The Attentive Brain

Stephen Grossberg[1]
Department of Cognitive and Neural Systems[2]
and
Center for Adaptive Systems
Boston University
677 Beacon Street
Boston, MA 02215

Invited Article for *American Scientist*
83, 438–449, 1995

Technical Report CAS/CNS-TR-95-012

Neural networks that match sensory inputs with learned expectations help to explain how humans see, hear, learn and recognize information.

Myriad signals relentlessly bombard our senses. These signals may arrive in disconnected pieces, yet we can integrate them as unified moments of conscious experience. The apparent singularity and coherence of an experience depends on how the brain processes environmental events. That processing concentrates on context. If you look at a complex picture, such as a photograph of a famous face (Figure 1A), you probably recognize it at a glance, but you might never recognize it by looking at it piece by piece (Figure 1B). Such context-dependent processing emerges because the brain typically operates on sensory data in parallel, or in batches.

Visual signals from a scene typically reach your eyes simultaneously, so parallel processing begins at the retina. Sounds that make up a word, on the other hand, reach your

[1] Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-92-J-0225), the Advanced Research Projects Agency (ONR N00014-92-J-4015), and the Office of Naval Research (ONR N00014-91-J-4100 and ONR N00014-92-J-1309).

[2] Acknowledgments: The authors wish to thank Diana Meyers and Cynthia Bradford for their valuable assistance in the preparation of the manuscript.

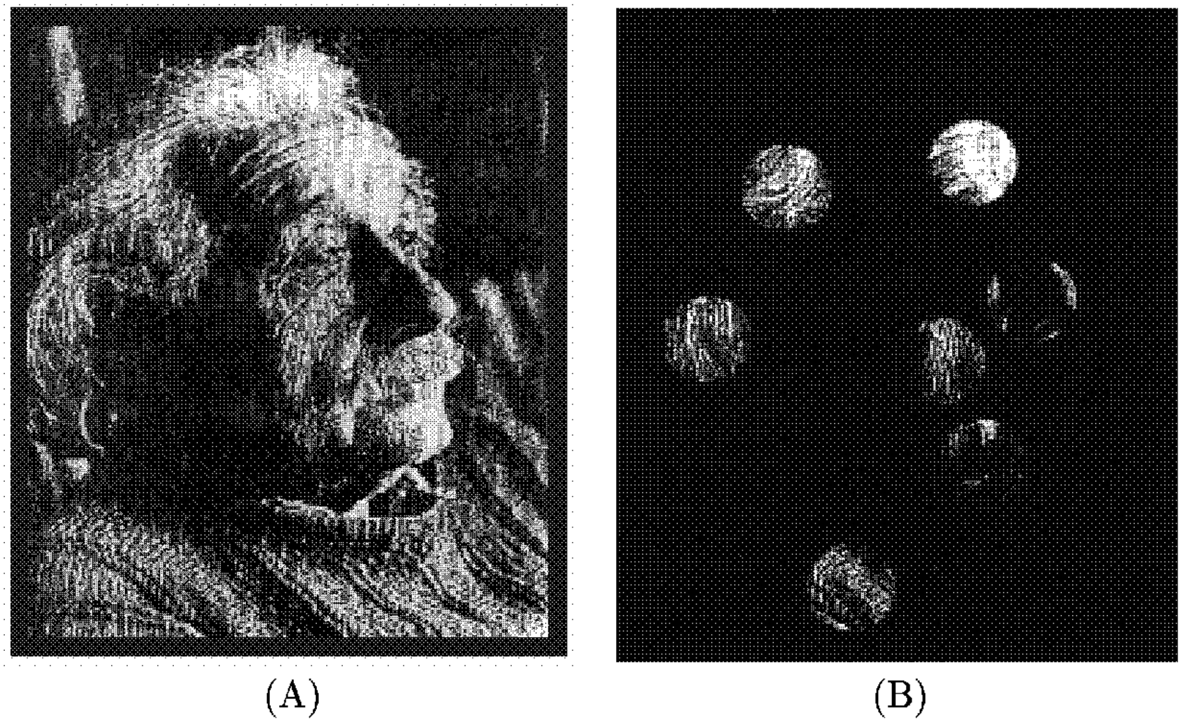

Figure 1: When Einstein's face (A) is seen through small apertures (B), its meaning as a face is greatly degraded.

ears sequentially. To process a pattern of sounds as a whole, it must be "recoded". Such a recoding, or processing stage, is often called a working memory, which stores short-term-memory traces. To identify familiar events, the brain compares short-term traces with stored categories. These categories are accessed using long-term-memory traces, which represent previous experiences that have been acquired through learning.

Somehow, we can rapidly learn new facts–placing them in long-term memory–without being forced just as rapidly to forget others. How does brain processing keep old memories stable and still maintain enough plasticity to learn new things? What I call the stability-plasticity dilemma must be solved by every brain system that attempts to learn about the flood of external signals.

I shall examine several challenging examples of visual and auditory data that suggest how the brain might solve the stability-plasticity dilemma. Each example can be explained by a computational approach called Adaptive Resonance Theory (ART), which I introduced 20 years ago. Despite the diversity of tasks that the brain must complete, the neural circuits that govern many processes seem to rely on a similar set of computational principles. In part, the connection lies in a fundamental characteristic of human brain function: Our perceptions are often matched against our expectations. In many examples where this is so, a similar type of circuit seems to exist.

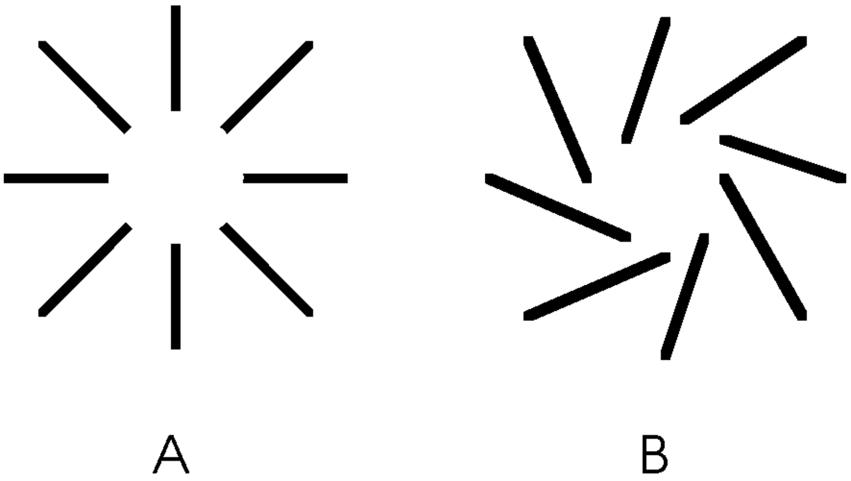

Figure 2: (A) The Ehrenstein pattern generates a circular illusory contour that encloses a circular disk of enhanced illusory brightness. (B) If the endpoints of the Ehrenstein pattern remain fixed while their orientations are tilted, then both the illusory contour and brightness vanish.

1 Sights and Sounds

What we perceive depends on how the nervous system processes a stimulus, such as a photograph or a song. That processing may alter some of the information in a stimulus, leading to a transformed perception of it. By comparing the characteristics of a stimulus and its perception, we may discover some of the principles of brain computation.

An intriguing perception emerges from an image called an Ehrenstein figure (Figure 2), which consists of black lines drawn in a radial pattern on white paper. The mind constructs an illusory circle inside the radiating lines, which makes the figure resemble a child's drawing of the sun–a bright white circular disk with black lines emerging as rays. In fact, the illusory disk appears brighter than the surrounding background. That perception is a collective, emergent property of all the lines that only develops when they are positioned suitably. Why do we see a bright disk that does not exist?

A higher level of visual processing relies on recognition categories, which control learned expectations of what we might see, such as a face or a letter. A key part of recognizing objects depends on how we learn categories that include different instances of a similar object, such as the same letter printed in different sizes or similar shapes. Some tasks–such as recognizing a particular face–require specific categories, and others–such as knowing that every person has a face–depend on comparatively general ones. How do we learn the breadth of a category?

In other situations, we may readily recognize a stimulus despite the fact that many similar stimuli exist simultaneously. You may experience this problem while talking to a friend in a noisy room. You can usually keep track of your conversation above the hubbub, even though sounds emitted by your friend probably overlap with other speakers' sounds. The challenge of separating a single voice from a jumbled mixture of sounds is called the

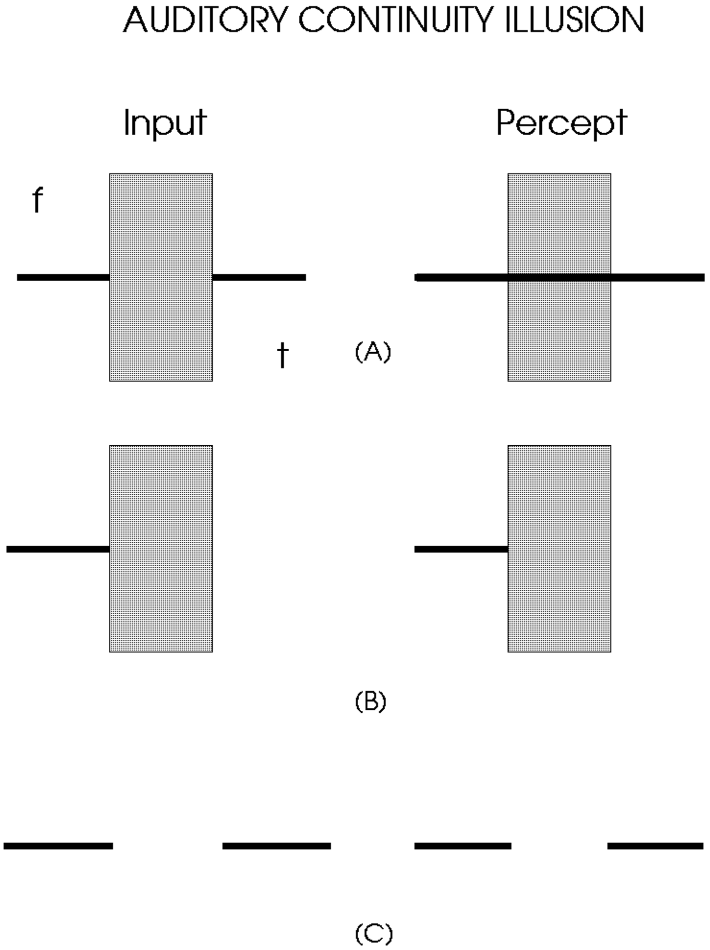

Figure 3: (A) Auditory continuity illusion: When a steady tone occurs both before and after a burst of noise, then under appropriate temporal and amplitude conditions, the tone is perceived to continue through the noise. (B) This does not occur if the noise is not followed by a tone. (C) Nor does it occur if two tones are separated by silence.

cocktail-party problem. How do we separate the voices into distinct sources, or auditory streams?

