RESONANT NEURAL DYNAMICS OF SPEECH PERCEPTION

Stephen Grossberg

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

September, 2002
Revised: December, 2002
Revised: April, 2003
Revised: June, 2003

Technical Report CAS/CNS-TR-02-008

Correspondence should be addressed to:
Stephen Grossberg
Department of Cognitive and Neural Systems
Boston University, 677 Beacon St., Boston, MA 02215
(617) 353-7858, FAX: (617) 353-7755, steve@bu.edu
http://www.cns.bu.edu/Profiles/Grossberg

Running title: Resonant Dynamics of Speech

ABSTRACT

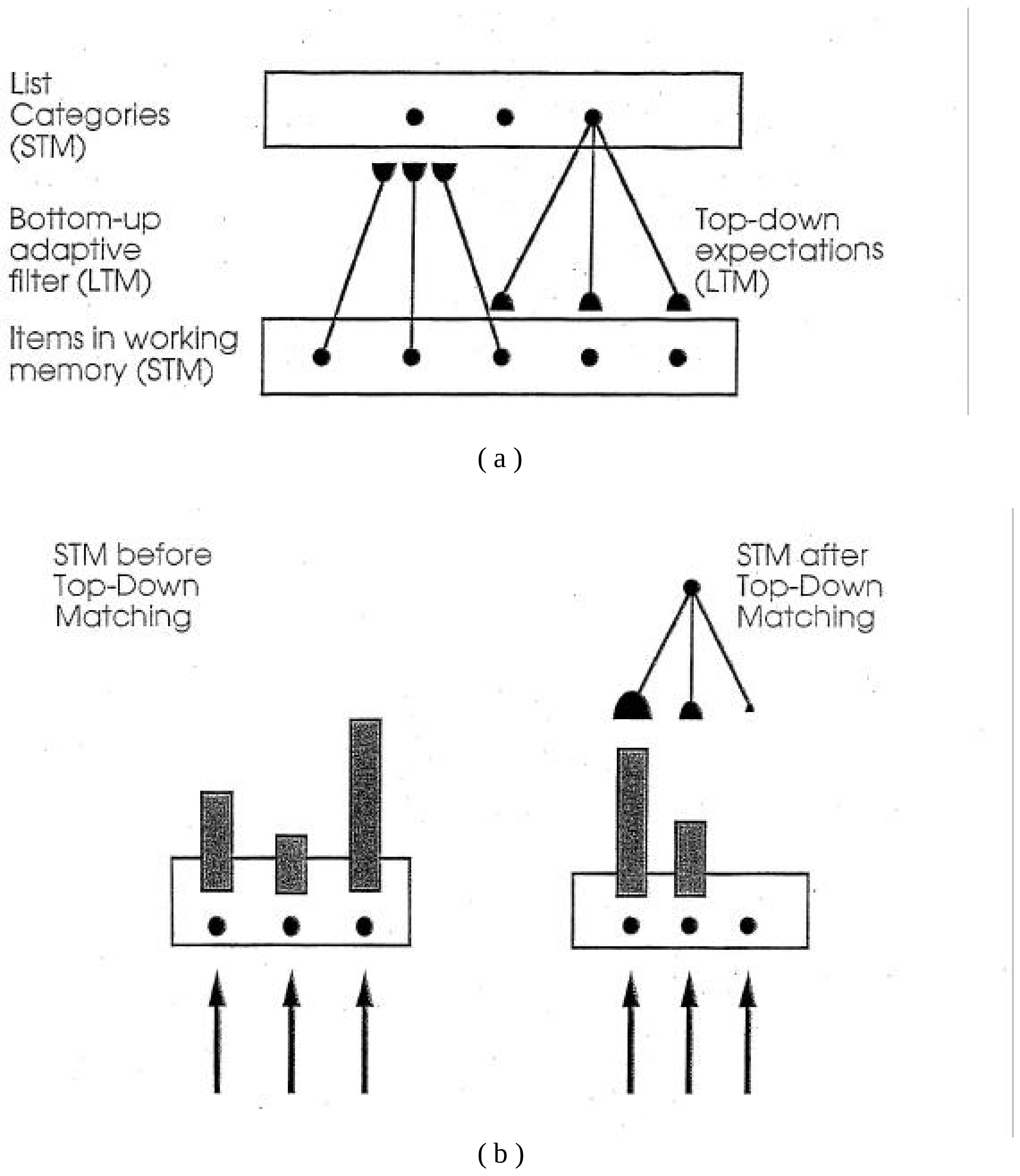
What is the neural representation of a speech code as it evolves in time? How do listeners integrate temporally distributed phonemic information across hundreds of milliseconds, even backwards in time, into coherent representations of syllables and words? What sorts of brain mechanisms encode the correct temporal order, despite such backwards effects, during speech perception? How does the brain extract rate-invariant properties of variable-rate speech? This article describes an emerging neural model that suggests answers to these questions, while quantitatively simulating challenging data about audition, speech and word recognition. This model includes bottom-up filtering, horizontal competitive, and top-down attentional interactions between a working memory for short-term storage of phonetic items and a list categorization network for grouping sequences of items. The conscious speech and word recognition code is suggested to be a resonant wave of activation across such a network, and a percept of silence is proposed to be a temporal discontinuity in the rate with which such a resonant wave evolves. Properties of these resonant waves can be traced to the brain mechanisms whereby auditory, speech, and language representations are learned in a stable way through time. Because resonances are proposed to control stable learning, the model is called an Adaptive Resonance Theory, or ART, model.

Key words: speech perception, word recognition, auditory scene analysis, consciousness, adaptive resonance, context effects, consonant perception, vowel perception, neural network, working memory, categorization, habituation, automatic gain control

1. Introduction: Learning, Expectation, Attention, and Resonance

How are brain events converted into behaviors and percepts? An answer to this question is needed if we are to understand how the brain controls behavior and how the brain is, in turn, shaped by environmental feedback that is experienced on the behavioral level. The nature of this connection also needs to be understood if we are to develop biologically plausible models of how the brain works.

The present article illustrates the hypothesis that conscious auditory and speech percepts are emergent properties that arise from resonant states of the brain. Such a resonance develops when bottom-up signals that are activated by environmental events interact with top-down expectations, or prototypes, that have been learned from prior experiences (Figure 1a). The top-down expectations control a matching process that selects those combinations of bottom-up features which are consistent with the learned prototype, while inhibiting those that are not (Figure 1b). In this way, an attentional focus starts to develop that concentrates activation on those feature clusters that are deemed important, based on past experience. The attended feature clusters, in turn, reactivate the cycle of bottom-up and top-down signal exchange. This reciprocal exchange of signals causes the selected cells to resonate with amplified and synchronized activities. Such a resonance binds the attended features together into a coherent brain state. Resonant states, rather than the activations that are due to bottom-up processing alone, are proposed to be the brain events that represent conscious behavior. The amplified and synchronous activations that occur during brain resonances are also proposed to enable the brain to learn quickly about important new information without just as rapidly forgetting all previously learned information. Adaptive Resonance Theory, or ART, is a cognitive and neural theory that is being developed to explain how the brain processes such resonant events (Grossberg, 1980, 1999b).



(a)

(b)

Figure 1. ART matching: (a) Auditory items activate short term memory (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 long-term memory (LTM) traces that influence the competitive selection of the list categories that are stored in STM. The list categories, in turn, activate top-down expectation signals that are also read out of LTM. These expectations, or prototypes, are matched against the active STM pattern in working memory. (b) This matching process selects STM activations that are supported by contiguous LTM traces (large hemi-disks) and suppresses those that are not. [Reprinted with permission from Grossberg (1999b).]

2. Phonemic Restoration and Conscious Speech Perception

A classical example of such a matching process occurs during phonemic restoration (Warren & Sherman, 1974; Samuel, 1981; Warren, 1984). Suppose that a broad-band noise is followed immediately by the rapidly presented words *eel is on the* ...” If that string of words is followed by the word *orange*, then under proper temporal conditions, subjects hear *peel is on the orange*. If the word *wagon* completes the sentence, *wheel is on the wagon* is heard. If the final word is *shoe,* then *heel is on the shoe* is heard. Such experiences show that a bottom-up stimulus alone, such as *noise-eel*, may not always determine a conscious perception. Rather, the percept may be determined by the sound that one *expects* to hear in that auditory context on the basis of previous language experiences.

To explain such percepts, we need to understand why *noise-eel* is not heard before the last word of the sentence is even presented. This may be explained by the fact that, if the resonance has not developed fully before the last word is presented, then this last word, when grouped together with earlier words of the sentence, can influence the selection of the top-down expectations that determine the conscious percept. We also need to explain how the expectation can convert *noise-eel* into a percept of *heel*. This is attributed to the top-down matching process that selects expected feature clusters for attentive processing while suppressing unexpected ones. In the *noise-eel* example, those spectral components of the noise are suppressed that are not part of the expected consonant sound. This selection process directly influences conscious phonetic percepts. It is not merely a process of symbolic inference. For example, if a silent interval replaces noise, then only silence is heard. Moreover, if the sentence *eel is on the shoe* is heard, then that sentence conveys an entirely different meaning than the sentence *heel is on the shoe*. Such data indicate that representations of phonetics and meaning must be able to intimately interact. Finally, if a reduced set of spectral components is used in the noise, then a correspondingly degraded consonant sound is heard (Samuel, 1981). This last property shows that the expectation selects consistent formants while suppressing inconsistent ones.

