HOW HALLUCINATIONS MAY ARISE FROM BRAIN MECHANISMS OF LEARNING, ATTENTION, AND VOLITION

Stephen Grossberg*

Department of Cognitive and Neural Systems
and
Center for Adaptive Systems
Boston University
677 Beacon Street
Boston, MA 02215

July, 1999
Revised: October, 1999
Technical Report CAS/CNS TR-99-020

Invited article for the
Journal of the International Neuropsychological Society, in press

ABSTRACT

This article suggests how brain mechanisms of learning, attention, and volition may give rise to hallucinations during schizophrenia and other mental disorders. The article suggests that normal learning and memory are stabilized through the use of learned top-down expectations. These expectations learn prototypes that are capable of focusing attention upon the combinations of features that comprise conscious perceptual experiences. When top-down expectations are active in a priming situation, they can modulate or sensitize their target cells to respond more effectively to matched bottom-up information. They cannot, however, fully activate these target cells. These matching properties are shown to be essential towards stabilizing the memory of learned representations. The modulatory property of top-down expectations is achieved through a balance between top-down excitation and inhibition. The learned prototype is the excitatory on-center in this top-down network. Phasic volitional signals can shift the balance between excitation and inhibition to favor net excitatory activation. Such a volitionally-mediated shift enables top-down expectations, in the absence of supportive bottom-up inputs, to cause conscious experiences of imagery and inner speech, and thereby to enable fantasy and planning activities to occur. If these volitional signals become tonically hyperactive during a mental disorder, the top-down expectations can give rise to conscious experiences in the absence of bottom-up inputs and volition. These events are compared with data about hallucinations. The article predicts where these top-down expectations and volitional signals may act in the laminar circuits of visual cortex, and by extension in other sensory and cognitive neocortical areas, and how the level of abstractness of learned prototypes may covary with the abstractness of hallucinatory content. A similar breakdown of volition may lead to declusions of control in the motor system.

Key Words: hallucinations, learned expectations, attention, learning, adaptive resonance theory.

How Hallucinations May Arise from Basic Brain Mechanisms

Hallucinations occur when we involuntarily experience a percept in the absence of an external stimulus. They are a familiar positive symptom of schizophrenia (Sartorius et al., 1974), among other mental disorders. What is the source of these disabling experiences? The present article suggests that they arise from neocortical mechanisms that play a key role in how we can learn about a changing world throughout life.

These mechanisms include learned top-down sensory expectations that are used to modulate, prime, and match incoming bottom-up information (Grossberg, 1980, 1999b). Such top-down expectations help to focus attention upon expected clusters of sensory features. These top-down expectations learn prototypes that incorporate the information that is essential for perceiving or recognizing information at different levels of cortical processing. When these expectations are activated in the absence of bottom-up information, they are usually prevented from causing hallucinations through a balance between top-down excitation and inhibition (see Figure 1). Such top-down expectations can modulate, sensitize, or prime the processing of bottom-up information, but cannot by themselves cause suprathreshold activation of their target cells. How and why such a modulatory balance is accomplished is discussed in greater detail below. In addition, recent modeling work on the laminar organization of visual cortex (Grossberg, 1999a; Grossberg and Raizada, 1999) has suggested some of the possible circuits by which these top-down expectations control the focusing of visual attention, and how these attentive mechanisms interact with the bottom-up flow of visual information (see Figure 2).

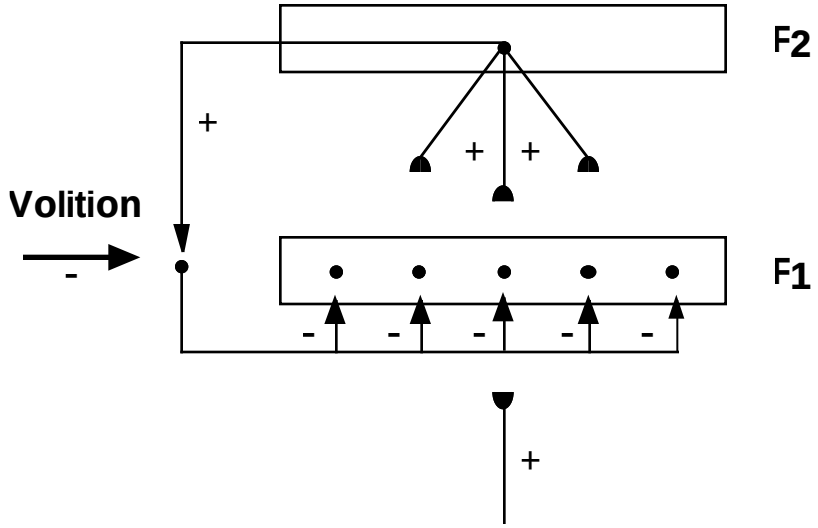

Figure 1. The ART Matching Rule is achieved by a top-down on-center off-surround network. The excitatory on-center (plus signs) encodes a learned prototype in its adaptive weights, or long-term memory traces (hemidisks). This prototype learns from bottom-up inputs, which can fully activate targets cells when top-down signals are off. The inhibitory off-surround (minus signs) is balanced against the on-center so that top-down signals, by themselves, cannot fully activate their target cells. When both bottom-up and top-down signals are active, only the cells in the top-down on-center that are also receiving bottom-up inputs can fire. Other cell activities are inhibited. A volitional signal can disrupt the top-down excitatory-inhibitory balance to enable top-down signals, by themselves, to cause suprathreshold activation.

This article suggests that, under normal behavioral conditions, a volitional signal can be phasically turned on that can alter this balance to favor top-down excitation (Figure 1). Through this means, top-down expectations can create conscious experiences in the absence of bottom-up information. When this occurs, we can experience visual imagery, inner speech, or other internal conscious states that are very useful towards understanding our past experiences and planning new

experiences. Thus, the ability of volitional signals to convert modulatory top-down priming signals into suprathreshold activations provides a great evolutionary advantage to those who possess it.

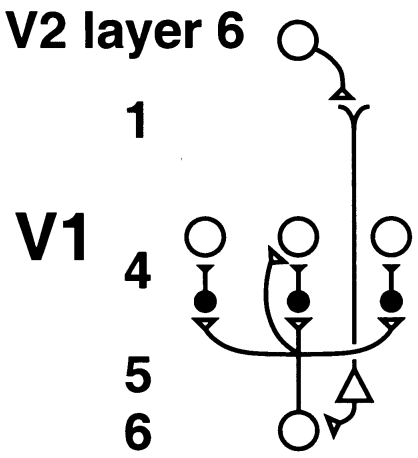

Figure 2. Model circuit for how a top-down on-center off-surround network is realized within the laminar circuits of visual cortical areas V1 and V2. Similar circuits are proposed to occur in other sensory and cognitive cortical areas. Open circles and triangles denote excitatory cells and pathways, respectively; closed black circles and triangles denote inhibitory cells and pathways, respectively. A *folded feedback* circuit carries top-down attentional signals from layer 6 of V2 to layer 4 of V1 via an on-center off-surround pathway from layer 6 to 4 of V1.Corticocortical feedback axons from layer 6 in V2 tend to terminate in layer 1 of V1 (Salin and Bullier, 1995, p. 110) where they can, for example, excite apical dendrites of layer 5 pyramidal cells whose axons send collaterals into layer 6. From layer 6, the feedback is then "folded" back into the feedforward flow of information from layer 6 to 4 of V1 via an on-center off-surround pathway (Bullier et al., 1996). See Grossberg (1999a) and Grossberg and Raizada (1999) for a model of how this circuit is embedded within the bottom-up, horizontal, and top-down (both intracortical and intercortical) interactions within visual cortex.