A simple version of generating streams appears in the auditory-continuity illusion. Suppose that you hear a steady tone that shuts off just as broadband noise turns on, and that the noise shuts off just as the tone turns on again–or tone, noise, tone. Under certain conditions, you will "hear" the tone from start to finish, even during the noise (Figure 3A). Nevertheless, if the tone does not turn on a second time, you will not hear it during the noise (Figure 3B). In your perception, the tone then turns off when the noise starts. If there is no noise between the two tones, then the tone does not bridge the silent interval (Figure 3C). How does your brain know whether there will or will not be a second tone later on, which determines whether it will continue the first tone through the noise? How does it use the noise to construct a tonal perception?

A similar phenomenon exists at a higher level of audition. It is called phonemic restoration. Suppose that you hear a noise followed immediately by the words "eel is on the.... If that string of words is followed by the word "orange," you hear "peel is on the orange." If the word "wagon" completes the sentence, you hear "wheel is on the wagon." If the final word is "shoe," you hear "heel is on the shoe." Richard Warren of the University of Wisconsin, Milwaukee, and his colleagues developed that marvelous example 25 years ago. It shows that a stimulus alone, such as "noise-eel," may not determine what you perceive. How do you hear the sound that you expect to hear, based on previous language experience?

The auditory-continuity illusion and phonemic restoration suggest that the brain can work "backward in time," allowing a later auditory stimulus to determine a perception of an earlier stimulus. I shall argue that these phenomena exist because, in each case, sensory data activate an expectation that focuses attention, much as "peel" emerges from "noise-eel." The attentional focus emerges as part of a "resonance" that leads to a conscious perception. If the resonance has not developed fully before future data arise, then those data can influence the expectations that determine the conscious perception. These auditory phenomena are related to the bright Ehrenstein disk and object recognition through the action of such expectations.

2 Adaptive Resonance Theory

The processing in the brain that generates a perception or recognition event from a stimulus can be investigated with a neural network, which is a computer-based model of neural mechanisms. An ART neural network includes two primary types of memory processes: short-term memory and long-term memory. Short-term memory captures stimuli; long-term memory stores learned information.

In an ART neural network, information can flow from short-term to long-term memory during learning or from long-term to short-term memory during recall (Figure 4A). Each processing layer encodes information in short-term memory via patterns of activation across a network of neurons. Long-term memory is encoded in adaptive weights within the pathways that join neurons in different layers. These weights multiply the signals in the pathways before they are added at their target neurons.

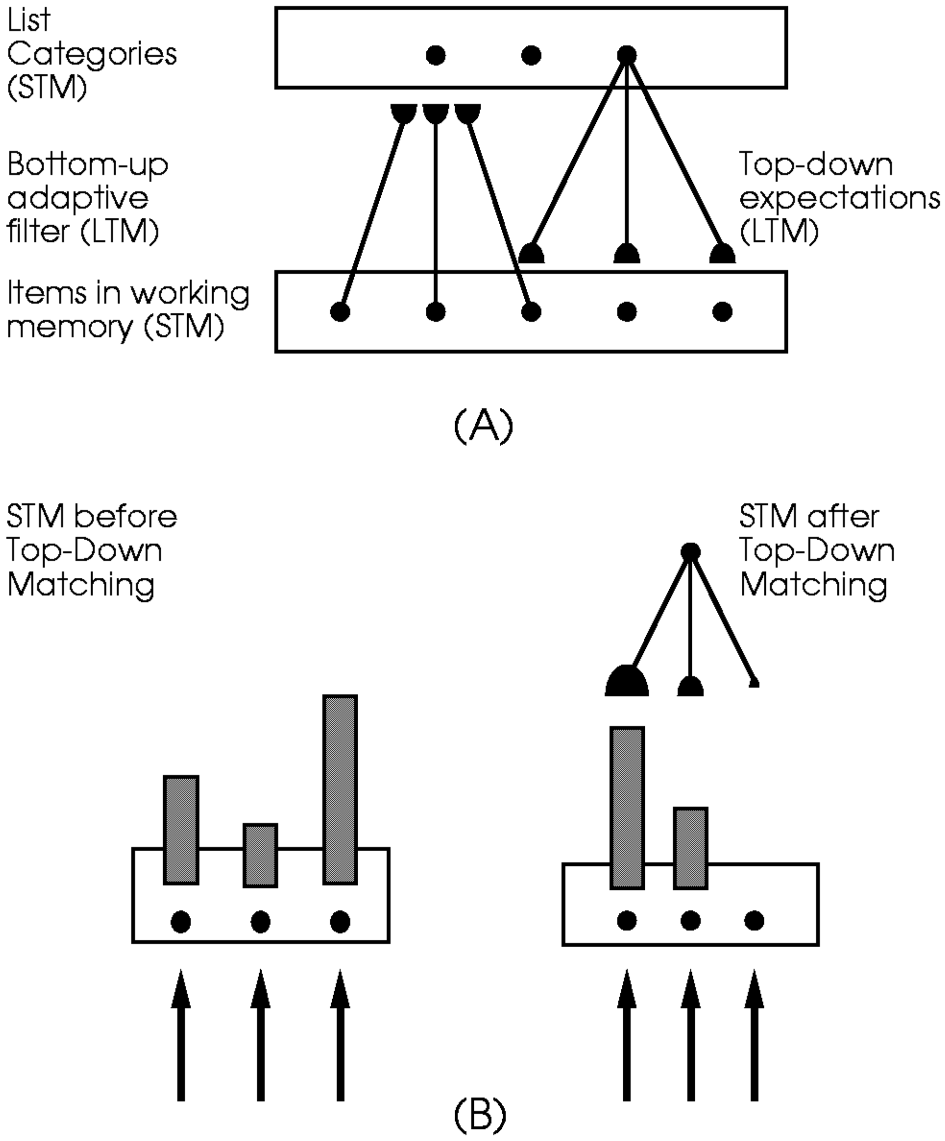

Figure 4: (A) Auditory items activate STM traces in a working memory, which send bottom-up signals towards a level at which list categories, or chunks, are activated in STM. These bottom-up signals are multiplied by learned LTM traces which influence the selection of the list categories that are stored 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. (B) This matching process confirms and amplifies STM activations that are supported by contiguous LTM traces, and suppresses those that are not.

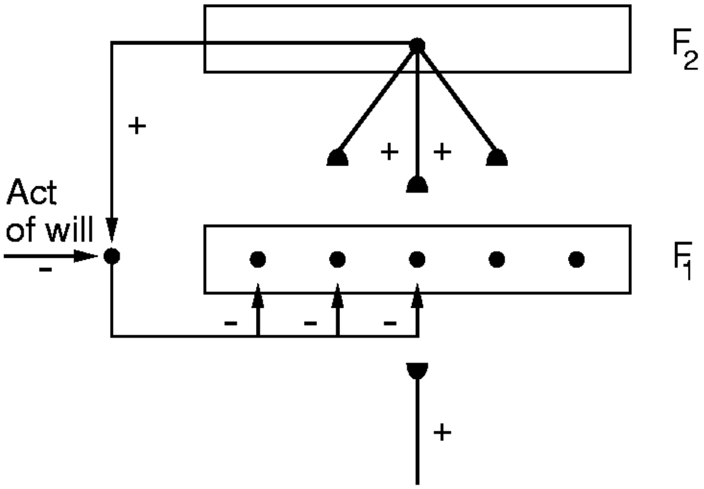

Figure 5: One way to realize the ART matching rule using top-down activation of nonspecific inhibitory interneurons.

In the case of phonemic restoration, feature detectors in short-term memory encode a sound stream. Activating short-term memory generates output signals via bottom-up pathways. The adaptive weights in these pathways help to select category neurons for activation at the next processing level. The category neurons, in turn, generate top-down outputs. The adaptive weights in the top-down pathways encode learned expectations, or prototypes (Figure 4B). These prototypes initiate a matching process that compares sensory inputs with learned expectations, and leads to selections such as “peel” from “noise-eel.”

An ART neural network operates according to specific matching rules (Figure 5). So-called top-down priming means that a top-down expectation, in the absence of any bottom-up input, sensitizes cells that would ordinarily respond to a particular class of stimuli and suppresses others. Conversely, if a cell receives a large enough bottom-up signal, in the absence of any top-down input, then the cell can generate a output, which is called automatic activation. During matching, a cell that receives convergent bottom-up and top-down inputs becomes active. On the other hand, a cell gets suppressed when it receives only a small, or zero, top-down expectation input, even if it receives a large bottom-up input.

Top-down processing selectively amplifies some features of a stimulus and suppresses others, which helps to focus attention on information that matches our expectations. That focusing process helps to filter out some parts of the flood of sensory signals that would otherwise overwhelm us and to prevent them from destabilizing our previously learned memories. Thus top-down expectations may solve the stability-plasticity dilemma by focusing attention and preventing spurious signals from accidentally eroding our previously learned memories.