3. Neural Substrates of Attention, Matching, and Learning

ART has predicted that a matching law with just these properties is needed for the brain to be able to stably learn new perceptual and cognitive representations, including auditory representations, without experiencing “catastrophic forgetting” (Grossberg, 1980, 1999b). The type of matching which is evident in speech percepts like phonemic restoration is thus proposed to illustrate brain mechanisms for rapidly learning speech and language representations. ART predicts that such matching is realized by a top-down modulatory on-center off-surround circuit (Figure 2a). The on-center is modulatory because its excitatory and inhibitory signals are approximately balanced. Such a top-down signal may increase the baseline activity of cells in the on-center, but cannot fire them by itself. The off-surround can strongly inhibit non-matched cues. Related modeling work has identified laminar cortical circuits that realize such a top-down circuit and are capable of carrying out the ART matching rule (Grossberg, 1999a; Raizada & Grossberg, 2001); see Figure 2b. It has also been shown how, when the modulatory property fails, auditory hallucinations may result (Grossberg, 2000a).

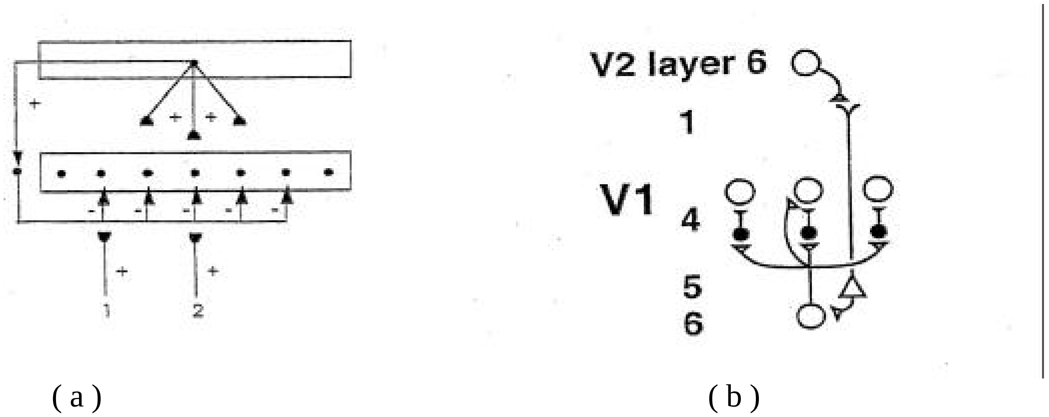


Figure 2. (a) The ART matching rule may be realized by a modulatory top-down on-center off-surround network. When only bottom-up signals 1 and 2 are active, they can activate their target cells. When only the top-down expectation is active, it cannot activate any cells. This is because a cell that receives an excitatory signal in the on-center (top-down + pathways) will also receive an approximately matched inhibitory signal (top-down – pathways). The baseline activity of such a cell may be slightly excited by a net advantage to the excitatory pathway. Cells that receive no excitatory signals can be suppressed. When both bottom-up and top-down signals are active, the bottom-up activation caused by pathway 2 can persist, and even be amplified, but the bottom-up activation caused by pathway 1 is suppressed by the off-surround. (b) One known anatomical pathway by which cortical cells from a higher level, in this case area V2 of the visual cortex, can attentionally modulate the activity of cells at a lower level, in this case area V1. In particular, a top-down output signal from layer 6 of the higher cortical area can activate apical dendrites of layer 5 pyramidal cells at the lower cortical area. These pyramidal cells can activate layer 6 of the lower cortical area, which in turn activates a modulatory on-center off-surround network of inputs to layer 4 of the lower cortical area (open symbols are excitatory on-center cells; closed symbols are inhibitory off-surround cells). This circuitous route from layer 6-to-6-to-4 is called "folded feedback" because it folds feedback from a higher cortical area into the bottom-up processing of signals within a lower cortical area. [Reprinted with permission from Grossberg & Raizada (2000).]

Recent neurophysiological experiments have begun to directly confirm the ART-predicted links between learning, top-down matching, attention and synchronous resonant dynamics. In particular, the claim that bottom-up sensory activity is enhanced when matched by top-down signals is in accord with an extensive neurophysiological literature showing the facilitatory effect of attentional feedback (e.g, Luck, Chelazzi, Hillyard, & Desimone 1997; Roelfsema, Lamme, & Spekreijse 1998). The on-center off-surround structure of top-down feedback has been demonstrated in the visual system both for V2-to-V1 feedback (Bullier, Hupé, James, & Girard 1996) and for V1-to-LGN feedback (Sillito, Jones, Gerstein, & West 1994). Zhang, Suga, & Yan (1997) have shown that feedback from auditory cortex to the medial geniculate nucleus and the inferior colliculus also has an on-center off-surround form. Temereanca & Simons (2001) have described evidence for such feedback in the rodent barrel system. Various other psychophysical and neurophysiological data have also shown that attention has a facilitatory on-center and suppressive off-surround (Downing, 1988; Steinman, Steinman, & Lehmkuhle, 1995; Caputo & Guerra, 1998; Mounts, 2000; Smith, Singh, & Greenlee, 2000; Vanduffel, Tootell, & Orban, 2000).

Ahissar & Hochstein (1993) have provided psychophysical evidence for the predicted role of attention in controlling adult plasticity and perceptual learning. Gao & Suga (1998) have reported neurophysiological evidence that acoustic stimuli caused plastic changes in the inferior colliculus of bats only when it received top-down feedback from auditory cortex. This plasticity was also found to be enhanced when the auditory stimuli were made behaviorally relevant. Krupa, Ghazanfar, & Nicolelis (1999) and Parker & Dostrovsky (1999) have found that cortical feedback also controls thalamic plasticity in the somatosensory system.

Various data have also linked the themes of top-down cortical feedback and learning to the prediction that resonant states tend to synchronize the activities of the resonating cells. Pollen (1999) and Engel, Fries, & Singer (2001) have provided excellent reviews of these ART-supportive literatures. In summary, recent neurophysiological and psychophysical data have provided significant additional experimental support for basic ART predictions that were first made in the 1970s.

4. How Can the Future Influence the Past while Perception Proceeds from Past to Future?

Given that a resonant event may lag behind the environmental stimuli that cause it, we need to develop a refined concept of how perceived psychological time is related to the times at which stimuli are presented. In particular, how can "future" events influence the perception of "past" events, yet time be perceived to always flow from past to future? ART suggests that this is accomplished by a resonant wave that develops from past to future while it incorporates future constraints into its top-down decision process until each event in the resonance equilibrates. In other words, future events can influence past events if the future events occur in the time interval after the past events are first registered by the brain, yet before the past events resonate and become conscious.

In order to represent such a time-dependent resonant process, we need to distinguish the external input rate from the internal processing rate at which the resonance evolves. Because external events may, in principle, occur at arbitrary times, the brain's rate process must have a finer time scale than any detectable external rate. It must also be faster than the resonance time scale that emerges as a result of bottom-up and top-down interactions. That is why differential equations are used to describe ART models. Differential equations are the universally accepted mathematical formalism in science that is used to describe events that are evolving in real time. A related question concerns how future events can influence past events without smearing over all the events that intervene. In particular, how can silent intervals seem to separate the words in the sentence *heel is on the shoe* given that the influence of *shoe* must cross all of the preceding sounds to influence *heel*? Here again the nature of the top-down matching process is paramount. This matching process can *select* feature components that are consistent with its prototype, but it cannot *create something out of nothing*. Without bottom-up input, the top-down on-center is merely modulatory. As a result, when the signal includes physical silence, a silent interval is perceived, as silence remains silence no matter how active the top-down prototypes may be.

5. Perceived Silence is a Discontinuity in the Rate at which Resonance Evolves

The opposite concern must also be considered: How can sharp word boundaries be perceived even if the sound spectrum that represents a sequence of words exhibits no silent intervals between them? ART proposes that an apparent discontinuity between words will be heard whenever there is a temporal delay between the resonances that represent the individual words.

In other words, apparent silence of this type is a discontinuity in the rate at which resonance evolves. This hypothesis helps to account for the fact that there is no simple relationship between the presence or absence of energy in the waveform at a given time, and the experience of apparent silence associated with perceived word boundaries at that time. Data of Repp (1980), admittedly for nonsense syllables rather than real words, illustrate the complexity of the relationship between the duration of physical silence in a particular context and what is perceived. The syllables [ib], [ga] and [ba], synthesized without bursts and with identical parameters except for the first three formant frequencies, were combined so that [ib] was followed by either [ga] or [ba], and the duration of the silent interval between the two syllables varied in equal steps, 10 ms for [ib]-[ga] and 20 ms for [ib]-[ba]. The ranges over which the silences varied also differed for the two sequences, as shown in Figure 3. Listeners made forced-choice responses, ‘b’ or ‘bg’ in the single/cluster condition, and ‘b’ or ‘bb’ in the single/geminate condition. As Figure 3 shows, listeners heard a single consonant, [iga] or [iba], at the shorter silent intervals in each series, and two consonants, [ibga] or [ibba], at the longer intervals. Moreover, they heard two distinct consonants in the sequence [ibga] at the same silent interval at which they heard only one consonant (i.e. [iba], not [ibba]) in the single-geminate condition, in which the spectral cues in each syllable were consistent with a single place of consonantal articulation. ART explains such data as an example of *resonant reset* of /b/ by /g/ in the case of [ibga], and *resonant fusion* of the first /b/ with the second /b/ in the case of [ibba], which fills a physically silent interval with the percept of a single place of plosive articulation; see Figure 4.