During a mental disorder like schizophrenia, it is proposed that the phasic volitional signal may become tonically hyperactive. As a result, top-down sensory expectations can generate conscious experiences that are not under the volitional control of the individual who is experiencing them. The net effect is a hallucination. Since the top-down expectations learn prototypes that incorporate the critical features that are used to bind sensory features into conscious experiences, these hallucinations, just like the imagery and inner speech that are generated under normal conditions, are sufficient to generate conscious experiences with vivid perceptual content.

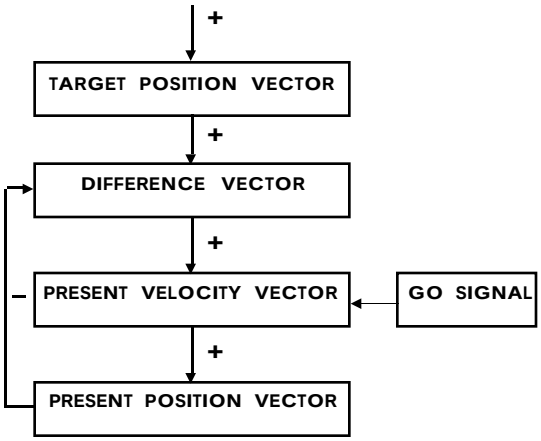

Figure 3. When an arm movement is being planned, a Target Position Vector (TPV) specifies where the arm should move and a Present Position Vector (PPV) controls an outflow signal that specifies the arm's present position. The PPV is subtracted from the TPV to define a Difference Vector (DV) which determines the direction and distance that the arm must move to bring the arm to the new location specified by the TPV. A new TPV can be loaded without moving the arm. Activation of a volitional GO signal opens a multiplicative gate which enables the DV to be expressed at a speed that is scaled by GO-signal amplitude. This stage computes a Present Velocity Vector that is integrated by the PPV. The arm then moves towards the TPV until the DV equals zero. See Bullock, Cisek, and Grossberg (1998) for a discussion of where in the brain these quantities are computed.

Top-down motor expectations are also proposed to exist (Bullock and Grossberg, 1988; Bullock et al., 1998). For example, they can code the desired final, or target, position of a limb, such as an arm, during a reaching movement (see Figure 3). Such expectations can also be read-out as priming events that do not, in themselves, cause a movement (Georgopoulos et al., 1986). Volitional motor signals, called GO signals, can convert these top-down expectations into realized movements (Figure 3). This article suggests that, when these GO signals become tonically hyperactive, an individual can experience delusions of control, because the movements occur involuntarily. Such volitional signals are proposed to be generated in the basal ganglia; e.g., Horak and Anderson (1984a, 1984b).

Some Properties of Hallucinations

Auditory hallucinations are most frequently experienced during schizophrenia, including hearing one's own thoughts being spoken aloud, or hearing voices speaking to or about the patient. Visual, tactile, and olfactory hallucinations have also been reported. These include seeing frightening faces, being strangled, and experiencing food as repulsive. Frith (1998) has noted three properties of these experiences: They occur in the absence of sensation; self-generated activity is perceived to come from an external source; and the voice (in the case of an auditory hallucination) is perceived to come from an agent who is attempting to influence the patient. Frith hypothesized that these effects are due to disruption of the interactions between prefrontal cortex

and posterior brain regions. It is suggested below how learned top-down expectations acting from prefrontal cortex to posterior sensory cortices, among other top-down pathways, help to ensure the stability of normal learning and memory.

Other properties of hallucinations will also clarified by the hypothesis that modulatory top-down expectations are a source of hallucinations when they become imbalanced, including: Why unstructured auditory inputs, such as white noise, can increase the severity of hallucinations, while listening to speech or music helps to reduce them (Margo et al., 1981). This effect suggests that auditory hallucinations are formed within the same systems that are usually used to analyze auditory sensations. It is traced below to the manner in which top-down expectations are matched against bottom-up sensory information in order to focus attention upon salient combinations of features, and in so doing, to stabilize learning and memory about a changing world. Also clarified is the fact that what the external "voices" say tends to match the content of whispers or sub-vocal speech that are produced by the patient (Gould, 1949). This property even more directly supports the idea that hallucinations may be caused by imbalanced top-down expectations. In summary, the literature about such results collectively suggests that hallucinations may use the same systems that are usually used when people listen to external speech or generate inner speech.

Adaptive Resonance Theory

I propose that various of these properties may be mechanistically understood as follows. Adaptive Resonance Theory, or ART, proposes a solution of the *stability-plasticity dilemma*, or the problem of how can learning continue into adulthood without causing catastrophic forgetting (Carpenter and Grossberg, 1991; Grossberg, 1980, 1999b; Grossberg and Merrill, 1996; Grossberg and Stone, 1986). Otherwise expressed, how can we learn quickly without being forced to forget just as quickly? ART suggests that bottom-up adaptive processing of sensory information, from the world ever deeper into the brain, needs to be supplemented by top-down learned expectations in order to solve this problem.

ART suggests that bottom-up activation may, by itself, automatically activate target cell populations. In addition, top-down expectations are needed to learn prototypes that obey three properties: First, they select consistent bottom-up signals. Second, they suppress inconsistent bottom-up signals. Together these properties begin to focus attention on a set of critical features that are consistent with the learned expectation. Third, a top-down expectation, by itself, cannot fully activate target cells. It can only sensitize, modulate, or prime them to respond more easily and vigorously if they are matched by consistent bottom-up inputs. Were this not the case, we would hallucinate events that were not really there whenever we activated a top-down expectation. Such modulatory top-down expectations are proposed to exist, for example, in the top-down feedback from prefrontal to sensory cortices.