Nevertheless, if a top-down expectation influences a bottom-up stimulus, what keeps the modified bottom-up signals from reactivating their top-down expectations in a continuing cycle of bottom-up and top-down feedback? Nothing. Once that reciprocal feedback equilibrates, the bottom-up and top-down signals lock the activity patterns in a resonant

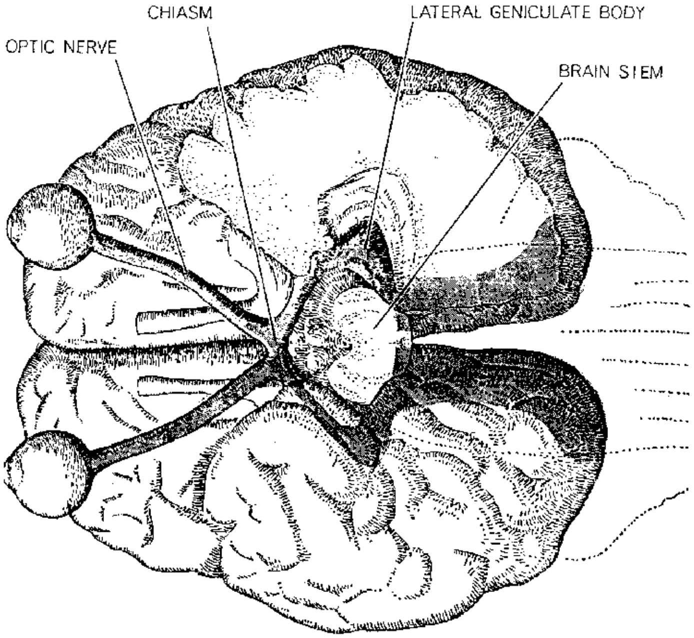

Figure 6: Light registered on the photosensitive retina of the eye is processed by the lateral geniculate nucleus before activating the visual cortex. The visual system appears in this representation of the human brain as viewed from below. Visual pathway from retinas to cortex via the lateral geniculate body is shown in gray.

state. I claim that only resonant states of the brain can achieve consciousness, and that the time needed to develop resonance helps to explain why an event's perception takes so long.

Adaptive resonance theory helps to explain why, as philosophers have asked for many years, humans might be "intentional" beings–planning future behavior and expecting its consequences. An ART neural network achieves its stablity by learning expectations about the world that are continually matched against world data. Those intentions lead to attention. That is, expectations start to focus attention on data worthy of learning.

During phonemic restoration, future evidence (such as the words orange or wagon) can influence the percepts of past sounds (such as the noise) if they occur before the resonance equilibrates. The future words can focus attention on the correct sound ("p" or "wh") in the noise by using top-down expectations that realize the ART matching rule. Only these sounds enter consciousness as the final resonance emerges.

In summary, top-down signals represent the brain's learned expectations of what bottom-up signal patterns should be, based on past experience. A matching process reinforces and amplifies features in a bottom-up pattern that are consistent with top-down expectations, or hypotheses, and suppresses features that are inconsistent. That matching step initiates the process whereby the brain selectively pays attention to experiences that it expects, binds them into coherent internal representations through resonant states and incorporates them in its knowledge about the world.

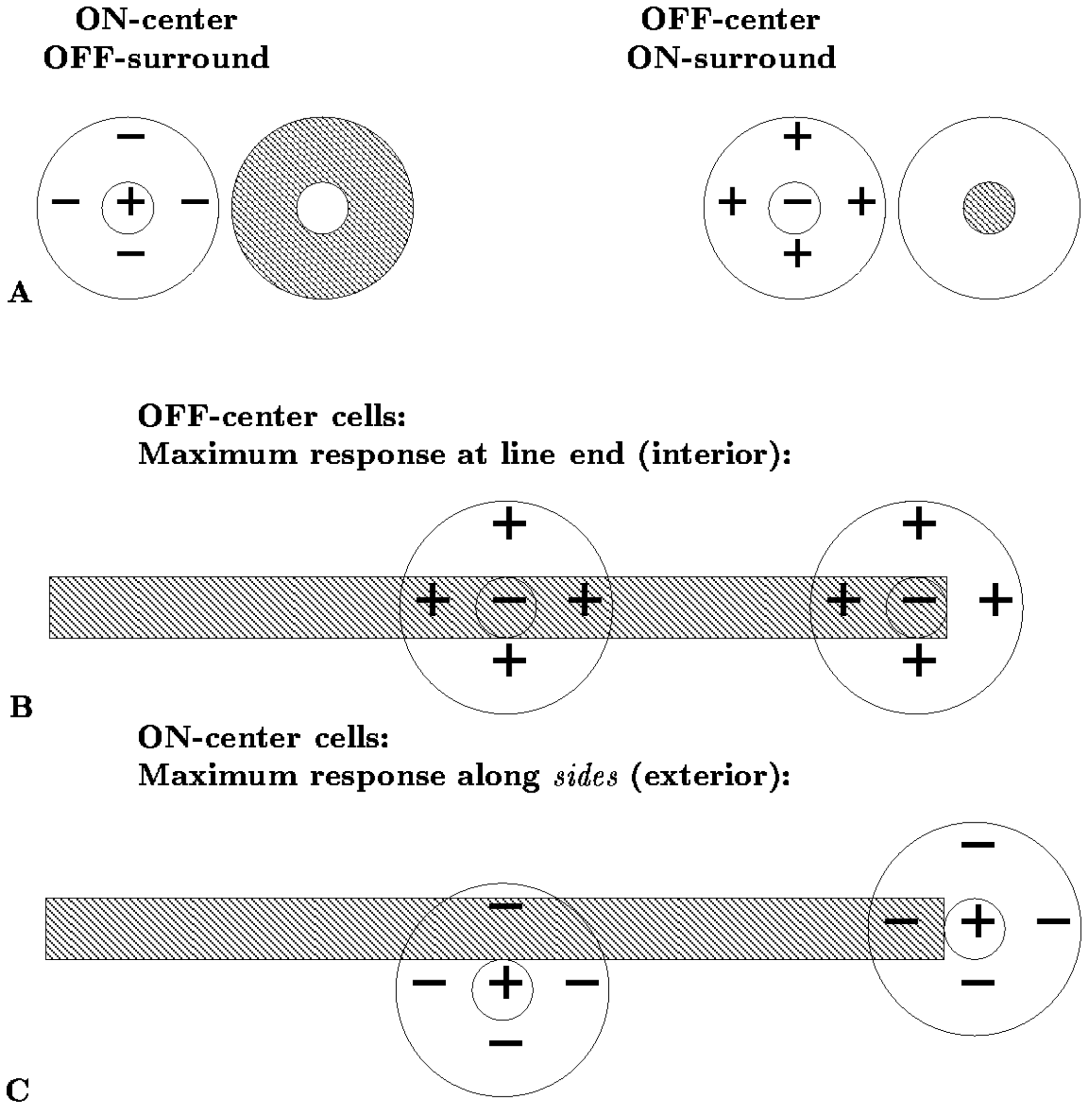
ON-center
OFF-surround
OFF-center
ON-surround
A
OFF-center cells:
Maximum response at line end (interior):
B
ON-center cells:
Maximum response along *sides* (exterior):
C

Figure 7: Retinal center-surround cells and their optimal stimuli (A). The ON cell, on the left, responds best to a high luminance disk surrounded by a low luminance annulus. The OFF cell, on the right, responds best to a low luminance disk surrounded by a high luminance annulus (B). OFF cells respond to the inside of a black line. The OFF cell centered at the line end responds more strongly than the OFF cell centered in the middle, because the surround region of the former cell is closer to optimal. In (C) ON cells respond to the white background just outside the black line. The amount of overlap of each ON cell's surround with the black line affects the strength of the cell's response. As seen in the ON cell's optimal stimulus (C), the more of the surround that is stimulated by a black region, the better the ON cell will respond. Thus, an ON cell centered just outside the side of the line will respond better than a cell centered just outside the end of the line.

3 Brightness Buttons

How does an ART network explain the enhanced brightness of an Ehrenstein disk? John Kennedy of the University of Toronto attempted to explain that perception with "brightness buttons," or proposed bright areas at the ends of dark lines. My colleagues and I proposed that these brightness buttons could, in turn, activate a process of surface filling in whereby the brightness signals diffuse across the visual field until they hit a circular illusory contour. A bright circular disk could thereby be generated.

In the mid-1980s, my colleagues Michael Cohen, Ennio Mingolla and Dejan Todorović and I began to explain many visual percepts with a neural model of how such visual boundaries and surfaces form. In that model, boundaries separate a button-containing region from other parts of a scene. Such a boundary may be generated by image edges, textures or shading, and it may produce illusory contours, such as an Ehrenstein circle. Although our model explained correctly and predicted many facts about illusory contours and perceptions of brightness, it also predicted that the Ehrenstein disk should look darker than its surround, which it does not. My colleagues Alan Gove and Mingolla and I then realized that Ehrenstein disks would look bright if we added a feedback loop from the visual cortex to the lateral geniculate nucleus, a waystation between the retina and the visual cortex (Figure 6).

In 1976 I had predicted that such a feedback loop should exist for quite different reasons. Cells in the visual cortex combine inputs from both eyes to carry out binocular vision. These cells learn their binocular properties at an early stage of development. I predicted that top-down ART matching stabilizes this learning process as it realizes a type of "automatic" attentional processing in the lateral geniculate nucleus. Only recently did my colleagues and I realize that it could also influence percepts of brightness.

The model begins with processing in the lateral geniculate nucleus, where a visual cell's receptive field consists of circular regions. Some of these cells, called ON cells, possess a so-called on-center, off-surround receptive field, which is stimulated by light near the cell's location (the on-center) and inhibited by light in more distant locations (the off-surround). A so-called OFF cell has an off-center, on-surround receptive field, which is inhibited by light near the cell's location and stimulated by more distant light (Figure 7A).