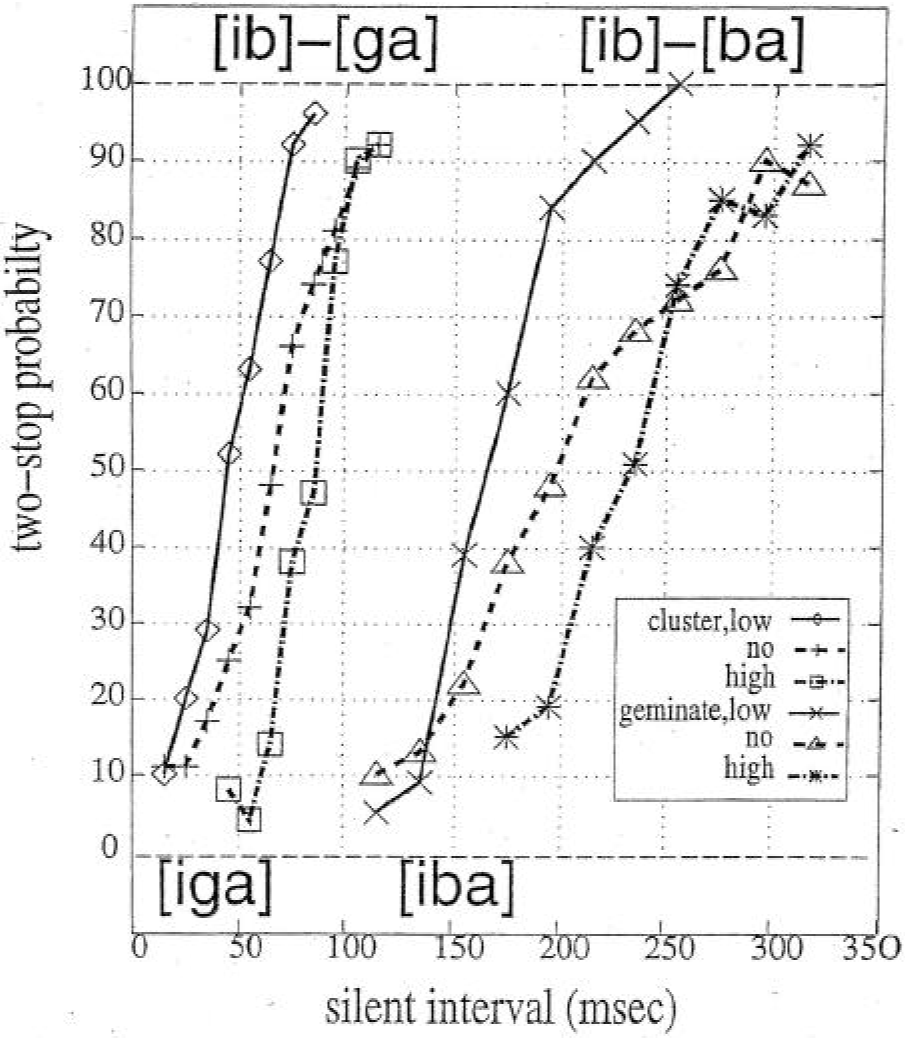


Figure 3. Probability (%) of perceiving two consonants in response to the VC-CV pairs [ib]-[ga] and [ib]-[ba], instead of a single intervocalic consonant, [iga] or [iba], when the mean silent interval is varied. ‘Cluster’ and ‘geminate’ refer to the [ib]-[ga] and [ib]-[ba] series respectively. ‘Low’, ‘no’, and ‘high’ refer to stimulus frequency effects built into the design. See text for details. [Adapted with permission from Repp (1980).]

Given that a resonance can be reset by a mismatching later event, it remains to explain how resonances end when there are no later mismatching events. The ART model proposes that chemical transmitters are released within the resonating pathways of the model. Persistent release of these transmitters during resonance inactivates, habituates, or depresses them, thereby weakening the feedback signals that support the resonance and ending it. Thus, a resonance can be terminated by either of two mechanisms: *resonant reset* or *habituative collapse*.

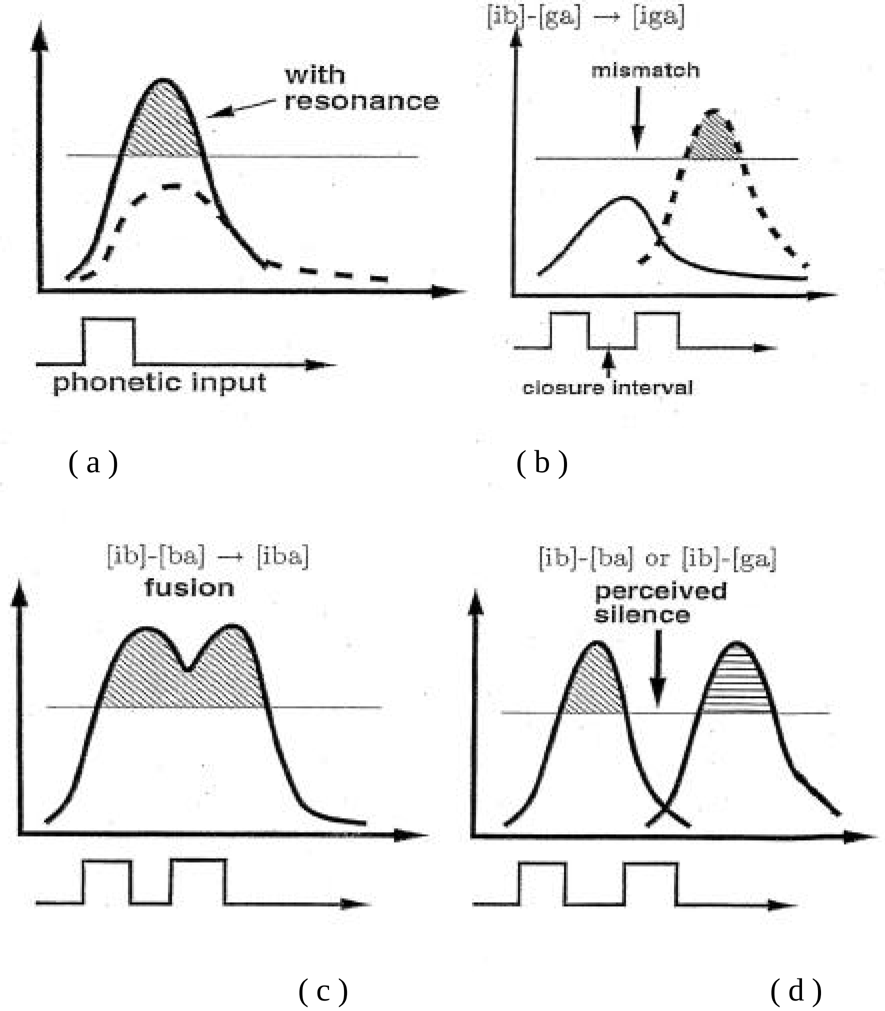


Figure 4. Heuristic summary of the resonances that are proposed to explain the data in Figure 3. (a) Response to a single stop with (solid line) and without (dashed line) resonant feedback. The ordinate represents category node activity and the abscissa represents time. The horizontal line represents an activation above which resonance occurs, along with a conscious percept. Resonant activation is shaded. (b) Reset due to phonologic mismatch. Here the activity corresponding to /b/ is reset by mismatch with /b/ before /b/ can resonate. Only the /g/ sound reaches resonance, leading to a percept of /iga/. (c) It takes a while for the /b/ activity to grow large enough to resonate, but a second occurrence of /b/ can more quickly boost already resonant activity. If the second /b/ occurs before the resonance in response to the first /b/ can collapse, then fusion of the two /b/ sounds can occur over the intervening silent interval, leading to prolongation of the /iba/ resonance through time. (d) A sufficiently long silent interval allows a two-stop percept to be heard. Habituative collapse of the /ib/ resonance before the /ba/ or /ga/ resonance develops leads to a percept of both syllables separated by a silent interval. [Adapted with permission from Grossberg, Boardman, & Cohen (1997).]

Figure 3 also illustrates an important property of *variable-rate* speech perception. Note that there are three category-boundary curves for each of the VC-CV syllables [ib]-[ga] and [ib]-[ba]. This is because the mean physically silent interval was varied in each of the three cases for a fixed pair of VC-CV syllables. A low anchor condition had its distribution of silent intervals skewed to be shorter, a high anchor condition had them skewed to be longer, and a no anchor condition had no skew. The three curves show that listeners track the mean physically silent interval in each case in order to make their judgements. These rate-*dependent* category boundary shifts have been explained by assuming that the brain is trying to create rate-*independent* speech and language representations. In particular, humans can understand variable-rate speech without having to represent it internally at every rate, which would create an uncontrollable combinatorial explosion and undeciferable decoding problem. The model assumes that there exists a rate-dependent gain control process which can speed up or slow down the integration rate of the resonance in response to the speech rate. The model uses the rate of speech, notably of speech transients, to control the integration rate. Figure 5 summarizes the ARTPHONE model of Grossberg, Boardman, & Cohen (1997) which quantitatively simulated the Repp (1980) data, and showed how to explain many other data of this type, by using a combination of resonant feedback, rate-dependent gain control, and habituative transmitter gating.

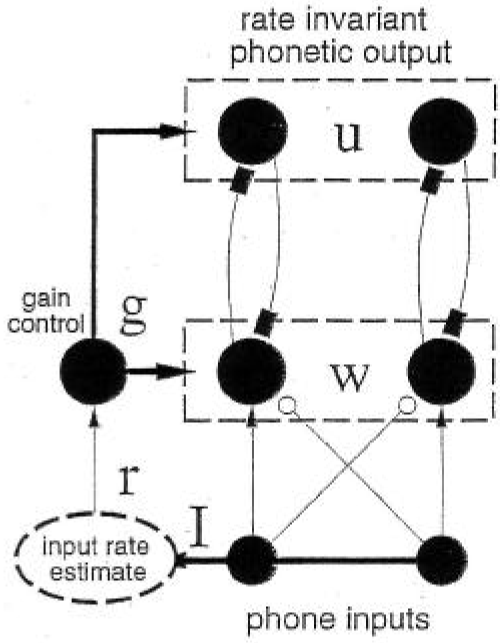


Figure 5. The ARTPHONE model: Working memory item activities (w) excite list chunk activities (u) through previously learned bottom-up pathways. List chunk activities send top-down excitatory feedback down to their item source cells. Bottom-up and top-down pathways are modulated by habituative transmitter gates (filled squares). Item cells receive inputs in an on-center off-surround anatomy. The sum of all the inputs (I) is averaged through time to generate a time-averaged estimate of input rate (r) that adjusts the working memory gain (g). Because this gain tracks the speech rate, it adjusts the integration rates of the working memory and chunking network to compensate for changes in speech rate. Excitatory paths are marked with arrowheads, inhibitory paths with small open circles. [Reprinted with permission from Grossberg, Boardman, & Cohen (1997).]