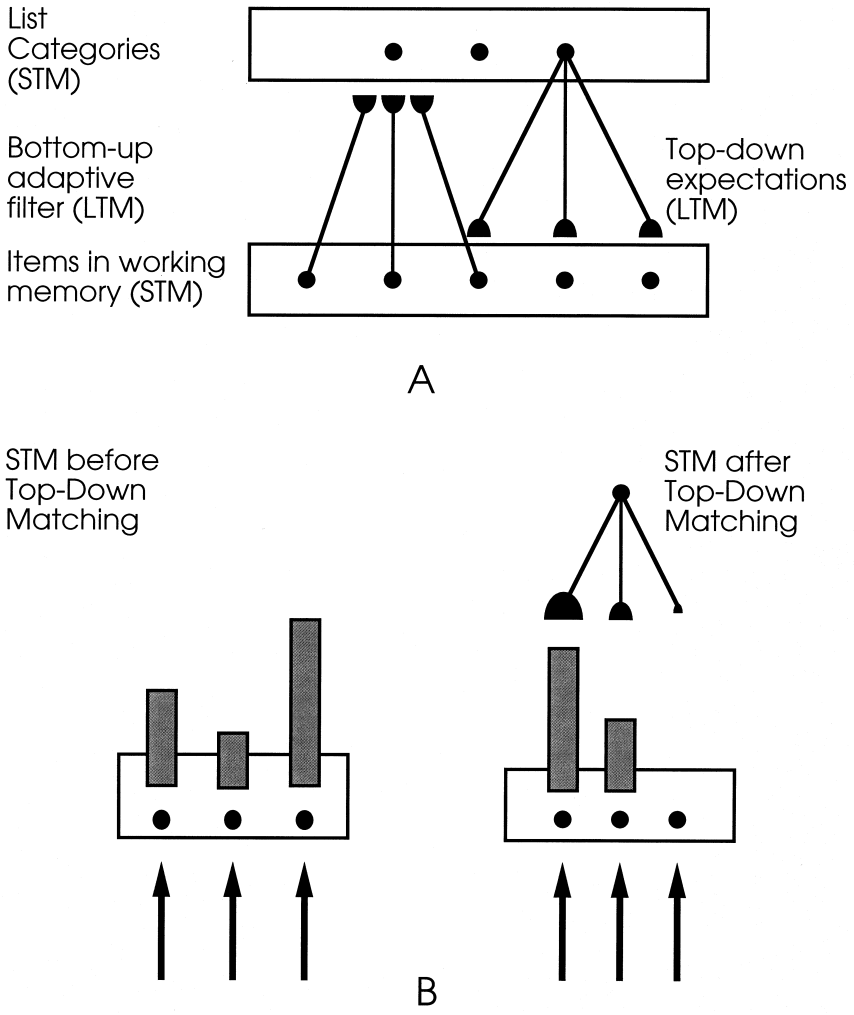

Figure 4. (A) During auditory perception, bottom-up inputs activate the short-term memory (STM) traces of item representations in a working memory. Stored working memory items, in turn, send bottom-up signals towards a subsequent processing stage at which list categories, or chunks, are activated in STM. These bottom-up signals are multiplied by learned LTM traces which influence the selection of the list categories that are stored in STM. The list categories, in turn, activate LTM-modulated top-down expectation signals that are matched against the active STM pattern in working memory using a circuit like the one in Figure 1. **(B)** This matching process confirms and amplifies those STM item activations that are in the prototypes of the active top-down expectations, and suppress those that are not. The height of the histograms in the left panel indicates the activation level of target cells in response to bottom-up inputs alone. The corresponding height in the right panel indicates the activation level of these cells after a top-down prototype is read-out. The size of the hemidisks represents the size of the LTM traces in the prototype. Attention starts to focus on the leftmost pair of cells when their activities are selected and amplified. The rightmost cell activity is suppressed, since it has no top-down signal.

Figure 4a illustrates the interplay of these bottom-up and top-down processes, and Figure 4b schematizes what is meant by attentional focusing. The reader can, for example, interpret the bottom-up and top-down interaction in Figure 4a in terms of the reciprocal activation that occurs between sensory cortices and prefrontal cortex, or due to other cortico-cortical or thalamocortical interactions, even those between visual cortical area V1 and the Lateral Geniculate Nucleus (Sillito

et al., 1994). In Figure 4a, distributed behavioral items are temporarily stored in short-term memory as they send signals through bottom-up pathways to a subsequence processing stage. Adaptive weights, or long-term memory traces, in the synaptic knobs of these pathways (the hemi-disks in Figure 4) multiply these signals before they are added up at target cells. The target calls, in turn, compete among themselves to select one, or a small number, of winning cells that receive the largest total bottom-up inputs. In various applications, these winning cells may represent a recognition category that responds selectively to particular combinations of items stored in short-term memory. Because of the hierarchical organization of such networks during the progressive elaboration of bottom-up information, the "features" within a distributed activation pattern on a given processing level may be the "categories" for a previous processing level, much as information gets progressively more abstractly encoded as it ascends the ventral processing stream through visual cortex, inferotemporal cortex, and prefrontal cortex (e.g., Desimone and Ungerleider, 1989; Goldman-Rakic et al., 1991).

If only these bottom-up signals are considered, then the winning cells activate learning within the abutting synaptic knobs. These synaptic knobs are proposed to learn a time-average of the signal amplitudes that are active in the pathways leading to these synapses when their target cells win the competition for activation. This type of model is called competitive learning or a self-organizing map (Grossberg, 1976a, 1978; Kohonen, 1989; von der Malsburg, 1973). Grossberg (1976a) proved theorems about conditions under which stable learning is achieved in these systems, before demonstrating that they undergo catastrophic forgetting under general input conditions. Grossberg (1976b) then introduced ART, including the role to learned top-down expectations, to suggest how the benefits of self-organizing maps could be preserved without the disadvantages–which many other learning models share–of catastrophic forgetting.

Learned top-down expectations with the properties listed above are one of the key mechanisms needed to prevent catastrophic forgetting. Figure 4b schematizes the property of ART Matching whereby a learned top-down expectation can select and amplify the activities of cells that are within the learned prototype of the active category. In Figure 4b, the size of the synaptic knobs codes for the size of the adaptive weights in the top-down pathways. Because the rightmost cell has a very small top-down weight, it does not receive any top-down signal, and thus its activation is inhibited, even though it receives a large bottom-up input. Due to this matching process, the network starts to "pay attention" to the activation pattern on the two remaining cells. In other words, the top-down expectation selects and amplifies cells whose activities are "confirmed" by the "hypothesis" that is represented by the top-down prototype, and suppresses the activities of cells which are not.

Figure 1 illustrates the type of circuit which is suggested by ART to carry out this top-down matching function. It is a top-down on-center off-surround network. Figure 1 notes that bottom-up inputs can, by themselves, activate their target cells. If a top-down expectation is turned on, then it tries to excite the cells that lie within its on-center prototype. But it also inhibits cells within a broader off-surround. The excitation and inhibition received by cells in the on-center are balanced so that these cells cannot be fully activated by top-down signals. At best, they can be slightly excited, or subliminally primed. When a bottom-up input, as in pathway 2, is within the prototype's on-center, then its target cell can remain active, and even be amplified, when the top-down expectation turns on. This is because the bottom-up input together with the top-down on-center can overcome the effects of the top-down inhibition; it is a case of two-against-one. When a bottom-up input, as in pathway 1, receives only a top-down inhibitory signal, it is shut off in the top-down mode, even if it receives a strong bottom-up excitatory input; it is a case of one-against-one.

Recent modeling work has suggested how this top-down circuit is realized within the laminar circuits of the neocortex (Grossberg, 1999a); see Figure 2. In particular, it is suggested that output signals from layer 6 of a higher cortical area can attentionally prime layer 4 of a lower cortical area via a top-down on-center off-surround network that passes through layer 6 of the lower cortical area. Several routes exist whereby feedback from higher cortical levels can reach layer 6 before being relayed through the on-center off-surround network between layer 6 and 4. Consider, for definiteness, the case of feedback from visual cortical area V2 to area V1 (Lamme et

al., 1998). Here, the majority of feedback signals from layer 6 in V2 pass into layer 1 (Cauller, 1995; Rockland and Virga, 1989). There they stimulate the apical dendrites of layer 5 pyramidal cells whose axons send collaterals in layer 6 (Lund and Boothe, 1975; Gilbert and Wiesel, 1979). Reversible deactivation studies of monkey V2 have shown that feedback from V2 to V1 does indeed have an on-center off-surround form (Bullier et al., 1996). In addition, the V1 layer whose activation is most reduced by cutting off V2 feedback is layer 6 (Sandell and Schiller, 1982).

The top-down input to layer 1 can also influence the dendrites of pyramidal cells in layer 2/3, which carry out perceptual grouping through their horizontal connections (Das and Gilbert, 1999; Hirsch and Gilbert, 1991). But here feedback from V2 seems to have much less effect (Sandell and Schiller, 1982) possibly because inhibitory interneurons with dendrites in layer 1 are present in layer 2/3 but not in layer 6 (Lund, 1987; Lund et al., 1988; Lund and Wu, 1997). Thus, there are at least two routes whereby top-down signals can attentionally modulate the processing of bottom-up information. Both of these routes seem to have a modulatory effect on their target cells due to a balance between excitation and inhibition. Both may, in principle, be imbalanced by a volitional signal during normal fantasy activities, and this volitional signal may become hyperactive during hallucinatory episodes. Thus, it would be of great interest to test for the existence and source of pathways to layer 4, as well as to layer 2/3, which are capable of upsetting this balance as a result of volition, possibly by inhibiting the inhibitory interneurons (Figure 1), so that the top-down prototypes can be supraliminally expressed.