These cells could produce contrast between a black line and a white background. An OFF cell positioned with its receptive center inside a black line will be activated (Figure 7B). Furthermore, an OFF cell near a line's end will be even more strongly activated, because more of its surround lies in a white background. An ON cell, on the other hand, gets stimulated when its center lies outside of a black line (Figure 7C). An ON cell positioned with its center just beyond the side of the line, but not at an end, will respond most strongly. The ON cells, therefore, enhance brightness along the sides of a black line and OFF cells enhance the darkness just inside the ends of a black line. In other words, these cells alone do not produce brightness buttons. They could make Ehrenstein disks look dark.

Feedback from the visual cortex is sensitive to a line's orientation and is more active near a line end, because of a process called endstopping (Figure 8B). Feedback could enhance the contrast at the ends of a black line and reduce it along the sides, thereby causing cells in the lateral geniculate nucleus to make brightness buttons (Figures 8C and

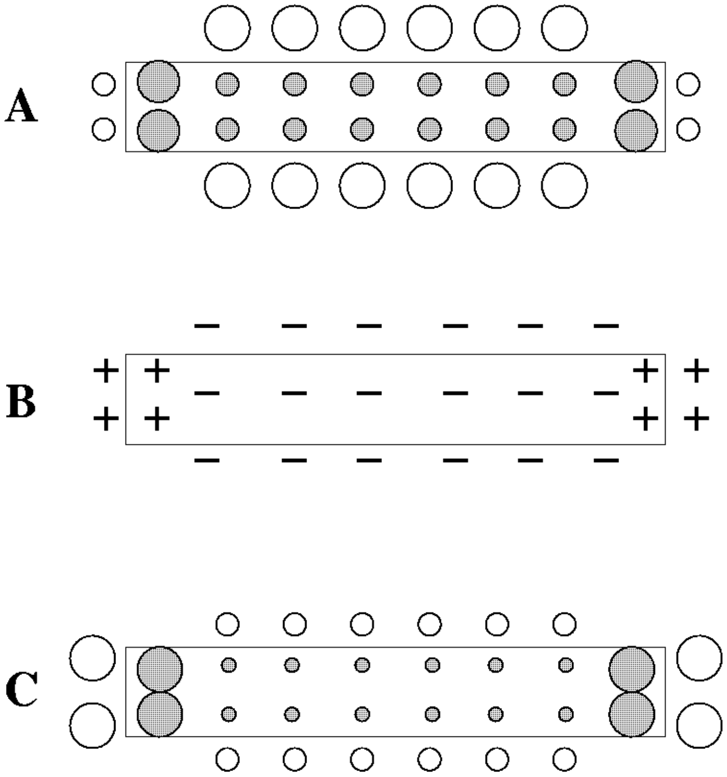

Figure 8: Schematic diagram of brightness button formation in the model. In (A) the distribution of model LGN cell activities prior to receiving any feedback, in response to a black bar is illustrated. Open circles code ON cell activity; filled circles code OFF cell activity. (B) shows the effect of feedback in bottom-up LGN activations. (C) shows the LGN activity distribution after feedback. A brightness button is formed outside both ends of the line.

9B). The model's boundary-completion network then "connects" neighboring line ends within its "cortex," thereby generating a circular illusory contour inside the lines (Figure 9C). Next, a diffusion process responds to the brightness buttons to fill in a uniform level of enhanced brightness within the bounding illusory contour. The result is an Ehrenstein disk with uniformly enhanced brightness relative to its surround (Figure 9D).

Direct experimental evidence demonstrates that cortical feedback can alter the properties of lateral-geniculate- nucleus cells as proposed. In 1987 Adam Sillito and his colleagues at University College, London, showed that cortical feedback in a cat tunes cells in its lateral geniculate nucleus to respond best to lines with a specific length. In 1986 Chris Redies of MPI Entwicklungsbiologie, Germany, and his colleagues found that some visual cells in a cat's lateral geniculate nucleus and cortex respond best at line ends. In other words, the cells respond more strongly to line ends than line sides, just like our model does. In addition, a remarkable 1994 article in Nature by Sillito and his colleagues provides neurophysiological data that suggest that feedback from the cortex to the lateral geniculate nucleus resembles the matching and resonance of an ART network. They found that cortical feedback changes the output of specific lateral-geniculate-nucleus cells, thereby increasing "the gain of the input for feature- linked events detected by the cortex ... the cortico- thalamic input is only strong enough to exert an effect on those dLGN [dorsal lateral geniculate nucleus] cells that are additionally polarized by their retinal input ... the feedback circuit searches for correlations that support the 'hypothesis' represented by a particular pattern of cortical activity.

4 Making Matches and Testing Hypotheses

Feedback in an ART network also focuses attention in models of visual object recognition. Here attention can be controlled flexibly in a task-sensitive way (Figure 10). Such an ART network consists, as before, of an attentional subsystem that learns to form categories and expectations in response to sensory inputs. In addition, there is an orienting subsystem that is activated by novel events and enables the attentional subsystem to learn about them in a stable way (Figure 10). In other words, interacting attentional and orienting subsystems permit an ART system to solve the stability-plasticity dilemma in response to large amounts of sensory data.

Processing begins when a sensory input stimulates short-term memory, in the attentional subsystem, and the orienting subsystem. The short-term memory contains a network of nodes, or cell populations, each of which is activated by a particular combination of features in an input. An input pattern gets registered as a short-term-memory activation pattern, which stimulates the long-term memory through bottom-up processing and inhibits the orienting subsystem (Figure 11A). As long as the short-term-memory pattern resembles the sensory input, the orienting subsystem remains idle, because of a balance of excitation from the input and inhibition from short-term memory. The long-term memory traces in the bottom-up pathways activate another network of nodes that represents recognition codes, or categories. In other words, the short-term-memory pattern at the first level uses its best match in long-term memory to activate a category at the second level that represents the sensory input.

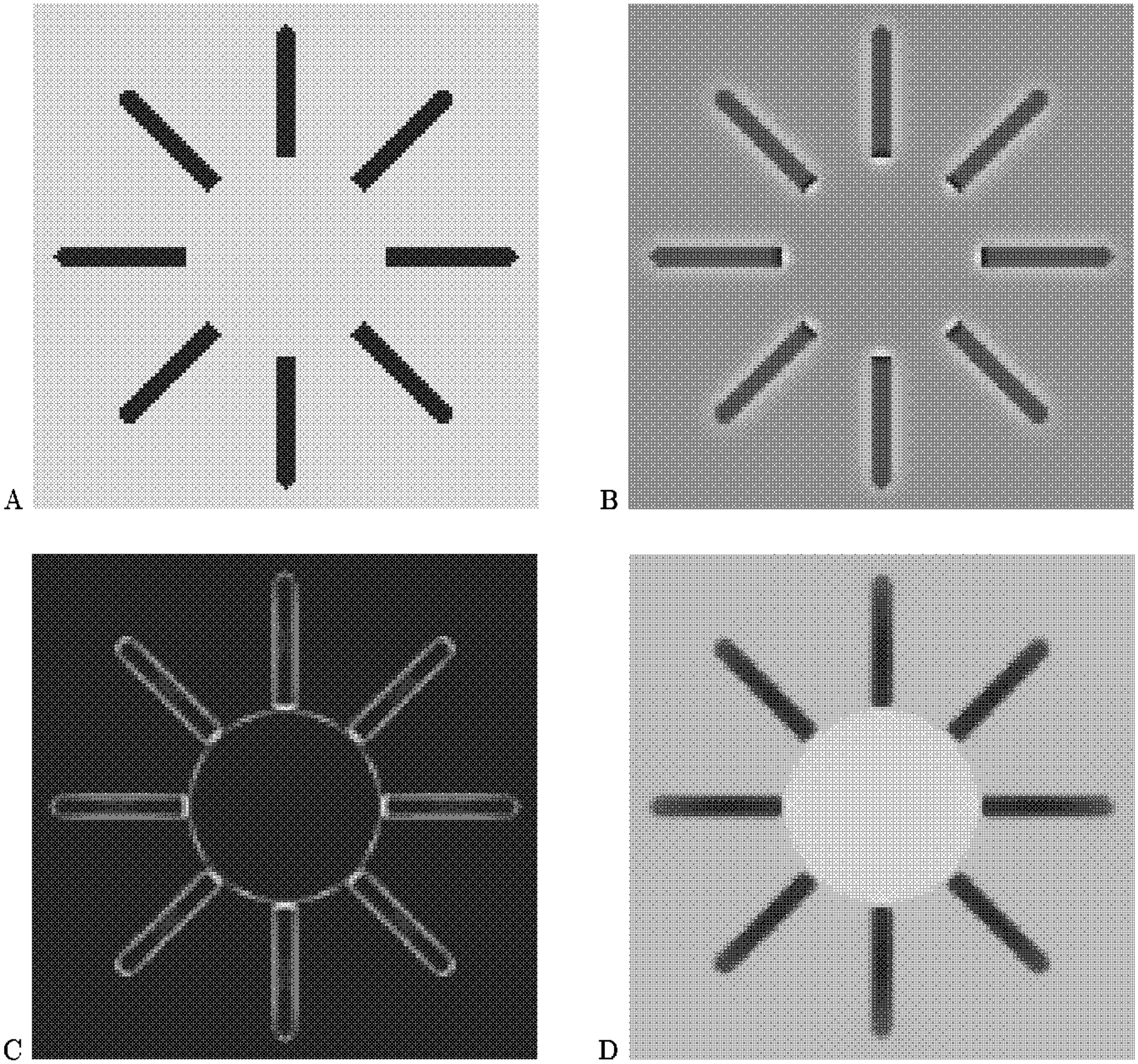

Figure 9: (A) The Ehrenstein figure. (B) The LGN stage response. Both ON and OFF cell activities are coded as rectified deflections from a neutral gray. Note the brightness buttons at the line ends. (C) The equilibrium boundaries. (D) In the filled-in surface brightness, the central disk contains larger activities than the background, corresponding to the perception of increased brightness.