The example in Figure 3 may at first seem to be in conflict with the lesson learned from phonemic restoration that a bottom-up signal, such as the noise in “noise-eel,” is needed to hear a word like “heel” in response to the action of a top-down expectation: The top-down expectation, by itself, cannot “create something out of nothing”. How, then, can the sound /b/ in Figure 3 fill such a long silent interval during which it is not actively being generated by spectral energy from the outside world? ART proposes that this can happen because sounds are stored in a working memory (Miller, 1956; Baddeley, 1986) which enables them to be acted upon by later-acting, top-down expectations, much as the noise is stored in “noise-eel” until the last word of the rapidly presented sentence “noise-eel is on the shoe” can occur and lead to the percept of “heel.”

6. Resonance between STORE Working Memory Items and List Chunk Categories

Working memories cannot work well unless they operate at correctly defined processing levels and obey correctly constrained laws. In particular, it is not sufficient to posit that phonemes, letters, and words each have their own processing levels, as proposed in the Interactive Activation Model (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). These levels are inadequate because they cannot learn stable representations of words in an unsupervised fashion, and are not consistent with various data about word recognition, as was noted in Grossberg (1984, 1986). Rather, processing levels that compute more abstract groupings of auditory signals are needed. In particular, a working memory is posited that represents sequences of *items* that have been unitized through prior learning experiences. Familiar feature clusters that are presented within a brief time interval become items by being categorized, or unitized, at a processing stage that occurs before the working memory stage. As item categories are processed through time, they input to the working memory stage at which multiple items are simultaneously stored as part of an evolving spatial pattern of activations across a network of item representations. The working memory hereby recodes a temporally occurring sequence of events into an evolving spatial pattern of activation. This spatial pattern represents both item information (which items are stored) and temporal order information (the order in which they are stored). Individual items can be recalled when a rehearsal wave nonspecifically activates the entire working memory. The rehearsal wave allows the most active items to be recalled first, after which they inhibit their own representations using recurrent inhibitory feedback, so that less active items can also be recalled in the order of their relative activity (Grossberg, 1978a, 1978b; Koch & Ullman, 1985).

A number of articles have modeled the design principles governing such *item-and-order working memories* (Grossberg, 1978a, 1978b; Cohen & Grossberg, 1986) and have used them to explain data about free recall (Grossberg, 1978a, 1978b), reaction time during sequential motor performance (Grossberg & Kuperstein, 1986/1989; Boardman & Bullock, 1991), errors in serial item and order recall that are due to rapid attention shifts (Grossberg & Stone, 1986a), errors and reaction times during lexical priming and episodic memory experiments (Grossberg & Stone, 1986b), and data concerning word superiority, phonemic restoration, and backward effects on speech perception (Grossberg, 1986). Such a wide range of data fall under the purview of these working memory models because they all satisfy two simple postulates (Grossberg, 1978a, 1978b; Bradski, Carpenter, & Grossberg, 1992, 1994). These postulates predict how unitized representations that are activated by lists of items in working memory can be learned in a stable way; see Figure 1. These representations are called “list chunks” because they are selectively activated by prescribed lists of items in working memory. List chunks may be learned in response to *any* grouping of items that can successfully and persistently activate such a chunk. List chunks can, in principle, code familiar words, as well as phonemes, letters, and syllables; cf., Greenberg (2003), in this volume. All these language units can be represented within a single chunking network.

The key postulate, which I have called the *LTM Invariance Principle* (Grossberg, 1978a, 1978b), proposes how working memories, which encode a type of *short-term* memory, are designed in a way to enable the stable learning and *long-term* memory of list chunks. For example, after having learned the words *my* and *self*, suppose that the word *myself* is temporarily stored in working memory for the first time. How does a listener learn a new word representation for *myself* that is stored in long-term memory, without erasing the previously learned word representations for *my* and *self*? When such learning occurs in an unsupervised fashion in real time, as it does when a child learns a language, a poorly designed working memory could easily cause catastrophic forgetting of *my* and *self* when learning *myself*.

The second postulate requires that the maximal total activity of the working memory is finite and, indeed, independent of the total number of active chunks. This postulate implies the limited capacity of working memory: Since the total activity in working memory cannot increase indefinitely with the number of chunks that might be activated, some chunks must be eliminated from working memory in order to enable other chunks to be stored.

Working memories that satisfy these postulates have been called STORE (Sustained Temporal Order REcurrent) models. Remarkably, specialized recurrent on-center off-surround networks naturally satisfy both of the STORE postulates. Recurrent on-center off-surround networks are ubiquitous in the brain because they enable distributed input patterns to be processed without a loss of sensitivity by their target cells (Grossberg, 1980). Thus, designing a working memory, which may seem at the outset to be a highly sophisticated task, reduces to adapting an ancient neural design that is needed to process all sorts of spatially distributed data.

7. The Temporal Chunking Problem and Multiple-Scale Chunks in Masking Fields

In all the examples of ART speech and language processing, working memories interact reciprocally with a categorization network that represents list chunks (Figure 1). These list chunks may represent the items themselves or larger groupings of items of variable length--such as phonemes, letters, syllables, or words—that can all be represented at the same masking field level. Which of these chunks will be activated or learned in any given situation depends upon contextual properties of the language. A chunking network called a *masking field* is designed to select those list categories that are most predictive of the temporal context that the items, taken together, collectively generate across the working memory (Grossberg, 1978a, 1986; Cohen & Grossberg, 1986, 1987; Grossberg & Myers, 2000). The word *masking* connotes that active list chunks which represent longer sequences of items, and thus a less ambiguous temporal context, can inhibit or *mask* chunks that represent shorter sequences of items, more than conversely. This property gives a chunk that codes a longer novel list like *myself* an *a priori* advantage over chunks that code shorter lists like *my* and *self*. If the shorter lists are familiar, and thus already coded by shorter chunks, then their adaptive filters in the masking field will have been tuned by learning to more strongly activate these chunks than before they were familiar. The masking advantage of the chunk that represents *myself* is needed for it to successfully compete with these amplified familiar chunks of shorter lists like *my* and *self*, even before the list chunks that *myself* activates can be tuned through learning to that word.

A second important property of a masking field is that a list chunk does not fire unless it receives enough bottom-up evidence. Otherwise, incompletely activated large chunks could always win over small chunks due to their greater masking potency. Larger chunks can be subliminally *primed* by incomplete evidence, and thereby readied to fire when that evidence becomes sufficient, but do not mask smaller chunks if insufficient evidence is available. Thus, even if *my* and *myself* are both familiar words, the *myself* chunk may be primed by inputting the word *my*, but cannot fire suprathreshold signals, thereby enabling the chunk for *my* to survive the otherwise overwhelming inhibition from the *myself* chunk. As the word *my* is supplanted by presentation of the complete word *myself*, the pattern of active list chunks shifts through time, until *myself* has enough evidence to mask the chunks for *my*, *self*, *elf*, etc.

A third property is that a masking field can represent lists only up to a finite maximal length, which is determined by the way in which the connections of the adaptive filter develop from the item field, or working memory, to the masking field, or chunking network (Cohen & Grossberg, 1986). Lists that are longer than the maximal length are necessarily represented by more than one list chunk. Shorter lists of items can also be represented by more than one list chunk, but the set of active chunks becomes more selective as the item list becomes more predictive (Cohen & Grossberg, 1986, 1987).

By implementing all of these properties, a masking field provides a solution of what I have called the *Temporal Chunking Problem* (Grossberg, 1984; Cohen & Grossberg, 1986); namely, why is not every list of items coded in terms of already familiar chunks, like *my* and *self*? How can a new, heretofore non-existent, representation of a novel word, like *myself*, begin to form, under the type of unsupervised learning conditions that typify a child's language learning experiences, despite the competitive salience of its previously learned parts, like *my* and *self*? Masking Fields help to solve the Temporal Chunking Problem by enabling larger chunks to form despite the salience of smaller, previously chunked lists, and STORE working memories embody the LTM Invariance Principle so that learning of these larger chunks does not force catastrophic forgetting of smaller, previously learned, chunks. The chunks that win the masking field competition activate their top-down expectations and thereby selectively amplify and focus attention upon consistent working memory items, while suppressing inconsistent working memory items. Feedback establishes a resonance which temporarily boosts the activation levels of selected working memory items and list chunks, thereby creating an emergent conscious percept.

We earlier noted that item working memories that embody the LTM Invariance Principle can be defined using recurrent on-center off-surround networks. How can a masking field of list chunks that solves the Temporal Chunking Problem be designed? Remarkably, masking fields can also be defined using recurrent on-center off-surround networks, albeit *multiple-scale* networks that can represent lists of variable length (Grossberg, 1978a; Cohen & Grossberg, 1986, 1987). This is a satisfying conclusion, because it clarifies how a masking field network that represents list chunks of a previous working memory in a network hierarchy can also serve as the working memory for a still higher level of list chunks in the hierarchy that can represent list chunks of list chunks. And so on. The ability of larger chunks to mask smaller chunks in a masking field translates into the statement that the multiple size chunks are *self-similar* with respect to one another. In other words, the bigger cells that represent longer lists can more strongly inhibit the smaller cells that represent shorter lists within the masking field, and can provide stronger excitatory priming for items at the previous working memory. All parameters of one cell size are thus (approximate) multiples of the parameters of other cell sizes. This self-similarity property can grow in the network using simple developmental rules (Cohen & Grossberg, 1986).