Adaptive Resonance: Linking Expectation, Attention, and Stable Learning

Figure 4 summarizes the ART proposal that stable learning is linked to top-down expectations and attention. In other words, our capacity to act like "intentional" beings is linked to our ability to continue learning rapidly and in a stable fashion throughout life. Figure 5 indicates that attended feature clusters reactivate their bottom-up pathways. These, in turn, reactivate their top-down pathways. A feedback resonance hereby develops between the attended information on multiple levels of the network. This resonance is a context-sensitive state that designates that the patterns of activation which are carried by the resonance are worthy of being learned by the system's adaptive weights, or long-term memory traces. ART proposes that all conscious states are resonant states, and hereby suggests a link between learning, intention, attention, and consciousness (see Grossberg (1999b) for a recent review).

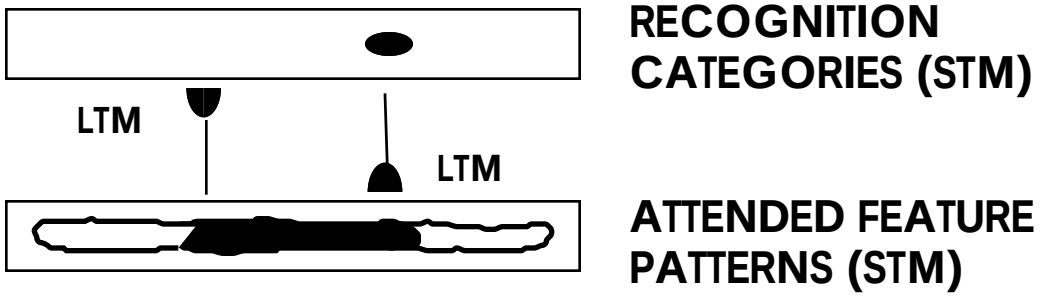

Figure 5. When a subset of cells is selected by active top-down prototypes, they can continue to send bottom-up signals to the category representations at the next processing stage. The categorical cells can, in turn, continue to send top-down signals to the attended feature pattern. A resonant state hereby locks in, whose cells are synchronized, amplified, and prolonged long enough for the LTM traces which support the resonant cells to learn from their activity patterns.

The resonant state synchronizes, amplifies, and prolongs the cell activations that it includes. This change of energy and time scales enables the more slowly varying adaptive weights to learn the activities that they can correlate. By discarding signals other than those that are attended, the resonant state helps to prevent catastrophic forgetting. It is because the resonance triggers learning that the theory is called *adaptive* resonance theory.

How Hyperactive Volitional Signals May Cause Hallucinations

This description clarifies how the top-down expectations can learn prototypes which are the important memories that the system uses for recall. When such expectations are used only for top-down priming and matching, they organize the processing of bottom-up information, and help to focus attention upon salient combinations of features. But they cannot, by themselves, lead to suprathreshold experiences. As noted above, ART predicts that evolution has exploited the top-down balance between excitation and inhibition to enable internal imagery, speech, planning, and other fantasy activities to occur. In particular, suppose that an inhibitory volitional signal (see Figure 1) can inhibit the inhibitory interneurons in the top-down on-center off-surround cortical priming networks during a time when internal fantasy is desired. One way in which this might be achieved is via inhibitory signals that inhibit the off-surround cells in layer 4 (Figure 2). Then the top-down expectations can, by themselves, activate target cells and lead to resonant, hence conscious, activity. Since these top-down expectations learn the important feature clusters to which the network pays attention, they can read-out enough information to generate a meaningful internal fantasy. I propose that these volitional signals can sometimes become tonically hyperactive, and can fire by themselves, without willful control by the individual. When this happens, the top-down expectations can elicit conscious perceptual experiences that can occur without bottom-up input and are not under the control of the individual, and are therefore experienced as hallucinations.

Hallucinations during Audition and Speech

ART models have been used to explain a great deal of psychophysical data about auditory and speech perception (e.g., Boardman et al., 1999; Cohen and Grossberg, 1986; Grossberg, 1986, 1999c; Grossberg et al., 1997; Grossberg and Myers, 1999; Grossberg and Stone, 1986). In these applications, the conscious resonances emerge through feedback interactions between working memories–that store sequences of auditory inputs as evolving spatial patterns of activity–and list chunking, or grouping, networks–that form unitized representations of different sequences of the events that are represented by the working memory. These unitized representations encode phonetically important units, notably words and syllables. Higher-level list chunks can represent combinations of these unitized representations.

In this situation, hallucinations would use the same systems that are usually used when listening to external speech or to generate inner speech. Here the top-down expectations are read-out by activation of the list chunks. How external “voices” tend to match the content of whispers or sub-vocal speech (Gould, 1949) is clarified by ART, because both exploit the top-down expectations of the individual. Here it is useful to distinguish sensory expectations from motor expectations within a perception-action cycle (Figure 6): Top-down sensory expectations help to unitize the contents of bottom-up sensory signals, whereas top-down motor expectations help to unitize the motor gestures that are used to read-out articulatory movements. Under normal conditions, sensory expectations of self-generated sounds are subliminally primed when motor expectations are used to produce speech. With a hyperactive volitional system, these subliminal primes can become suprathreshold.