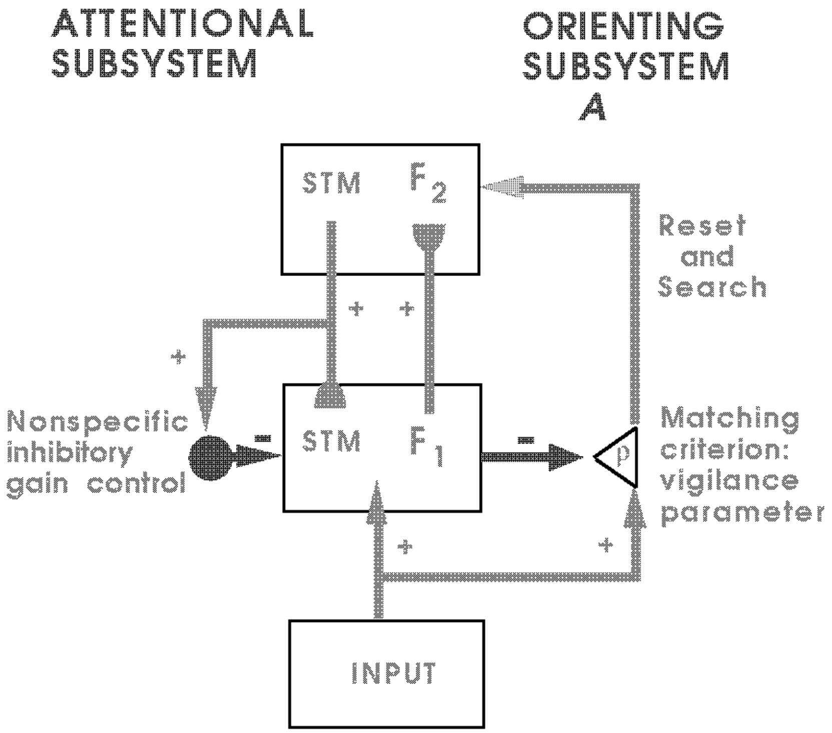

Figure 10: An example of a model ART circuit in which attentional and orienting circuits interact. Level $\mathcal{F}_1$ encodes a distributed representation of an event by a short term memory (STM) activation pattern across a network of feature detectors. Level $\mathcal{F}_2$ encodes the event using a compressed STM representation of the $\mathcal{F}_1$ pattern. Learning of these recognition codes occurs at the long term memory (LTM) traces within the bottom-up and top-down pathways between levels $\mathcal{F}_1$ and $\mathcal{F}_2$. The top-down pathways read-out learned expectations whose prototypes are matched against bottom-up input patterns at $\mathcal{F}_1$. The size of mismatches in response to novel events are evaluated relative to the vigilance parameter ρ of the orienting subsystem $\mathcal{A}$. A large enough mismatch resets the recognition code that is active in STM at $\mathcal{F}_2$ and initiates a memory search for a more appropriate recognition code. Output from subsystem $\mathcal{A}$ can also trigger an orienting response. (A) Block diagram of circuit. (B) Individual pathways of circuit, including the input level $\mathcal{F}_0$ that generates inputs to level $\mathcal{F}_1$. The gain control input g_1 to level $\mathcal{F}_1$ helps to instantiate the matching rule (see text). Gain control g_2 to level $\mathcal{F}_2$ is needed to instate a category in STM.

Activating such a category may be interpreted as "making a hypothesis" about an input. The "winning" category elicits an output that excites the feature detectors through top-down processing. The top-down signal thereby plays the role of a learned expectation, and activating that expectation may be interpreted as "testing the hypothesis."

The top-down signals also activate an attentional-gain-control channel that nonspecifically inhibits all of the feature detectors (Figure 11B). Unless a feature detector receives a large, excitatory, learned-expectation signal, it is shut off by this inhibitory signal. This produces a modified short-term-memory pattern, which encodes only the input features that the network deems relevant to the hypothesis based on its past experience. From then on, the network "pays attention to" the modified short-term- memory pattern. If the modified short-term-memory pattern closely resembles the original sensory input, the orienting subsystem still produces no output. In addition, the modified short-term-memory feature pattern reactivates the category via bottom-up signals, which reactivates the short-term-memory pattern via top-down signals and so on. A resonance hereby develops that binds spatially distributed features into either a stable equilibrium or a synchronous oscillation, much like Reinhard Eckhorn of the Phillips-University, Germany, Wolf Singer of the Max Planck Institute for Brain Research in Frankfurt and their colleagues have shown in experiments on the visual cortex.

How does the network react when there is a mismatch? In that case, the top-down expectation, or prototype, does not match the short-term-memory feature pattern. Then the attentional subsystem produces a modified short- term-memory feature pattern that includes only the few, if any, parts that match the top-down expectation. The significantly modified short-term-memory pattern weakens or removes the inhibition on the orienting subsystem, allowing it to turn on (Figure 11C). Then, the orienting subsystem sends a "reset" signal that clears activity in the category level and the subsequent levels that it feeds. Finally, the network retrieves its original short-term-memory feature pattern and initiates a memory search for a better category (Figure 11D). That cycle continues until a match surfaces or a new category is selected to learn about a novel situation.

5 Novelty and Generalization

Whether or not resonance develops depends on the level of mismatch, or novelty, that a network tolerates. Novelty is a measurement of how well a given short- term-memory feature pattern matches the expectation read out by the category that it evokes. The criterion of an acceptable match is defined by an internally controlled parameter that my colleague, Gail Carpenter, and I call vigilance, which the orienting subsystem computes. Vigilance weighs how similar a short-term- memory pattern must be to a prototype in order to generate resonance.

Either a higher level of vigilance or a lower level of matching can prevent resonance. If the orienting subsystem sends a reset signal, a new bout of hypothesis testing, or memory search, begins. During a search, the orienting subsystem interacts with the attentional subsystem to rapidly reset mismatched categories and to select a different representation with which to categorize novel events, without risking unselective forgetting of previous knowledge. A search may produce a familiar category that is similar enough to the input

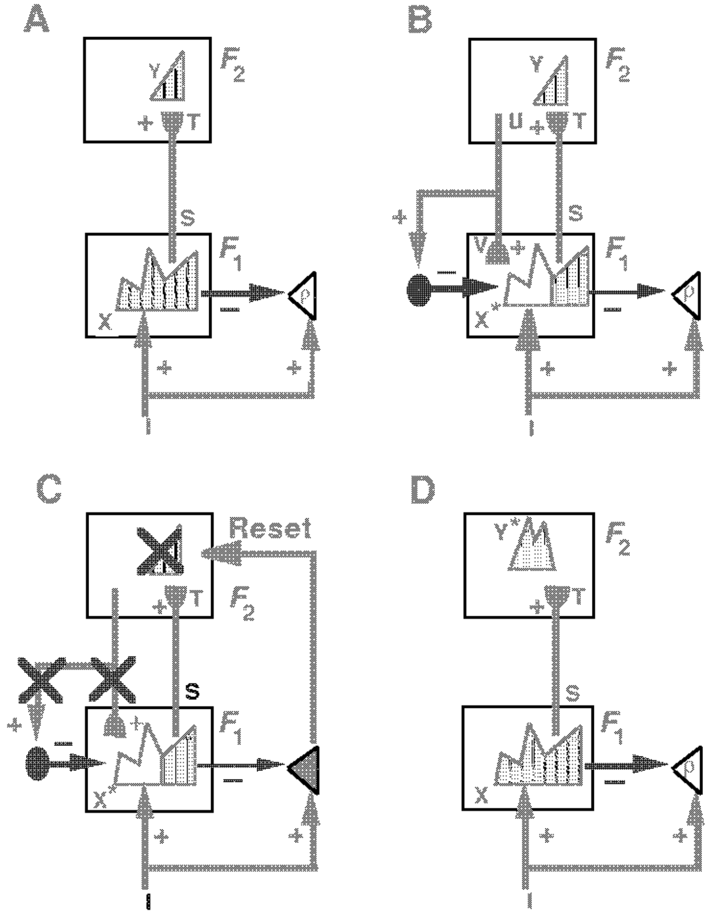

Figure 11: ART search for a recognition code: (A) The input pattern **I** is instated across the feature detectors at level $\mathcal{F}_1$ as a short term memory (STM) activity pattern **X**. Input **I** also nonspecifically activates the orienting subsystem $\mathcal{A}$. STM pattern **X** is represented by the hatched pattern across $\mathcal{F}_1$. Pattern **X** both inhibits $\mathcal{A}$ and generates the output pattern **S**. Pattern **S** is multiplied by long term memory (LTM) traces and added at $\mathcal{F}_2$ nodes to form the input pattern **T**, which activates the STM pattern **Y** across the recognition categories coded at level $\mathcal{F}_2$. (B) Pattern **Y** generates the top-down output pattern **U** which is multiplied by top-down LTM traces and added at $\mathcal{F}_1$ nodes to form the prototype pattern **V** that encodes the learned expectation of the active $\mathcal{F}_2$ nodes. If **V** mismatches **I** at $\mathcal{F}_1$, then a new STM activity pattern $\mathbf{X}^*$ is generated at $\mathcal{F}_1$. $\mathbf{X}^*$ 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 $\mathbf{X}^*$ causes a decrease in the total inhibition from $\mathcal{F}_1$ to $\mathcal{A}$. (C) If inhibition decreases sufficiently, $\mathcal{A}$ releases a nonspecific arousal wave to $\mathcal{F}_2$, which resets the STM pattern **Y** at $\mathcal{F}_2$. (D) After **Y** is inhibited, its top-down prototype signal is eliminated, and **X** can be reinstated at $\mathcal{F}_1$. Enduring traces of the prior reset lead **X** to activate a different STM pattern $\mathbf{Y}^*$ at $\mathcal{F}_2$. If the top-down prototype due to $\mathbf{Y}^*$ also mismatches **I** at $\mathcal{F}_1$, then the search for an appropriate $\mathcal{F}_2$ code continues until a more appropriate $\mathcal{F}_2$ representation is selected. Then an attentive resonance develops and learning of the attended data is initiated.

to satisfy the resonance criterion. The representation in short-term memory may then be refined by attentional focusing and learned by the active prototype. If the input is too different from any previously learned category, then an uncommitted category is selected to learn about the new data.