I note in passing that the above discussion is consistent with the fact that some acoustical properties of *my* as a word may subtly differ from those of *my* as a syllable of *myself*, such as prosodic differences. A full discussion of these differences is beyond the scope of this review. However, it can be noted that a key goal of this part of the system, as noted below, is to derive a speech representation that is independent of speaker rate and various other prosodic variations. This representation is then used to encode such variations in a parallel working memory. This design principle is called the Factorization of Order and Rhythm (Grossberg, 1986).

8. The Magical Number Seven, Word Superiority, and Backward Effects

Self-similar multiple-scale chunking networks have explained and predicted a variety of data about list categorization, including the Miller (1956) "Magical Number Seven, Plus or Minus Two" and the Samuel, van Santen, & Johnston (1982, 1983) data about word length effects during word superiority experiments (Grossberg, 1978a, 1984). The Magical Number Seven summarized data about the maximal number of chunks that could simultaneously be stored in working memory, independent of chunk size. This property naturally follows from self-similarity, since each winning subnetwork of chunks has similar properties to all the other subnetworks, and has a limited capacity. The word length effect says that a letter is better recognized as it is embedded in longer words of lengths from 1 to 4. This property also follows from self-similarity; namely, from the fact that chunks which code longer lists can provide stronger excitatory priming of their items in working memory. The data of Wheeler (1970) inspired discovery of the word length effect. These data challenge the idea that letters and words are represented on separate processing levels. Were this true, then letters such as I and A that are also words would be better recognized than other letters, since they would selectively receive excitatory top-down feedback from the word level. This is not true. Within a masking field, familiar letters that are words, as well as familiar letters that are not words, have list chunks, and thus a top-down advantage for letters that are words is not expected.

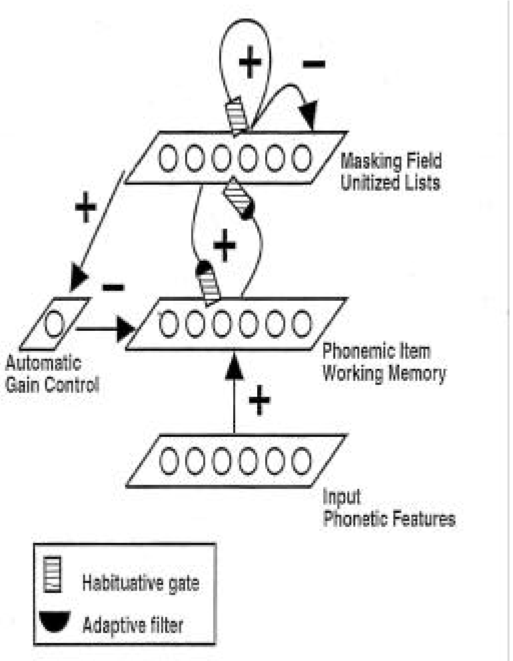


Figure 6. Masking field and ARTWORD model: (a) A masking field is organized so that longer item sequences, up to some optimal length, selectively activate cells with more potent masking, or inhibitory, properties. Familiar individual items, as well as item sequences, may be represented in the masking field. See text for details. [Reprinted with permission from Cohen & Grossberg, 1986.] (b) The ARTWORD model incorporates the same types of mechanisms as in the ARTPHONE model, with the addition of a multiple-scale masking field. [Reprinted with permission from Grossberg & Myers (2000).]

Later studies exploited the fact that, because resonances kick in later than their initial working memory activations, they can be influenced by information presented after relatively long intervening silent intervals. Variations in the durations of speech sounds and silent pauses can hereby produce different perceived groupings of words, and future sounds can influence how we hear past sounds. For example, the ARTWORD model (see Figure 6) was developed in Grossberg & Myers (2000) to quantitatively simulate context-sensitive speech categorization data wherein variable-length list groupings can strongly influence conscious percepts. The ARTWORD model generalized the earlier ARTPHONE model by incorporating a multiple-scale masking field categorization network. ARTWORD was used to quantitatively simulate psychophysical data showing that increasing the silent interval between the words *gray chip* may result in the percept *great chip*, whereas increasing the duration of fricative noise in *chip* may alter the percept to *great ship* (Repp, Liberman, Eccardt, & Pesetsky, 1978); see Figure 7. These patterns reflect those of natural speech, and are to be expected in a forced-choice design when only the durations of silence and aperiodicity varied. Nonetheless, data of this kind are paradoxical for several reasons, and we need to account for them. For example, how can increasing the silent interval between the two words *gray* and *chip* cause *gray* to sound like *great*? Even allowing for the fact that the spectro-temporal patterns here reflect the patterns made in natural speech to distinguish these words, might not one expect greater temporal separation between two familiar words to make them easier to distinguish, not to enable a sound from the later word to leap over a 100 ms silent interval to change the percept of one familiar word, *gray*, to another one, *great*? Likewise, how can increasing the duration of the fricative noise in *chip* make it easier for the noise to leap over the silent interval between it and the earlier word *gray* to generate the percept *great* while leaving behind the remainder of the word *chip* to which it is temporally contiguous, leading to the percept *ship*?

The resonant dynamics of STORE working memory and masking field interactions account naturally for these context effects. For example, when the word *gray* is heard, a masking field can fire the *gray* chunk but can only prime the chunk for *great*. When the interval of silence between the two words is relatively long, a *resonant transfer* from the *gray* to the *great* chunk becomes more likely because the *gray* chunk can resonate longer with its working memory items, and thus the transmitters in the bottom-up and top-down pathways that support this resonance can become more habituated, thereby weakening activation of the *gray* chunk, along with its ability to inhibit other competing chunks. The *great* chunk can be quickly activated when the fricative noise in *chip* occurs because it was already significantly primed by the auditory signals that activated *gray*. When the *great* chunk finally fires, it can more easily inhibit the *gray* chunk after the longer interval, thereby altering the percept to *great*.

Such data and their explanatory resonances emphasize how future sounds can leap over a physical interval of silence to join a past word. The same type of resonance mechanisms were used in the ARTPHONE simulations of the data in Figure 3. These and many other long-domain context-sensitive percepts (see Coleman, 2003; Hawkins, 2003; and Local, 2003; all in this volume) can now be given a unified explanation in terms of how the evolving resonance between working memory items and list chunks naturally leads to such context effects.

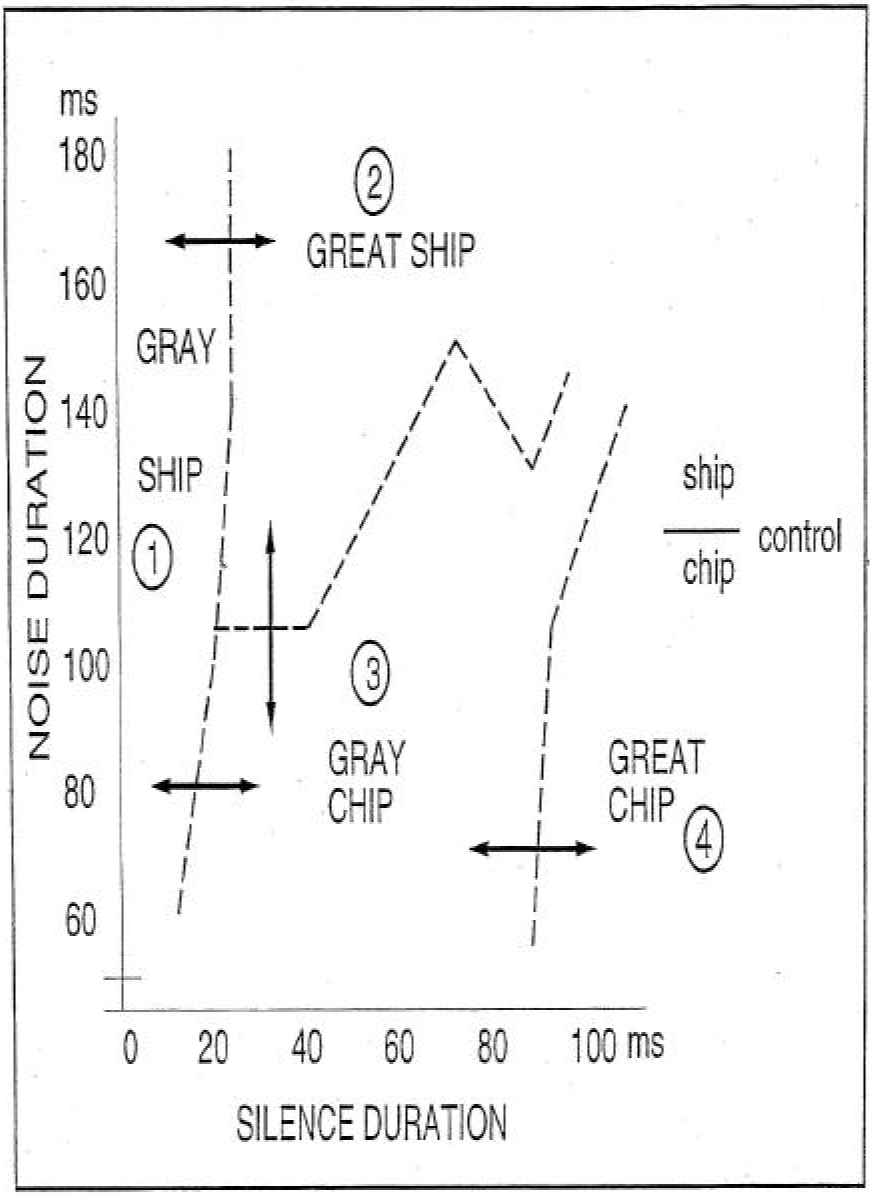


Figure 7. (a) Perceptual boundaries reported in Repp et al. (1978). Four distinct percepts, in regions (1) through (4), may be perceived when silence duration and/or fricative noise duration are varied. See text for details. [Adapted with permission from Repp et al. (1978).]