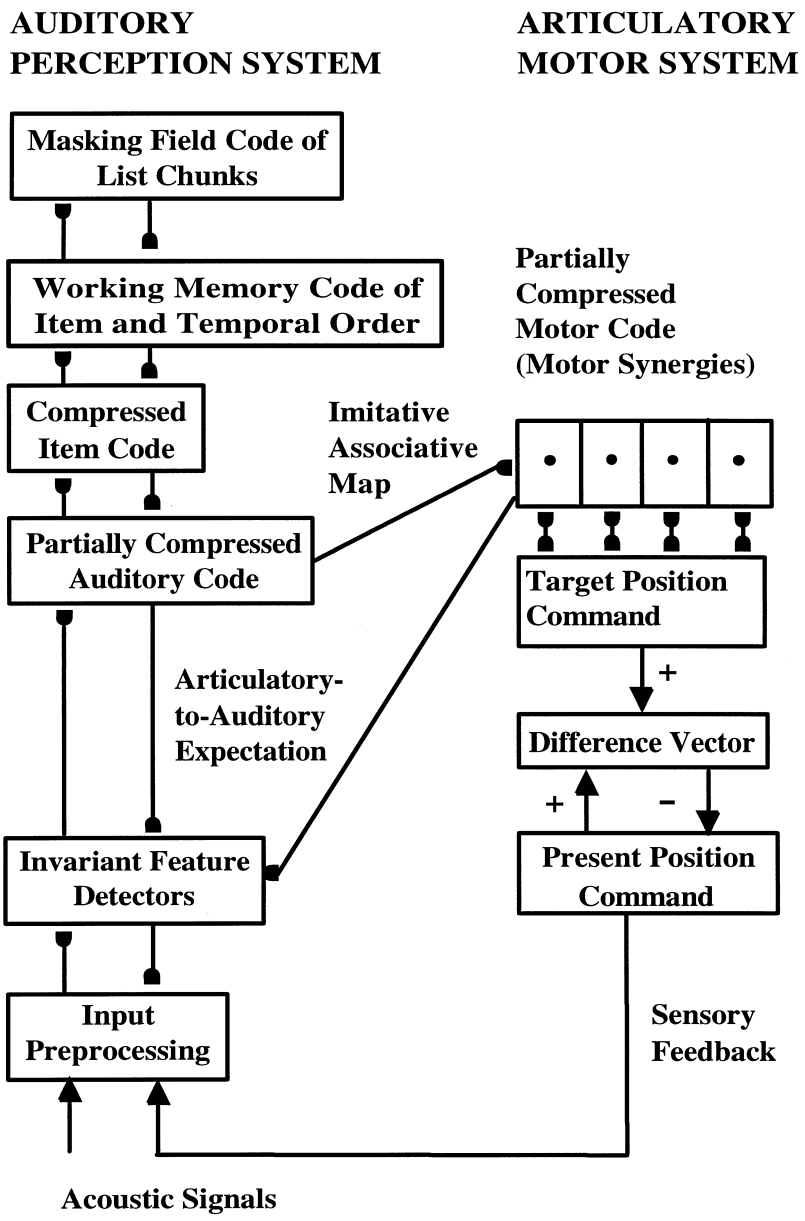

Figure 6. Schematic of how auditory and articulatory pathways are linked in a perception-action cycle for purposes of auditory perception and speech production. Reciprocal bottom-up and top-down circuits are proposed to exist in both the auditory and the articulatory systems. In the auditory system, these loops unitize auditory signals into familiar language units. In the articulatory system, they help to unitize the speech gestures that control spoken language. The systems are linked by an imitative map that enables auditory representations to control articulatory productions.

ART also clarifies why unstructured auditory inputs, such as white noise, can increase the severity of hallucinations, while listening to speech or music helps to reduce them (Margo et al., 1981). Both types of data are consistent with the ART Matching Rule (Figure 1), which selects those auditory signals that are consistent with the top-down expectation, while suppressing those that are not. In the case of white noise, the broad-band nature of the noise enables it to supply more bottom-up activity to the active components of the top-down expectation, and thereby to generate an even stronger hallucination. This property of the ART Matching Rule also plays a key role in

explaining data about normal auditory and speech perception (e.g., Grossberg, 1999b, 1999c). Data about phonemic restoration are particularly useful towards illustrating how this happens. Suppose, for example, that a listener hears a noise of suitable duration followed immediately by the words "eel is on the…." (Warren, 1984; Warren and Sherman, 1974). If this string of words is followed by the word "orange," then "noise-eel" sounds like "peel." If the word "wagon" completes the sentence, then "noise-eel" sounds like "wheel." If the final word is "shoe," then "noise-eel" sounds like "heel." If some formants of the expected sound are missing from the noise, then only a partial reconstruction is heard (Samuel, 1981a, 1981b). If silence replaces the noise, then only silence is heard.

Phonemic restoration vividly shows that the bottom-up occurrence of the noise is not sufficient for us to hear it. Rather, the sound that we *expect* to hear based upon our previous language experiences influences what we do hear, at least if the sentence is said quickly enough. Somehow, the brain works "backwards in time" to allow the meaning imparted by a later word to alter the sounds that we consciously perceive in an earlier word. The delayed perception is explained in ART by the amount of time that is needed to form a bottom-up/top-down resonance between working memory and list chunk representations; see Grossberg et al., (1997) and Grossberg and Myers (1999) for computer simulations of such resonances. In fact, ART claims that conscious speech is a *resonant wave* that evolves across the brain's working memories. The selection of those noise components that are consistent with the top-down expectations of this resonant wave is achieved by the ART Matching Rule. In particular, the matching process cannot "create something out of nothing." It can, however, selectively amplify the expected features in the bottom-up signal and suppress the rest. I suggest that a similar thing happens when white noise enhances a hallucination.

Reset, Memory Search, and Stable Learning and Memory

How does the ART Matching Rule explain how speech or music helps to reduce hallucinations (Margo et al., 1981)? In brief, a bottom-up input that does not match an internally generated top-down expectation can *reset* it and impose the representation mandated by the bottom-up input. This reset operation is hypothesized to be part of the machinery whereby the brain updates its information processing in response to new inputs. It also is predicted to help search for new, or more appropriate, representations with which to categorize novel information, and in so doing, to help stabilize learning and memory.

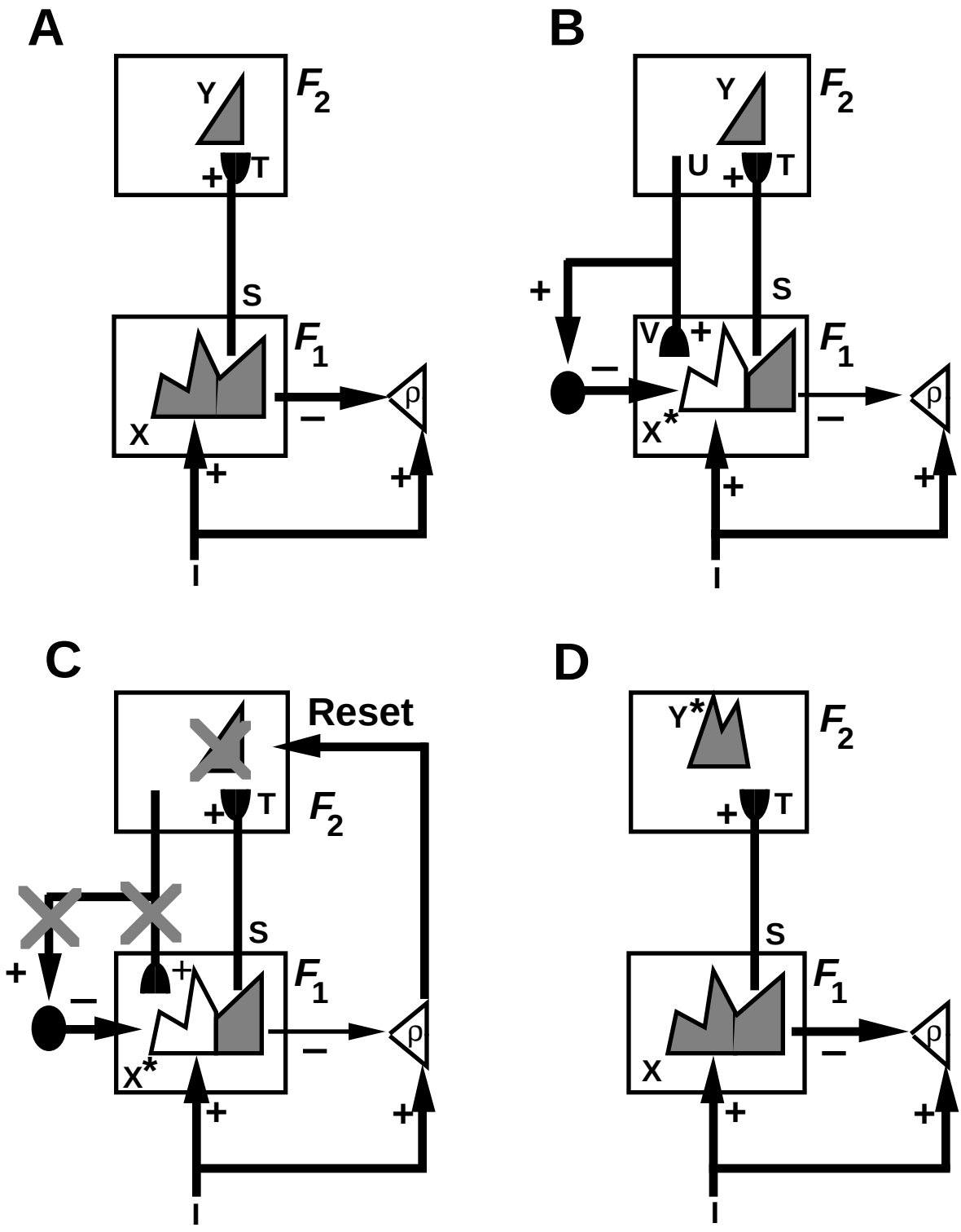

Figure 7. ART search for a recognition code: **(A)** The input pattern I is instated across the feature detectors at level F_1 as a short term memory (STM) activity pattern X. Input I also nonspecifically activates the orienting subsystem A; see Figure 8. STM pattern X is represented by the hatched pattern across F_1. Pattern X both inhibits A and generates the output pattern S. Pattern S is multiplied by long term memory (LTM) traces and added at F_2 nodes to form the input pattern