Vigilance can vary across learning trials. Thus recognition categories capable of encoding widely differing degrees of generalization or abstraction can be learned by a single ART network. Low vigilance leads to broad generalization, or abstract categories. High vigilance leads to narrow generalization, or more specific categories. In other words, a single ART network can employ abstract categories, such as knowing that everyone has a face, and more specific categories, such as recognizing an individual face, by simply adjusting its vigilance.

As sequences of inputs are practiced over learning trials, the search process eventually converges on stable categories. In 1987 Carpenter and I proved mathematically that familiar inputs directly access the category that provides the globally best match, and unfamiliar inputs engage the orienting subsystem to trigger memory searches for better categories until one gets selected.

The attentional subsystem of an ART network has been used to model aspects of inferotemporal cortex, a higher visual center, and the orienting subsystem can model part of the hippocampal system, which contributes to memory functions. The interpretation of ART dynamics in terms of inferotemporal cortex led Bob Desimone and his colleagues at the National Institutes of Mental Health to successfully test the prediction that cells in monkey inferotemporal cortex are reset after each trial in a working-memory task. To illustrate the implications of an ART interpretation of inferotemporal-hippocampal interactions, I shall review how disconnecting an ART models orienting subsystem creates a memory disorder with symptoms much like the medial-temporal amnesia that is caused in animals and humans after hippocampal- system lesions. Such lesions induce many symptoms, including unlimited anterograde amnesia (inability to remember events subsequent to the lesion), limited retrograde amnesia (ability to remember remote but not recent events) and abnormal reactions to novelty.

Disconnecting an ART network's orienting subsystem generates similar problems. Unlimited anterograde amnesia, for instance, develops because the network cannot carry out a memory search to learn a new recognition category. Limited retrograde amnesia arises because familiar events can directly access the correct recognition codes, but consolidating a new memory requires the orienting subsystem, using its sensitivity to novel events.

Similar behavioral problems have been identified in animals that lack a functional hippocampal system. David Gaffan of Oxford University noted that transecting a monkey's fornix, which carries information leaving the hippocampal system, "impairs ability to change an established habit ... in a different set of circumstances that is similar to the first and therefore liable to be confused with it." Likewise, a rat with a destroyed hippocampal system has difficulty orienting to novel cues. An ART network with a defective orienting subsystem responds similarly, because it cannot trigger a memory search to learn different representations for similar events or entirely new categories for novel stimuli. During normal learning, an ART network's orienting subsystem disengages automatically as events become familiar during the memory-consolidation process, which is consistent with the progressive reduction in novelty- related hippocampal potentials that normal rats develop during learning. These correlations between experimental results and an ART network illustrate how–as Stuart Zola-Morgan and Larry Squire of the University of California at San Diego have reported– memory consolidation and novelty detection may be mediated by the same neural structures.

6 Streams of Sound

The same ART principles also help to explain many auditory phenomena, such as variable-rate speech perception. Consider how people hear combinations of vowels and consonants in vowel-consonant consonant-vowel sequences. Bruno Repp at Haskins Laboratories has studied perception of the sequences [ib]-[ga] and [ib]-[ba] by varying the silent interval between the initial vowel-consonant syllable and the terminal consonant-vowel syllable. If the silence interval is short enough, [ib]-[ga] sounds like [iga] and [ib]-[ba] sounds like [iba]. Repp showed that the transition from perceiving [iba] to [ib]-[ba] requires from 100 to 150 milliseconds more silence than the transition from [iga] to [ib]-[ga], which is very long compared with the time needed to activate neurons (Figure 12). Why is this shift so large?

My colleagues Ian Boardman and Michael Cohen and I simulated these data with our ARTPHONE model. That model reveals how a resonant wave develops as a result of bottom-up and top-down signal exchanges between a short-term memory, which represents lists of individual speech items stored in a working memory, and a list categorization network that groups them together into learned language units, or chunks. The model suggests that a short silence between [ib] and [ga] produces a mismatch between [g] and [b], which rapidly resets the working memory, thereby preventing the [b] sound from reaching resonance and consciousness (Figure 13B). The syllables [ib]-[ba] generate a resonance from the first [b] that fuses with the subsequent resonance from the second [b]. This sounds like a single [b] and thereby greatly extends the perceived duration of [iba] across a silence interval (Figure 14A).

Nevertheless, if [ib] can fuse across time with [ba], how do we ever hear distinct [ib]-[ba] sounds when the silence gets long enough? After a resonance develops fully, it eventually collapses spontaneously because of habituation that goes on in the pathways that maintain the resonance via bottom-up and top-down signals. Thus, if the silence is long enough for resonant collapse of [ib], then a distinguishable [ba] resonance can develop subsequently and be heard (Figure 14B).

A similar type of resonant processing, at an earlier level of auditory processing, helps to explain cocktail-party separation of distinct voices into auditory streams. My colleagues Krishna Govindarajan, Lonce Wyse and Michael Cohen and I developed the ARTSTREAM model, which suggests how distinguishable auditory streams can be formed and separated (Figure 15). Here I shall concentrate on separating sounds of different frequencies (Figure 16), but more complex models use similar mechanisms to further separate different sounds based on additional characteristics, such as their location.

The ARTSTREAM model consists of two main processing levels: a spectral-stream level and a pitch-stream level. The incoming auditory signal gets preprocessed by the

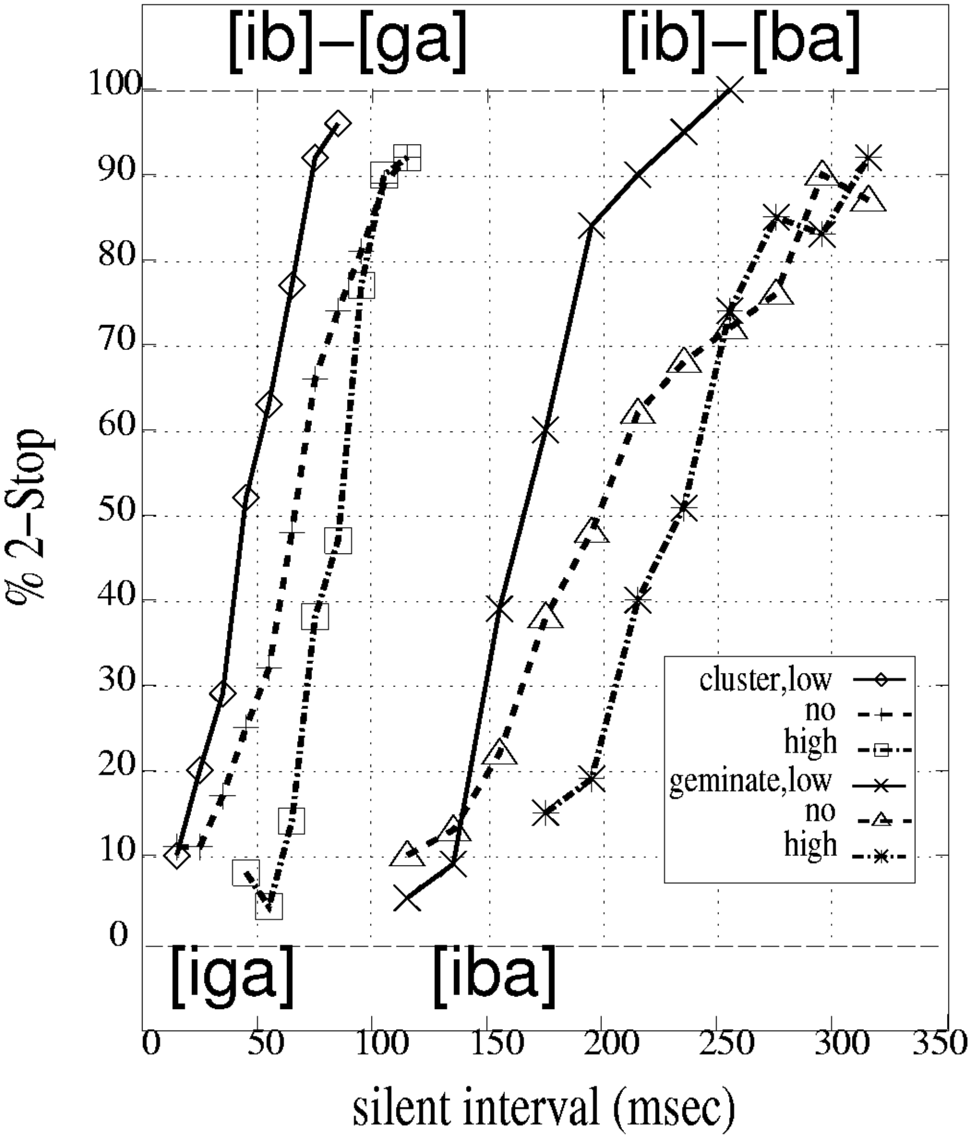

Figure 12: The left-hand curves represent the probability, under several experimental conditions, that the subject will hear [ib]–[ga] rather than [iga]. The right-hand curves do the same for [ib]–[ba] rather than the fused percept [iba]. Note that the perception of [iba] can occur at a silence interval between [ib] and [ba] that is up to 150 milliseconds longer than the one that leads to the percept [iga] instead of [ib]–[ga]. (Data are reprinted with permission from B.H. Repp (1980), Haskins Laboratories Status Report on Speech Research, **SR-61**, 151–165.)