9. How is a Temporally Invariant Speech Code Derived from Variable-Rate Speech?

In order to explain othe r types of auditory and speech data, preprocessing of acoustic signals plays a crucial role. As noted above, such preprocessing has been proposed to play a role in solving the problem of how humans can understand variable-rate speech without having to represent it internally at every rate, and thereby creating an uncontrollable combinatorial explosion and undeciferable decoding problem. The ARTPHONE model of Figure 5, and its elaboration in the ARTWORD model, suggested that preprocessing in the form of a rate-dependent gain control adjusts the integration rate of working memory and chunking networks to keep up with the overall speech rate.

In addition to this long-term change in processing rates, it has also been proposed that a rapidly acting gain control helps to generate more temporally-invariant representations within individual syllables. In particular, Cohen & Grossberg (1997) proposed that auditory signals are processed by parallel auditory streams, one of which responds preferentially to transient, and the other to sustained properties of the acoustic signal before being stored in parallel transient and sustained working memories. In fact, it is well known that there are cells in the brain's auditory system that are selectively sensitive to transient and sustained properties of acoustic waveforms (Britt & Starr, 1976; Sachs & Young, 1979; Young & Sachs, 1979; Moller, 1983; Delgutte & Kiang, 1984a, 1984b; Rhode & Smith, 1986; Pickles, 1988; Mendelson, Schreiner, Sutter, & Grasse, 1993; Tian & Rauschecker, 1994). Cohen & Grossberg (1997) suggested that this decomposition helps to partially separate coarticulated consonants and vowels into distinct, but parallel, working memories, and also to explain data about auditory-nerve processing. Decomposition of sensory signals into transient and sustained cell responses is also well known to occur within the visual system.

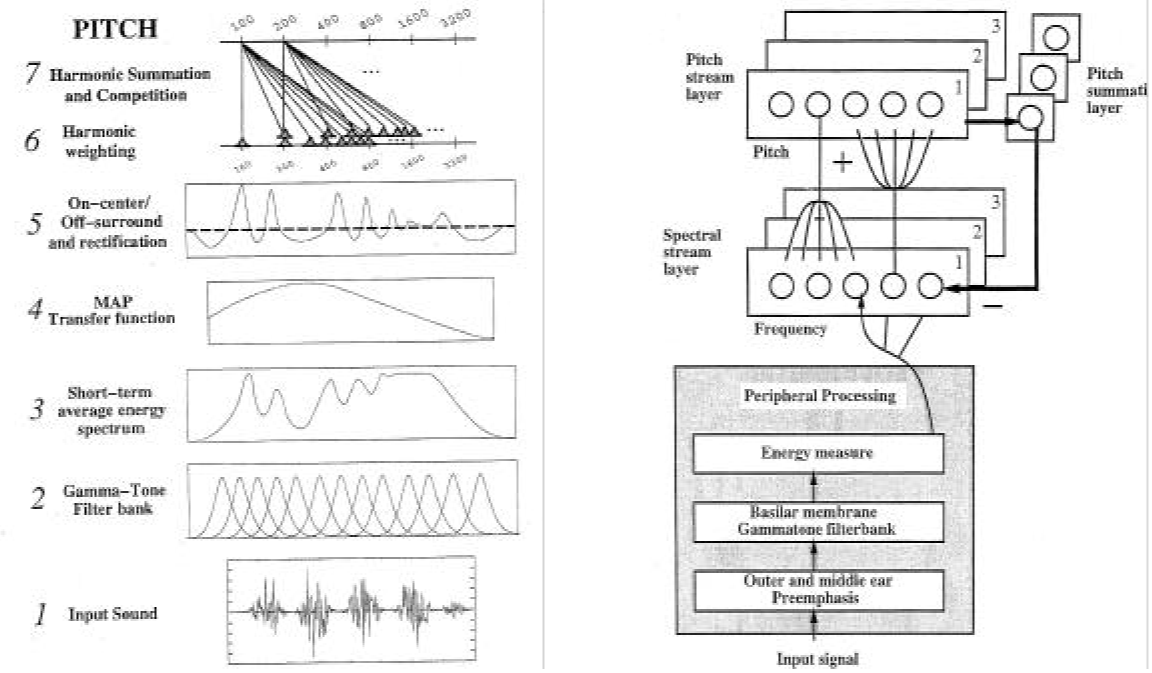
Boardman, Grossberg, Myers, & Cohen (1999) modeled these transient and sustained working memories in the PHONET model and used their interaction to realize a more rate-invariant representation of speech. In particular, increased activation of the transient working memory was proposed to increase the gain of the integration rate within the sustained working memory. Several experiments had earlier reported asymmetric vocalic context effects (Kunisaki & Fujisaki, 1977; Mann & Repp, 1980) from transient to sustained, but not conversely. Such asymmetric gain control in PHONET leads to a working memory representation of an individual syllable or word that is more invariant under changes in speech rate. For example, a faster speech rate may more strongly activate transients in the speech signal but leave less time to integrate sustained signals. Asymmetric gain control tends to correct this imbalance and to thereby generate *relative* activities across transient and sustained working memory representations that compensate for, and are thus invariant under, variable speaking rates. Invariant relative activations, in turn, lead to preserved categorization by the chunking network because the bottom-up adaptive filter pathways between the working memory and the list chunks tends to activate the same set of chunks when relative activities are preserved (e.g., Grossberg, 1980).

Parallel preprocessing into transient and sustained channels, followed by gain control from the transient to sustained channel, was used in PHONET to quantitatively simulate how, in CV syllables such as /ba/ and /wa/, an increase in the duration of the vowel /a/ can cause a switch in the percept of the preceding consonant from /w/ to /b/ (Miller & Liberman, 1979), and how a change in frequency extent (namely, total frequency change), but not rate, can also influence the /b/-/w/ distinction (Schwab, Sawusch, & Nusbaum, 1981).

The rapidly-acting gain control in the PHONET model is proposed to work together with the longer-persisting gain control of the working memory and chunking network integration rates of the ARTWORD model. Taken together, these two sorts of gain control begin to explain how a temporally-invariant speech representation may be created internally by brain dynamics from variable-rate speech signals that themselves do not exhibit obvious properties of invariance. Other types of preprocessing are also needed to prepare auditory signals for speech analysis, as the following section notes.

10. Resonance during Auditory Streaming

If adaptive resonance is indeed a mechanism to control rapid learning of perceptual and cognitive codes without catastrophic forgetting, then resonant interactions may be expected to occur at multiple levels of the auditory system. In fact, suitably designed resonance networks have also been helpful in explaining data about auditory streaming and the cocktail party problem, albeit at a different level of auditory processing (Grossberg, 1999b, 1999c). It is well known that pitch is one of the major properties of the auditory signal that is used to separate voices or instruments into separately perceived auditory streams (Bregman, 1990). Auditory signals are thus first preprocessed to extract the pitch of a voice or instrument by using a model of early auditory preprocessing called the SPINET model, or Spatial PItch NETwork, so-called because SPINET converts temporally-occurring auditory signals into spatial representations of pitch (Cohen, Grossberg, & Wyse, 1995); see Figure 8a. The SPINET model was validated by quantitatively simulating many psychophysical data about pitch perception by human observers. Unlike more traditional transform models of pitch, the SPINET model proposes how different spectral and pitch representations can be represented in a spatial map. A key hypothesis of SPINET is that harmonically-related spectral components (see stages 6 and 7 in Figure 8a) can activate a given pitch category through an adaptive filter that obeys laws similar to those which activate list categories in the speech models. As in the speech models, it is assumed that the selection of harmonics by the filter is due to learning, in particular, learning that is driven by the natural harmonic grouping of frequencies due to early auditory processing.



(a) (b)

Figure 8. SPINET and ARTSTREAM models: (a) SPINET model processing stages transform a sound stream into activations of spatially distributed pitch nodes. [Reprinted with permission from Cohen, Grossberg, & Wyse (1995).] (b) ARTSTREAM model processing stages. The spectral and pitch layers of the SPINET model (layers 6 and 7) are elaborated in the ARTSTREAM model into multiple representations, or strips of cells, and top-down ART matching also occurs. Bottom-up signals group harmonically-related spectral components into activations of pitch categories. Inhibition within each pitch stream enables only one pitch category to be active at any time in a given stream. Asymmetric inhibition across streams in the pitch stream layer is biased so that the winning pitch cannot be represented in another stream. The winning pitch category feeds back excitation to its harmonics in the corresponding spectral stream. This stream also receives nonspecific top-down inhibition from the pitch layer. ART matching is hereby realized. It suppresses those spectral components that are not harmonically related to the active pitch. Inhibition across spectral streams then prevents the resonating frequency from being represented in other streams as well. [Reprinted with permission from Grossberg (1999b).]

Because of its spatial representation, the SPINET model could be naturally extended to define a more comprehensive model of pitch-based auditory streaming, called the ARTSTREAM model. To do this, the spectral and pitch representations of the SPINET model are extended into multiple representations of each frequency and pitch across a spatial map (Figure 8b). These multiple representations give rise to "strips" of the map that are devoted to a single frequency or pitch. A second extension is that a top-down filter encodes the "expectations" that the pitch categories learn to expect. Each expectation codes the harmonics of the pitch that is represented by the pitch category. A spectral-pitch resonance can occur when a bottom-up adaptive filter and its top-down expectation generate a focus of attention. Such a resonance represents an auditory stream in the model.