T, which activates the STM pattern Y across the recognition categories coded at level F_2. **(B)** Pattern Y generates the top-down output pattern U which is multiplied by top-down LTM traces and added at F_1 nodes to form the prototype pattern V that encodes the learned expectation of the active F_2 nodes. If V mismatches I at F_1, then a new STM activity pattern X^* is generated at F_1. X^* is represented by the hatched pattern. It includes the features of I that are confirmed by V. Inactivated nodes corresponding to unconfirmed features of X are unhatched. The reduction in total STM activity which occurs when X is transformed into X^* causes a decrease in the total inhibition from F_1 to A. **(C)** If inhibition decreases sufficiently, A releases a nonspecific arousal wave to F_2, which resets the STM pattern Y at F_2. **(D)** After Y is inhibited, its top-down prototype signal is eliminated, and X can be reinstated at F_1. Enduring traces of the prior reset lead X to activate a different STM pattern Y^* at F_2. If the top-down prototype due to Y^* also mismatches I at F_1, then the search for an appropriate F_2 code continues until a more appropriate F_2 representation is selected. Then an attentive resonance develops and learning of the attended data is initiated. [Reprinted with permission from Grossberg and Merrill (1996).]

To see how reset and search occur, consider Figure 7. Figure 7 suggests how the bottom-up and top-down processes work together to control reset and memory search. For simplicity of exposition, the two processing levels depicted there are denoted by F_1 and F_2, where F_1 represents distributed feature patterns and F_2 unitizes them into categorical representations. As noted above, learning in ART does not occur when some winning F_2 activities are stored in STM. Instead, activation of F_2 nodes may be interpreted as "making a hypothesis" about an input at F_1. When F_2 is activated, it quickly generates an output pattern that is transmitted along the top-down adaptive pathways from F_2 to F_1. These top-down signals are multiplied in their respective pathways by LTM traces at the semicircular synaptic knobs of Figure 7B. The LTM-gated signals from all the active F_2 nodes are added to generate the total top-down feedback pattern from F_2 to F_1. It is this pattern that plays the role of a learned expectation. Activation of this expectation may be interpreted as "testing the hypothesis", or "reading out the prototype", of the active F_2 category. As shown in Figure 7B, ART networks are designed to match the "expected prototype" of the category against the bottom-up input pattern, or exemplar, to F_1. Cells that are activated by this exemplar are suppressed if they do not correspond to large LTM traces in the top-down prototype pattern. The resultant F_1 pattern encodes the cluster of input features that the network deems relevant to the hypothesis based upon its past experience. This resultant activity pattern, called X^* in Figure 7B, encodes the pattern of features to which the network "pays attention".

If the expectation is close enough to the input exemplar, then a state of resonance develops as the attentional focus takes hold. The pattern X^* of attended features reactivates the F_2 category Y which, in turn, reactivates X^*. The network locks into a resonant state through a positive feedback loop that dynamically links, or binds, X^* with Y. The resonance binds spatially distributed features into either a stable equilibrium or a synchronous oscillation (Grossberg, 1976b), much like the synchronous feature binding in visual cortex that has recently attracted so much interest (Eckhorn et al., 1988; Gray and Singer, 1989; Grossberg and Grunewald 1997).

In ART, the resonant state, rather than bottom-up activation, is predicted to drive the learning process. The resonant state persists long enough, at a high enough activity level, to activate the slower learning processes in the LTM traces. This helps to explain how the LTM traces can regulate the brain's fast information processing without necessarily learning about the signals that they process. Through resonance as a mediating event, the combination of top-down matching via the ART Matching Rule, and its attentional focusing properties, helps to stabilize ART learning and memory in response to an arbitrary input environment. The stabilizing properties of top-down matching may be one reason for the ubiquitous occurrence of reciprocal bottom-up and top-down cortico-cortical and cortico-thalamic interactions in the brain.

How is the Generality of a Perceptual Category Controlled?

How does reset and memory search enter into this description of resonant dynamics? This question is closely related to the question of what combinations of features or other information are bound together into object or event representations. ART provides an answer to this question that

overcomes problems faced by earlier models. In particular, ART systems learn prototypes, rather than exemplars, because the attended feature vector X^*, rather than the input exemplar itself, is learned. Both the bottom-up LTM traces that tune the category nodes and the top-down LTM traces that filter the learned expectation learn to correlate activation of F_2 nodes with the set of all *attended* X^* vectors that they have ever experienced. These attended STM vectors assign less STM activity to features in the input vector I that mismatch the learned top-down prototype V than to features that match V.

Given that ART systems learn prototypes, how can they also learn to recognize unique experiences, such as a particular view of a friend's face? The prototypes learned by ART systems accomplish this by realizing a qualitatively different concept of prototype than that offered by previous models. In particular, ART suggests how prototypes, and thus the feature bundles that should form the contents of hallucinatory experiences, form in a way that is designed to conjointly maximize category generalization while minimizing predictive error (Bradski and Grossberg, 1995; Carpenter and Grossberg, 1987a, 1987b; Carpenter et al., 1991; Carpenter et al., 1992). As a result, ART prototypes can automatically learn individual exemplars when environmental conditions require highly selective discriminations to be made. How the matching process achieves this is discussed below.

This matching property modulates the criterion of when mismatch between bottom-up and top-down information is too great for a resonance to develop. In ART, when such a mismatch occurs, then the F_2 category is quickly reset and a memory search for a better category is initiated (Figures 7C and 7D). This combination of top-down matching via the ART Matching Rule, attentional focusing, and memory search is what stabilizes ART learning and memory in an arbitrary input environment. The attentional focusing by top-down matching prevents inputs that represent irrelevant features at F_1 from eroding the memory of previously learned LTM prototypes. In addition, the memory search resets F_2 categories so quickly when their prototype V mismatches the input vector I that the more slowly varying LTM traces do not have an opportunity to correlate the attended F_1 activity vector X^* with them. Conversely, the resonant event, when it does occur, maintains and amplifies the matched STM activities for long enough and at high enough amplitudes for learning to occur in the LTM traces.