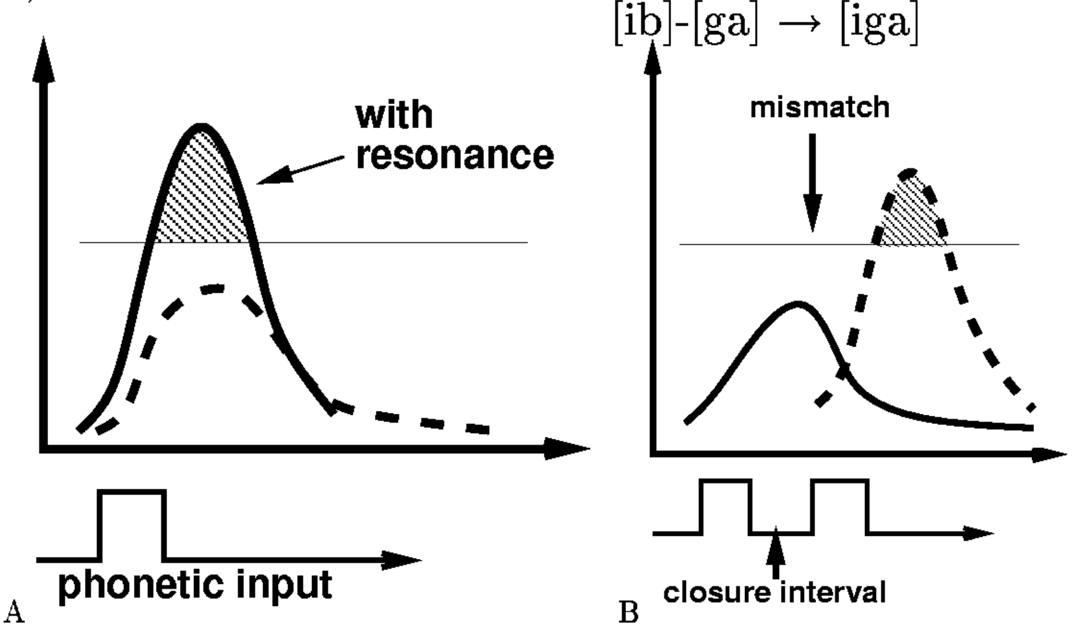

Figure 13: (A) Response to a single stop, such as [b] or [g], with and without resonance. Suprathreshold, resonant, activation is shaded. (B) Reset due to phonologic mismatch between [ib] and [ga]. (Reprinted from Grossberg, S., Boardman, I., and Cohen, M.A. (1995). Neural dynamics of variable-rate speech categorization. Technical Report CAS/CNS-TR-94-038, Boston, MA: Boston University. *Journal of Experimental Psychology: Human Perception and Performance*, in press.)

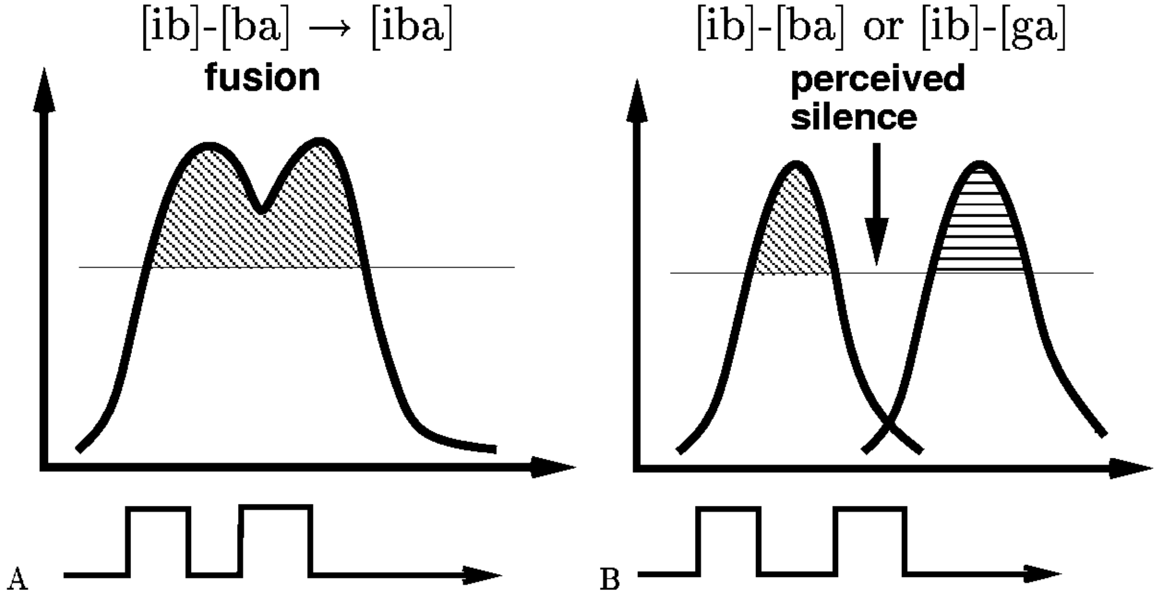

Figure 14: (A) Fusion in response to proximal similar phones. (B) Perceptual silence allows a 2-stop percept. (Reprinted from Grossberg, S., Boardman, I., and Cohen, M.A. (1995). Neural dynamics of variable-rate speech categorization. Technical Report CAS/CNS-TR-94-038. Boston, MA: Boston University. *Journal of Experimental Psychology: Human Perception and Performance*, in press.)

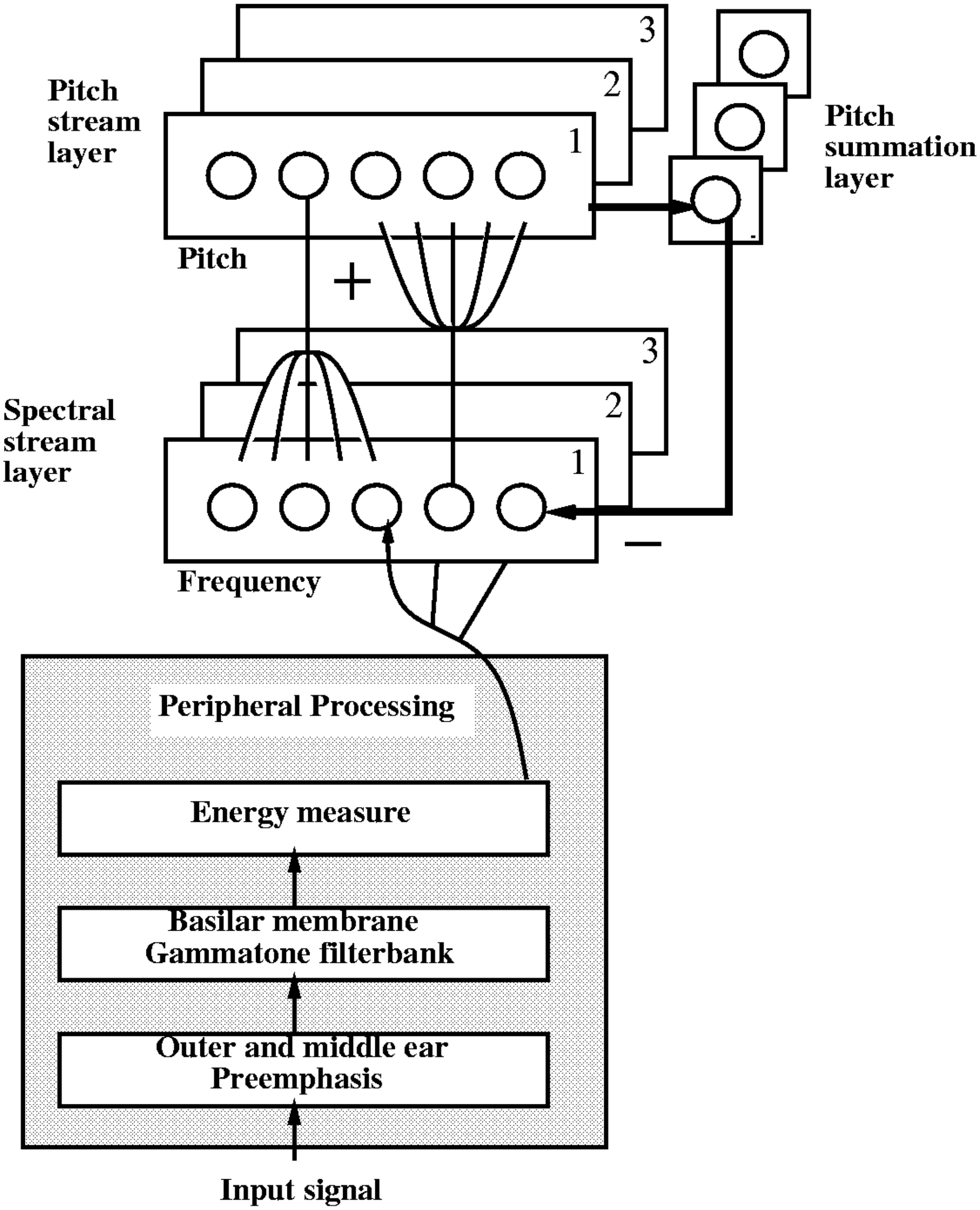

Figure 15: Block diagram of the ARTSTREAM auditory streaming model. Note the nonspecific top-down inhibitory signals from the pitch level to the spectral level that realize ART matching within the network. (Reprinted from Govindarajan, K.K., Grossberg, S., Wyse, L.L., and Cohen, M.A. (1994). A neural network model of auditory scene analysis and source segregation. Technical Report: CAS/CNS-TR-94-039. Boston, MA: Boston University.)