How can multiple streams simultaneously be experienced? Because of their spatial relationships, the multiple spectral and pitch representations can naturally interact across the network via cooperative and competitive feedback interactions that resonantly capture harmonic frequencies that belong to the same pitch source within a given stream, while enabling multiple streams to coexist simultaneously. In particular, exclusive allocation (Bregman, 1990), or the allocation of a frequency to only one stream, can arise as follows: When a given pitch captures its harmonics via a spectral-pitch resonance, these harmonics at the spectral level can inhibit the same harmonics within other streams using within-frequency cross-stream lateral inhibition. This competition prevents the other streams from representing these frequencies.

Other streaming data have also been simulated by the ARTSTREAM model, such as the auditory continuity illusion and how gliding frequencies, with and without noise, interact when they come together. The auditory continuity illusion (Miller & Licklider, 1950) is a classical example of streaming wherein properties of the ART matching rule fairly leap out from the page. Suppose that a steady tone shuts off just as a broadband noise turns on. Suppose, moreover, that the noise shuts off just as the tone turns on once again. When this happens under appropriate conditions, the tone seems to continue through the noise, which seems to occur in a separate auditory stream. This example suggests that the auditory system can actively extract those components of the noise that are consistent with the tone and use them to track the "voice" of the tone right through the noise.

Suppose, however, that the tone does not turn on again for a second time. Then the first tone cannot continue through the noise to the other end. It is perceived to stop before the noise stops. A comparison of these two cases raises the question: How does the brain use the information about a future event, the second tone, to continue the first tone through the noise? This seems to require that the brain can operate “backwards in time” to alter its decision as to whether or not to continue a past tone through the noise based on future events. The ARTSTREAM model proposes that this backwards effect is due to a spectral-pitch resonance between the spectral representation of the tone and its pitch category; see Figure 8b. Such a resonance takes awhile to develop after the first tone occurs, but it takes much less time for the second tone to re-excite it once it has already begun, much as in the case of the second /b/ during resonant fusion of [ib]-[ba]. The third property of the auditory continuity illusion shows that the ART matching rule is obeyed here, and exhibits the same computational property whereby it helped to explain phonemic restoration. In particular, suppose that no noise occurs between two temporally disjoint tones. Then, unlike the case when broadband noise separates the two tones, the tone is not heard across the silent interval. Instead, two temporally disjoint tones are heard. This case shows that the brain actively uses the noise to continue the tone through it. The matching law can thus select a noise component that is consistent with the pitch category, but it cannot create a spectral sound unless there is already pitch-consistent activation at the spectral level that was caused by bottom-up inputs.

11. Distinct but Interacting Streaming and Phonetic Processes

Taken together, the ARTSTREAM and PHONET/ARTWORD models support the hypothesis that auditory streaming and phonetic processes are distinct. Streaming includes the setting up of spectral-pitch resonances, whereas phonetic processing generates (working memory)-(list chunk) resonances in a different part of the brain. Due to the harmonic bottom-up and top-down filters that bind spectral components to pitch categories during auditory streaming (Figure 8a, level 7), the role of harmonics is more important during auditory streaming than during phonetic perception, as has been experimentally demonstrated by Remez, Rubin, Berns, Pardo, & Lang (1994) and Remez, Pardo, Piorkowski, & Rubin (2001); also see Remez (2003) in this volume. Coordinating these several processing levels into a complete auditory architecture remains an open problem. One exciting property is that both types of processing seem to use resonant processes. They, as it were, speak the same dynamical language, and thus one can begin to analyse how they interact in a context-sensitive manner.

12. The Intimate Link between Auditory Information Processing and Learning in the Brain

In order for the auditory system to efficiently integrate information across its several processing levels, the mechanisms that occur at these levels need to be computationally consistent. ART predicts that there is a deeper reason why multiple levels of auditory processing may all use resonant dynamics. ART predicts that resonant processes enable our brains to continue to learn about a changing world in a stable fashion throughout life. These processes include the learning of top-down expectations, the matching of these expectations against bottom-up data, the focusing of attention upon the expected clusters of information, and the development of resonant states between bottom-up and top-down processes as they reach an attentive consensus between what is expected and what is there in the outside world. It is suggested that all conscious states in the brain are resonant states, and that these resonant states trigger learning of sensory and cognitive representations. The auditory models outlined above are proposed to be specialized versions of ART mechanisms for stably learning about temporally evolving auditory information about the world. Said in another way, ART clarifies how humans can so quickly learn to perceive and understand speech and language without experiencing catastrophic forgetting. This hypothesis suggests that experiments which attempt to probe the brain’s designs for speech and language should attempt, wherever possible, to link studies of information processing with manipulations of learning.

In addition to ART-based explanations of temporal data from audition and speech, variants of predicted ART top-down expectation, attentional-priming, and matching circuits have also been used to quantitatively simulate psychophysical and neurobiological data from early vision and visual object recognition; see Figure 2b. Given that all of sensory and cognitive neocortex shares key laminar circuit properties, it will be interesting to test the hypothesis that similar circuits, suitably specialized, may operate in auditory thalamocortical circuits.

Although ART mechanisms are predicted to occur in many brain processes, it is not proposed that they occur in all brain processes. In particular, the present article is devoted to sensory and cognitive processing in the What processing stream of the brain, which learns to recognize *what* objects and events occur in the world by using temporal cortex, among other brain regions (Ungerleider & Mishkin, 1982). A parallel Where cortical processing stream learns to spatially localize *where* objects are in the world and to act upon them, by using parietal cortex, among other brain regions (Goodale & Milner, 1992). These What and Where streams obey top-down matching and learning laws that are often *complementary* to one another (Grossberg, 2000b). This enables sensory and cognitive representations in the What stream to use their ART-like processing to maintain their stability as we learn more about the world, while allowing spatial and motor representations to forget learned maps and gains that are no longer appropriate as our bodies develop and grow from infanthood to adulthood. Detailed neural models of sensory-motor control and procedural memory clarify why procedural memories are not conscious. they often use inhibitory matching and learning processes that cannot lead to resonance, and hence cannot lead to consciousness (Fiala, Grossberg, & Bullock, 1996; Contreras-Vidal, Grossberg, & Bullock, 1997; Grossberg, Roberts, Aguilar, & Bullock, 1997; Bullock, Cisek, & Grossberg, 1998; Cisek, Grossberg, & Bullock, 1998). How to coordinate What and Where processing during auditory perception is another open problem that is ready for theoretical synthesis.

Acknowledgements:
Supported in part by the Air Force Office of Scientific Research (AFOSR F49620-01-1-0397), the National Science Foundation (NSF IRI-97-20333), and the Office of Naval Research (ONR N00014-01-1-0624). The author wishes to thank Susanne Daley for her valuable assistance in the preparation of the manuscript.

REFERENCES

Ahissar, M. & Hochstein, S. (1993) Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences*, 90, 5718-5722.

Baddeley, A.D. (1986) *Working memory*. Oxford: Clarendon Press.

Boardman, I. & Bullock, D. (1991) A neural network model of serial order recall from short-term memory. *Proceedings of the International Joint Conference on Neural Networks*, Vol. II: 879–884, Seattle, Washington. Piscataway, NJ: IEEE Service Center.

Boardman, I., Grossberg, S., Myers, C., & Cohen, M. (1999) Neural dynamics of perceptual order and context effects for variable-rate speech syllables. *Perception & Psychophysics*, 61, 1477-1500.

Bradski, G., Carpenter, G.A., & Grossberg, S. (1992) Working memory networks for learning temporal order with application to three-dimensional visual object recognition. *Neural Computation*, 4, 270-286.

Bradski, G., Carpenter, G.A., & Grossberg, S. (1994) STORE working memory networks for storage and recall of arbitrary temporal sequences. *Biological Cybernetics*, 71, 469-480.

Bregman, A.S. (1990) *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Britt, R. & Starr, A. (1976) Synaptic events and discharge patterns of cochlear nucleus cells: II. Frequency-modulated tones. *Journal of Neurophysiology*, 39, 179-194.

Bullier, J., Hupé, J.M., James, A., & Girard, P. (1996) Functional interactions between areas V1 and V2 in the monkey. *Journal of Physiology*, 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.

Caputo, G. & Guerra, S. (1998) Attentional selection by distractor suppression. *Vision Research*, 38, 669-689.

Cisek, P., Grossberg, S., & Bullock, D. (1998) A cortico-spinal model of reaching and proprioception under multiple task constraints. *Journal of Cognitive Neuroscience*, 10, 425-444.

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.

Cohen, M.A. & Grossberg, S. (1987) Masking fields: A massively parallel neural architecture for learning, recognizing, and predicting multiple groupings of patterned data. *Applied Optics*, 26, 1866-1891.

Cohen, M.A. & Grossberg, S. (1997) Parallel auditory filtering by sustained and transient channels separates coarticulated vowels and consonants. *IEEE Transactions on Speech and Audio Processing*, 5, 301-318.

Cohen, M.A., Grossberg, S., & Wyse, L.L. (1995) A spectral network model of pitch perception. *Journal of the Acoustical Society of America*, 98, 862-879.

Coleman, J. (2003) Discovering the acoustic correlates of phonological contrasts. *Journal of Phonetics*, 31, 000-000.

Contreras-Vidal, J.L., Grossberg, S., & Bullock, D. (1997) A neural model of cerebellar learning for arm movement control: Cortico-spino-cerebellar dynamics. *Learning & Memory*, 3, 475-502.

Delgutte, B. & Kiang, N.Y.S. (1984a) Speech coding in the auditory nerve: I. Vowel-like sounds. *Journal of the Acoustical Society of America*, 75, 866-878.