Whether or not a resonance occurs depends upon the level of mismatch, or novelty, that the network is prepared to tolerate. Novelty is measured by how well a given exemplar matches the prototype that its presentation evokes. The criterion of an acceptable match is defined by an internally controlled parameter that is called *vigilance* (Carpenter and Grossberg, 1987a). The vigilance parameter is computed in the orienting subsystem A; see Figure 8. Vigilance weighs how similar an input exemplar I must be to a top-down prototype V in order for resonance to occur. Resonance occurs if $\rho|I| - |X^*| \leq 0$, where $|I|$ and $|X^*|$ measure the total amount of activity in the bottom-up input vector I and the matched activity pattern X^*, respectively. This inequality says that the F_1 attentional focus X^* inhibits A more than the input I excites it. If A remains quiet, then an F_1 ÖF_2 resonance can develop.

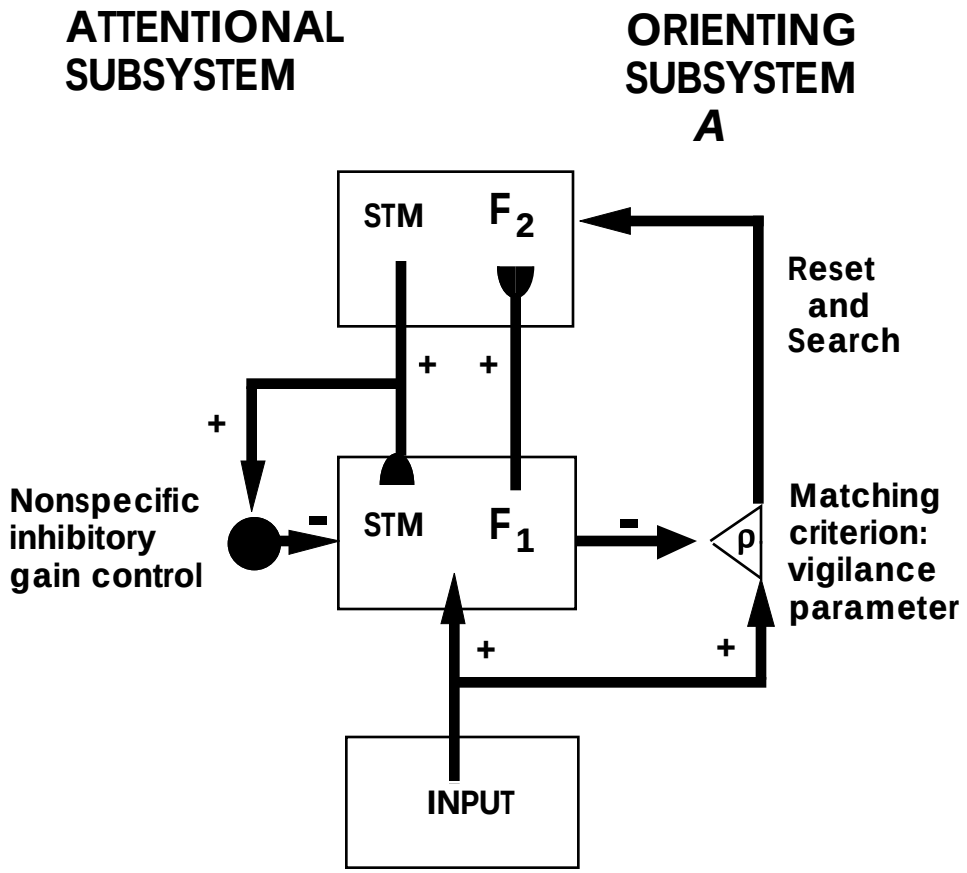

Figure 8. An example of a model ART circuit in which attentional and orienting circuits interact. Level F_1 encodes a distributed representation of an event by a short term memory (STM) activation pattern across a network of feature detectors. Level F_2 encodes the event using a compressed STM representation of the F_1 pattern. Learning of these recognition codes occurs at the long term memory (LTM) traces within the bottom-up and top-down pathways between levels F_1 and F_2. The top-down pathways read-out learned expectations whose prototypes are matched against bottom-up input patterns at F_1. The size of mismatches in response to novel events are evaluated relative to the vigilance parameter ρ of the orienting subsystem A. A large enough mismatch resets the recognition code that is active in STM at F_2 and initiates a memory search for a more appropriate recognition code, as described in Figure 7. Output from subsystem A can also trigger an orienting response.

Either a larger value of ρ or a smaller match ratio $|X^*|I|^{-1}$ makes it harder to satisfy the resonance inequality. When ρ grows so large or $|X^*|I|^{-1}$ is so small that $\rho|I| - |X^*| > 0$, then A generates an arousal burst, or novelty wave, that resets the STM pattern across F_2 and initiates a bout of hypothesis testing, or memory search. During search, the orienting subsystem interacts with the attentional subsystem (Figures 7C and 7D) to rapidly reset mismatched categories and to select better F_2 representations with which to categorize novel events at F_1, without risking unselective forgetting of previous knowledge. Search may select a familiar category if its prototype is similar enough to the input to satisfy the resonance criterion. The prototype may then be refined by attentional focusing. If the input is too different from any previously learned prototype, then an uncommitted population of F_2 cells is selected and learning of a new category is initiated.

Because vigilance can vary across learning trials, recognition categories capable of encoding widely differing degrees of generalization or abstraction can be learned. Low vigilance leads to broad generalization and abstract prototypes. High vigilance leads to narrow generalization and to prototypes that represent fewer input exemplars, even a single exemplar. ART hereby

clarifies how we can learn abstract prototypes with which to recognize abstract categories of faces and dogs, as well as "exemplar prototypes" with which to recognize individual faces and dogs. A single system can learn both, as the need arises, by increasing vigilance just enough to activate A if a previous categorization leads to a predictive error.

Thus the contents of a conscious percept can be modified by task-sensitive vigilance control signals. In a similar vein, ART predicts that when highly specific prototypes are learned, then specific hallucinations will be read-out by these prototypes, and when abstract prototypes are learned, then they will lead to similarly abstract hallucinations. This prediction adds a new twist to the idea that hallucinations are formed within the same systems that are usually used to analyze perceptual sensations. It suggests not only that learned top-down expectations control the perceptual contents of hallucinations, but also that these contents may be dynamically refined or generalized based on the task-selective constraints that were operative during individual learning experiences.

ACKNOWLEDGMENTS

*This work was supported in part by the Defense Advanced Research Projects Agency and the Office of Naval Research (ONR N00014-95-1-0409), and the National Science Foundation (NSF IRI-97-20333). The author wishes to thank Robin Amos and Diana Meyers for their valuable assistance in the preparation of the manuscript.

REFERENCES

Boardman, I., Grossberg, S., Myers, C. & Cohen, M. (1999). Neural Dynamics of Perceptual Order and Context Effects for Variable-Rate Speech Syllables. *Perception and Psychophysics*, in press.

Bradski, G. & Grossberg, G. (1995). Fast learning VIEWNET architectures for recognizing 3-D objects from multiple 2-D views. *Neural Networks, 8*, 1053-1080.

Bullier, J., Hupe, J.M., James, A. & Girard, P. (1996). Functional interactions between areas V1 and V2 in the monkey. *Journal of Physiology (Paris), 90*, 217-220.

Bullock, D., Cisek, P. & Grossberg, S. (1998). Cortical networks for control of voluntary arm movements under variable force conditions. *Cerebral Cortex, 8*, 48-62.

Bullock, D. & Grossberg, S. (1988). Neural dynamics of planned arm movements: Emergent invariants and speed-accuracy properties during trajectory formation. *Psychological Review, 95*, 49-90.

Carpenter, G.A. & Grossberg, S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing, 37*, 54-115.

Carpenter, G.A. & Grossberg, S. (1987b). ART 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics, 26*, 4919-4930.