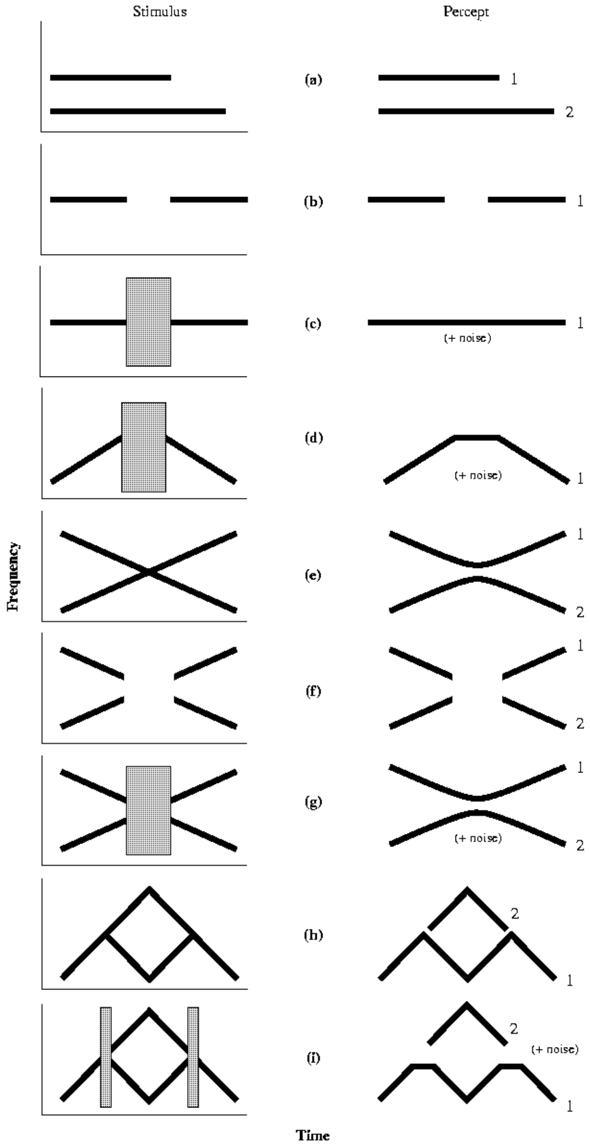

Figure 16: Illustrative stimuli and the listeners' percepts that ARTSTREAM model simulations emulate. The hashed boxes represent broadband noise. The stimuli consist of: (A) two inharmonic tones, (B) tone-silence-tone, (C) tone-noise-tone, (D) a ramp or glide-noise-glide, (E) crossing glides, (F) crossing glides where the intersection point has been replaced by silence, (G) crossing glides where the intersection point has been replaced by noise, (H) Steiger diamond stimulus, and (I) Steiger diamond stimulus where bifurcation points have been replaced by noise. (Reprinted from Govindarajan, K.K., Grossberg, S., Wyse, L.L., and Cohen, M.A. (1994). A neural network model of auditory scene analysis and source segregation. Technical Report: CAS/CNS-TR-94-039. Boston, MA: Boston University.)

ear's mechanical and neurophysiological filters, which divide sounds into groups of similar frequencies. The spectral, or frequency, components of a sound serve as input for multiple spectral-stream layers. The spectral- stream cells convert the incoming signal to a spatial map of frequencies. You might imagine that high frequencies stimulate cells at one end of a spectral-stream layer, low frequencies stimulate cells at the other end and intermediate frequencies stimulate cells in the middle of the layer. So a specific sound activates a specific pattern of cells.

Each spectral-stream layer passes a bottom-up signal to its pitch-stream layer. Between layers, the bottom-up pathways act like a type of harmonic sieve that filters the spectrum so that only allow certain harmonically related frequencies can pass. The filtered bottom-up signals activate multiple representations of a sounds pitch at the pitch-stream level. These pitch representations compete to select a single winning node, which becomes active. That node inhibits the redundant representations in other pitch streams, and sends top-down matching signals back to its spectral-stream level to excite spectral nodes whose frequencies are consistent with the selected pitch.

Now, a bottom-up signal alone fails to activate spectral nodes in the absence of an excitatory top-down signal because the active pitch node also stimulates a pitch-summation layer, which inhibits all nodes in its spectral-stream layer. Only a spectral-stream node that receives simultaneous bottom-up and top-down signals becomes fully activated. All other nodes in that spectral stream are inhibited, including spectral nodes that were previously activated by bottom-up signals but received no subsequent top-down pitch support. In other words, the frequency components that are consistent with the winning pitch node are amplified, and all others are suppressed, thereby leading to a spectral-pitch resonance within the stream of the winning pitch node.

In this way, the pitch layer binds together the frequency components that correspond to a prescribed auditory source. All the frequency components that are suppressed in this stream are freed to activate and resonate with a different pitch in a different stream. The net result is multiple resonances, each selectively grouping together the frequencies that correspond to a distinct auditory source.

Using the ARTSTREAM model, we have simulated many of the basic streaming perceptions, including the auditory-continuity illusion. It exists, I contend, because the spectral-stream resonance takes a length of time to develop that is commensurate to the duration of the subsequent noise. Once the tone resonance develops, the second tone can act quickly to support and maintain it throughout the duration of the noise, much as [ba] fuses with [ib] during perception of [iba].

7 Are ART Processes Universal?

In all of the examples discussed above–early vision, visual object recognition, auditory streaming and speech recognition–ART matching and resonance play a central role in models that help explain how the brain stabilizes its learned adaptations in response to changing environmental conditions. That type of matching can be achieved using a top-down, nonspecific inhibitory gain control that inhibits all target cells except those that also receive specific, excitatory top-down signals. Other brain processes also seem to utilize these mechanisms.

My colleagues Mario Aguilar, Dan Bullock and Karen Roberts and I have developed a model that uses ART principles to explain how the superior colliculus uses visual, auditory and planned-movement signals to control the fast eye movements, called saccades, whereby we rapidly look at new objects. That model explains behavioral and neural data about multimodal eye- movement control in terms of how the brain learns a map wherein visual, auditory and planned-movement commands can be represented consistently and compete for attention until a prescribed target location is selected.

Recent experiments from Marcus Raichle's lab at Washington University, using positron emission tomography (PET), support the idea that ART top-down priming also operates in human somatosensory cortex. In their experiments, attending to an impending stimulus to the fingers caused inhibition of nearby cortical cells that code for the face, but not of cells that code for the fingers. Likewise, priming of the toes produced inhibition of nearby cells that code for the fingers and face, but not of cells that code for the toes. Again, it appears that a combination of top-down, specific-excitatory and nonspecific-inhibitory signaling is at work. Thus early vision, visual object recognition, auditory streaming, speech recognition, eye-movement control and somatosensory representation may all incorporate variants of ART networks. These results suggest that a type of "automatic" attention operates even at early levels of brain processing, such as the lateral geniculate nucleus, but that some higher levels benefit from an orienting subsystem that can be used to flexibly reset attention and to facilitate voluntary control of top-down expectations.

Given this type of circuit, how could top-down priming be released from inhibition to enable us to voluntarily experience internal thinking and fantasies? That could be achieved through an "act of will" that activates cells that turn off top-down inhibition (Figure 5), such as that generated by the pitch-summation layer, thereby allowing cells to turn on when they receive top-down signals. These cells would then be free to generate self-initiated resonances.

Thus we arrive at an emerging picture of how the adaptive brain works, wherein issues of stability and plasticity are joined with properties of attention, intention, thinking, fantasy and consciousness. The mediating events are adaptive resonances that affect a dynamic balance between the complementary demands of stability and plasticity, and of expectation and novelty, whose maintenance throughout life in a changing world is one of the core challenges that we face in trying to live our lives fully and well.

References

Bregman, A. S. (1990). **Auditory Scene Analysis: The Perceptual Organization of Sound**. Cambridge, MA: MIT Press.

Carpenter, G. A. and Grossberg, S. (1994) (Eds.). **Pattern Recognition by Self-Organizing Neural Networks**. Cambridge, MA: MIT Press.

Carpenter, G. A. and Grossberg, S. (1993). Normal and amnesic learning, recognition, and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences* **16**, 131–137.

Desimone, R. (1992). Neural circuits for visual attention in the primate brain. In G. A Carpenter and S. Grossberg (Eds.), **Neural Networks for Vision and Image Processing**. Cambridge, MA: MIT Press.

Gove, A., S. Grossberg and E. Mingolla. (1995). Brightness perception, illusory contours, and corticogeniculate feedback. *Visual Neuroscience*, **12**, 1027–1052.

Govindarajan, K.K., Grossberg, S., Wyse, L.L., and Cohen, M.A. (1994). A neural network model of auditory scene analysis and source segregation. Technical Report: CAS/CNS-TR-94-039. Boston, MA: Boston University.

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics*, **55**, 48–120.

Grossberg, S., Boardman, I., and Cohen, M.A. (1995). Neural dynamics of variable-rate speech categorization. Technical Report CAS/CNS-TR-94-038, Boston, MA: Boston University. *Journal of Experimental Psychology: Human Perception and Performance*, in press.

Petry, S., and G. Meyer (1987) (Eds.). **The Perception of Illusory Contours**. New York: Springer-Verlag.

Repp, B. H. 1980. A range-frequency effect on perception of silence in speech. *Haskins Laboratories Status Report on Speech Research*, **SR–61**, 151–165.

Schiller, P. H. (1992). The On and OFF channels of the visual system. *Trends in Neurosciences*, **15**, 86–92.

Sillito, A. M., H. E. Jones, G. L. Gerstein and D. C. West. (1994). Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature*, **369**, 479–482.

Squire, L. R. (1987). **Memory and Brain**. New York: Oxford University Press.

Warren, R. M. (1970). Perception restoration of missing speech sounds. *Science*, **167**, 393–395.