (1980) Speech perception. Lang Speech 23: 45–66

Studdert-Kennedy M, Liberman AM, Harris KS, Cooper FS (1970) Motor theory of speech perception: a reply to Lane's critical review. Psychol Rev 77:234–249

Wheeler DD (1970) Processes in word recognition. Cogn Psychol 1:59–85

Young RK (1968) Serial learning. In: Dixon TR, Horton DL (eds) Verbal behavior and general behavior theory. Prentice-Hall, Englewood Cliffs, NJ