Carpenter, G.A. & Grossberg, S. (1991). *Pattern Recognition by Self-Organizing Neural Networks*. Cambridge: MIT Press.

Carpenter, G.A., Grossberg, S. Markuzon, N., Reynolds, J.H., & Rosen, D.B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks, 3*, 698-713.

Carpenter, G.A., Grossberg, S. & Reynolds, J.H. (1991). ARTMAP: Superviesed real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks, 4*, 565-588.

Cauller, L. (1995). Layer I of primary sensory neocortex: Where top-down converges upon bottom-up. *Behavioral Brain Research, 71*, 163-170.

Cohen, M.A. & Grossberg, S. (1986). Neural dynamics of speech and language coding: Developmental programs, perceptual grouping, and competition for short term memory. *Human Neurobiology, 5*, 1-22.

Das, A. & Gilbert, C.D. (1999). Topography of contextual modulations mediated by short-rante interactions in primary visual cortex. *Nature, 399*, 656-661.

Desimone, R. & Ungerleider, L.G. (1989). Neural mechanisms of visual processing in monkeys. In F. Boller and J. Grafman (Eds.), *Handbook of Neuropsychology* , Vol. 2, (pp. 267-299). Amsterdam: Elsevier.

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M. & Reitboeck, H.J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biological Cybernetics, 60*, 121-130.

Frith, C.D. (1998). The role of the prefrontal cortex in self-consciousness: the case of auditory hallucinations. In A.C. Roberts, T.W. Robbins and L. Weiskrantz (Eds.) *The Prefrontal Cortex: Executive and Cognitive Functions, (*pp. 181-194). Oxford: Oxford University Press.

Georgopoulos, A.P., Schwartz, A.B. & Kettner, R.E. (1986). Neuronal population coding of movement direction. *Science, 233,* 1416-1419.
Gilbert, C.D. & Wiesel, T.N. (1979). Morphology and intracortical projections of functionally characterised neurones in the cat visual cortex. *Nature, 280,* 120-125.
Goldman-Rakic, P.S., Funahashi, S. & Bruce, C.J. (1991). Neocortical memory circuits. *Quarterly Journal of Quantitative Biology, 55,* 1025-1038.
Gould, L.N. (1949). Auditory hallucinations and subvocal speech. *Journal of Nervous and Mental Disorders, 109,* 418-427.
Gray, C.M. & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences USA, 86,* 1698-1702.
Grossberg, S. (1976a). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23,* 121-134.
Grossberg, S. (1976b). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics, 23,* 187-202.
Grossberg, S. (1978). A theory of human memory: Self-Organization and performance of sensory-motor codes, maps, and plans. In R. Rosen and F. Snell (Eds.). *Progress in Theoretical Biology,* (pp. 233-374). New York: Academic Press.
Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review, 1,* 1-51.
Grossberg, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E.C. Schwab and H.C. Nusbaum (Eds.) *Pattern Recognition by Humans and Machines, Vol. 1: Speech Perception,* (pp. 187-294). New York: Academic Press.
Grossberg, S. (1999a). How does the cerebral cortex work? Learning, attention, and grouping by the lminar circuits of visual cortex. *Spatial Vision, 12,* 163-186.
Grossberg, S. (1999b). The link between brain learning, attention, and consciousness. *Consciousness and Cognition, 8,* 1-44.
Grossberg, S. (1999c). Pitch-based streaming in auditory perception. In N. Griffith and P. Todd (Eds.). *Musical Networks: Parallel Distributed Perception and Performance.,* (pp. 117-140). Cambridge: MIT Press.
Grossberg, S., Boardman, I., & Cohen, M.A. (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 481-503.
Grossberg, S. & Grunewald, A. (1997). Cortical synchronization and perceptual framing. *Journal of Cognitive Neuroscience, 9,* 117-132.
Grossberg, S. & Merrill, J.R.W. (1996). The hippocampus and cerebellum in adaptively timed learning, recognition, and movement. *Journal of Cognitive Neuroscience, 8,* 257-277.
Grossberg, S. & Myers, C. (1999). The resonant dynamics of conscious speech: Interword integration and duration-dependent backward effects. *Psychological Review.,* in press.
Grossberg, S. & Raizada, R.D.S. (1999). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. Technical Report CAS/CNS TR-99-008.
Grossberg, S. & Stone, G. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review, 93,* 46-74.
Hirsch, J.A. & Gilbert, C.D. (1991). Synaptic physiology of horizontal connections in the cat visual cortex. *Journal of Neuroscience, 11,* 1800-1809.
Horak, R.B. & Anderson, M.E. (1984a). Influence of globus pallidus on arm movements in monkeys. I. Effects of nainic acid-induced lesions. *Journal of Neurophysiology, 52,* 290-304.
Horak, R.B. & Anderson, M.E. (1984b). Influence of globus pallidus on arm movements in monkeys. II. Effects of stimulations. *Journal of Neurophysiology, 52,* 305-322.

Kohonen, T. (1989). *Self-Organization and Associative Memory*, 3rd edition. Berlin: Springer-Verlag.
Lamme, V.A.F., Super, H. & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology, 8*, 529-535.
Lund, J.S. (1987). Local circuit neurons of macaque monkey striate cortex: I. Neurons of laminae 4C and 5A. *Journal of Comparative Neurology, 257*, 60-92.
Lund, J.S. & Boothe, R.G. (1975). Interlaminar connections and pyramidal neuron organisation in the visual cortex, area 17, of the macaque monkey. *Journal of Comparative Neurology, 159*, 305-334.
Lund, J.S., Hawken, M.J. & Parker, A.J. (1988). Local circuit neurons of macaque monkey striate cortex: II. Neurons of laminae 5B and 6. *Journal of Comparative Neurology, 276*, 1-29.
Lund, J.S. & Wu, C.Q. (1997). Local circuit neurons of macaque monkey striate cortex: IV. Neurons of laminae1-3A. *Journal of Comparative Neurology, 384*, 109-126.
Margo, A., Hemsley, D.R. and Slade, P.D. (1981). The effects of varying auditory input on schizophrenic hallucinations. *British Journal of Psychiatry, 39*, 101-107.
Rockland, K.S. & Virga, A. (1989). Terminal arbors of individual "feedback" axons projecting from area V2 to V1 in the macaque monkey: A study using immunohistochemistry of anterogradely transported phaseolus vulgaris-leucoagglutinin. *Journal of Comparative Neurology, 285*, 54-72.
Salin, P. & Bullier, J. (1995). Corticocortical connections in the visual system: Structure and function. *Physiological Reviews, 75*, 107-154.
Samuel, A.G. (1981a). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 1124-1131.
Samuel, A.G. (1981b). Phonemic resotration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110*, 474-494.
Sandell, J.H. & Schiller, P.H. (1982). Effect of cooling area 18 on striate cortex cells in the squirrel monkey. *Journal of Neurophysiology, 48*, 38-48.
Sartorius, N., Shapiro, R. & Jablensky, A. (1974). The international pilot study of schizophrenia. *Schizophrenia Bulletin, 1*, 21-35.
Sillio, A.M., Jones, H.E., Gerstein, G.L. & West, D.C. (1994). Feature-linked synchronization of the thalamic relay cell firing induced by feedback from the visual cortex. *Nature, 369*, 479-482.
Von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik, 14*, 85-100.
Warren, R.M. (1984). Perceptual restoration of obliterated sounds. *Psychological Bullein, 96*, 371-383.
Warren, R.M. & Sherman, G.L. (1974). Phonemic restorations based on subsequence context. *Perception and Psychophysics, 16*, 150-156.