Delgutte, B. & Kiang, N.Y.S. (1984b) Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds. *Journal of the Acoustical Society of America*, 75, 879-886.

Downing, C.J. (1988) Expectancy and visual-spatial attention: effects on perceptual quality, *Journal of Experimental Psychology: Human Perception and Performance*, 14, 188-202.

Engel, A.K., Fries, P., & Singer, W. (2001) Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews: Neuroscience,* 2, 704-716.

Fiala, J.C., Grossberg, S., & Bullock, D. (1996) Metabotropic glutamate receptor activation in cerebellar Purkinje cells as substrate for adaptive timing of the classically conditioned eye-blink response. *Journal of Neuroscience,* 16, 3760-3774.

Gao, E. & Suga, N. (1998) Experience-dependent corticofugal adjustment of midbrain frequency map in bat auditory system. *Proceedings of the National Academy of Sciences*, 95, 12663-12670.

Goodale, M.A. & Milner, D. (1992) Separate visual pathways for perception and action. *Trends in Neurosciences*, 15, 10-25.

Grossberg, S. (1978a) A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In *Progress in theoretical biology, Vol. 5.* (R. Rosen & F. Snell, editors), pp. 233-374. New York: Academic Press. Reprinted in Grossberg, S. (1982), *Studies of mind and brain.* Kluwer/Reidel Press.

Grossberg, S. (1978b) Behavioral contrast in short-term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology,* 3, 199-219.

Grossberg, S. (1980) How does a brain build a cognitive code? *Psychological Review,* 87, 1-51.

Grossberg, S. (1984) Unitization, automaticity, temporal order, and word recognition. *Cognition and Brain Theory*, 7, 263-283.

Grossberg, S. (1986) The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In *Pattern recognition by humans and machines, Vol. 1: Speech perception* (E.C. Schwab & H.C. Nusbaum, editors), pp.187-294. New York: Academic Press.

Grossberg, S. (1999a) How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spatial Vision,* 12, 163-185.

Grossberg, S. (1999b) Pitch-based streaming in auditory perception. In *Musical networks: Parallel distributed perception and performance* (N. Griffith & P. Todd, editors), pp.117-140. Cambridge, MA: MIT Press.

Grossberg, S. (1999c) The link between brain learning, attention, and consciousness. *Consciousness and Cognition,* 8, 1-44.

Grossberg, S. (2000a) How hallucinations may arise from brain mechanisms of learning, attention, and volition. *Journal of the International Neuropsychological Society,* 6, 579-588.

Grossberg, S. (2000b) The complementary brain: Unifying brain dynamics and modularity. *Trends in Cognitive Sciences*, 4, 233-246.

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. & Kuperstein, M. (1986/1989) *Neural dynamics of adaptive sensory-motor control: Ballistic eye movements.* Amsterdam: North-Holland.

Grossberg, S. & Myers, C. (2000) The resonant dynamics of conscious speech: Interword integration and duration-dependent backward effects. *Psychological Review*, 107, 735-767.

Grossberg, S. & Raizada, R.D.S. (2000) Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research,* 40, 1413-1432.

Grossberg, S., Roberts, K., Aguilar, M., & Bullock, D. (1997) A neural model of multimodal adaptive saccadic eye movement control by superior colliculus. *The Journal of Neuroscience,* 17, 9706-9725.

Grossberg, S. & Stone, G. (1986a) Neural dynamics of attention switching and temporal-order information in short-term memory. *Memory & Cognition*, 14, 451-468.

Grossberg, S. & Stone, G. (1986b) Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review,* 93, 46-74.

Hawkins, S. (2003) Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31, 000-000.

Koch, C. & Ullman, S. (1985) Shifts in selective visual attention: Towards the under-laying neural circuitry. *Human Neurobiology,* 4, 219-227.

Krupa, D.J., Ghazanfar, A.A., & Nicolelis, M.A. (1999) Immediate thalamic sensory plasticity depends on corticothalamic feedback. *Proceedings of the National Academy of Sciences,* 96, 8200-8205.

Kunisaki, O. & Fujisaki, H. (1977) On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics,* 85-91.

Local, S. (2003) Variable domains and variable relevance: Interpretive phonetic experiments. *Journal of Phonetics*, 31, 000-000.

Luck, S.J., Chelazzi, L., Hillyard, S.A., & Desimone, R. (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology*, 77, 24-42.

Mann, V. & Repp, B. (1980) Influence of vocalic context on perception of the [zh]-[s] distinction. *Perception and Psychophysics*, 28, 213-228.

McClelland, J. & Rumelhart, D. (1981) An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375-407.

Mendelson, J., Schreiner, C., Sutter, M., & Grasse, K. (1993) Functional topography of cat primary auditory cortex: Responses to frequency-modulated sweeps. *Experimental Brain Research*, 94, 65-87.

Miller, G.A. (1956) The magical number seven plus or minus two. *Psychological Review*, 63, 81-97.

Miller, G.A. & Licklider, J.C.R. (1950) Intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, 22, 167-173.

Miller, J.L. & Liberman, A.M. (1979) Some effect of later-occurring information on the perception of stop consonant and semivowel. *Perception and Psychophysics*, 25, 457-465.

Moller, A.R. (1983) *Auditory physiology*. New York: Academic Press.

Mounts, J.R. (2000) Evidence for suppressive mechanisms in attentional selection: Feature singletons produce inhibitory surrounds. *Perception & Psychophysics*, 62, 969-983.

Parker, J.L. & Dostrovsky, J.O. (1999) Cortical involvement in the induction, but not expression, of thalamic plasticity. *Journal of Neuroscience,* 19, 8623-8629.

Pickles, J.O. (1988) *An introduction to the physiology of hearing* (2nd edition). San Diego: Acadmic Press.

Pollen, D.A. (1999) On the neural correlates of visual perception. *Cerebral Cortex*, 9, 4-19.

Raizada, R.D.S. & Grossberg, S. (2001) Context-sensitive binding by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast. *Visual Cognition*, 8, 431-466.

Remez, R.E. (2003) Establishing and maintaining perceptual coherence: Unimodal and multimodal evidence. *Journal of Phonetics,* 31, 000-000.

Remez, R.E., Pardo, J.S., Piorkowski, R.L., & Rubin, P.E. (2001) On the bistability of sine wave analogues of speech. *Psychological Science*, 12, 24-29.

Remez, R.E., Rubin, P.E., Berns, S.M., Pardo, J.S., & Lang, J.M. (1994) On the peceptual organization of speech. *Psychological Review*, 101, 129-156.

Repp, B.H. (1980) A range-frequency effect on perception of silence in speech. *Haskins Lab Status Report*, SR-61, 151-165.

Repp, B.H., Liberman, A.M., Eccardt, T., & Pesetsky, D. (1978) Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance,* 4, 621-637.

Rhode, W.S. & Smith, P.Y. (1986) Physiological studies on neurons in the dorsal cochlear nucleus of the cat. *Journal of Neurophysiology*, 56, 287-307.

Roelfsema, P.R., Lamme, V.A. & Spekreijse, H. (1998) Object-based attention in the primary visual cortex of the macaque monkey. *Nature,* 395, 376-381.

Rumelhart, D.E. & McClelland, J.L. (1982) An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review,* 89, 60-94.

Sachs, M.B. & Young, E.D. (1979) Encoding of steady state vowels in the auditory nerve: Representations in terms of discharge rate. *Journal of the Acoustical Society of America*, 66, 470-479.

Samuel, A.G. (1981) Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General,* 110, 474-494.

Samuel, A.G., van Santen, J.P.H., & Johnston, J.D. (1982) Length effects in word perception: We is better than I but worse than you or them. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 91-105.

Samuel, A.G., van Santen, J.P.H., & Johnston, J.D. (1983) Reply to Matthei: We really is worse than you or them, and so are ma and pa. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 321-322.

Schwab, E.C., Sawusch, J.R., & Nusbaum, H.C. (1981) The role of second formant transitions in the stop-semivowel distinction. *Perception and Psychophysics,* 21, 121-128.

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

Smith, A.T., Singh, K.D., & Greenlee, M.W. (2000) Attentional suppression of activity in the human visual cortex. *Neuroreport,* 11, 271-277.

Steinman, B.S., Steinman, S.B., & Lehmkuhle, S. (1995) Visual attention mechanisms show a center-surround organization. *Vision Research*, 35, 1859-1869.

Temereanca, S. & Simons, D.J. (2001) Topographic specificity in the functional effects of corticofugal feedback in the whisker/barrel system. *Society for Neuroscience Absracts*, 393.6.

Tian, B. & Rauschecker, J. (1994) Processing of frequency-modulated sounds in the cat's anterior auditory field. *Journal of Neurophysiology*, 71, 1959-1975.

Ungerleider, L.G. & Mishkin, M. (1982) Two cortical visual systems: separation of appearance and location of objects. In *Analysis of Visual Behavior*. (D.L. Ingle et al., editors), pp. 549-586. Cambridge, MA: MIT Press.

Vanduffel, W., Tootell, R.B. & Orban, G.A. (2000) Attention-dependent suppression of metabolic activity in the early stages of the macaque visual system. *Cerebral Cortex*, 10, 109-126.

Warren, R.M. (1984) Perceptual restoration of obliterated sounds. *Psychological Bulletin*, 96, 371-383.

Warren, R.M. & Sherman, G.L. (1974) Phonemic restorations based on subsequent context. *Perception and Psychophysics*, 16, 150-156.

Wheeler, D.D. (1970) Processes in word recognition. *Cognitive Psychology*, 1, 45-65.

Young, E.D. & Sachs, M.B. (1979) Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers. *Journal of the Acoustical Society of America*, 66, 1381-1403.

Zhang, Y., Suga, N., & Yan, J. (1997) Corticofugal modulation of frequency processing in bat auditory system. *Nature*, 387, 900-903.