Laminar cortical dynamics of conscious speech perception: Neural model of phonemic restoration using subsequent context in noise

Stephen Grossberg[a] and Sohrob Kazerounian
Center for Adaptive Systems, Boston University, 677 Beacon Street, Boston, Massachusetts 02215

(Received 10 February 2010; revised 12 December 2010; accepted 18 April 2011)

How are laminar circuits of neocortex organized to generate conscious speech and language percepts? How does the brain restore information that is occluded by noise, or absent from an acoustic signal, by integrating contextual information over many milliseconds to disambiguate noise-occluded acoustical signals? How are speech and language heard in the correct temporal order, despite the influence of contexts that may occur many milliseconds before or after each perceived word? A neural model describes key mechanisms in forming conscious speech percepts, and quantitatively simulates a critical example of contextual disambiguation of speech and language; namely, phonemic restoration. Here, a phoneme deleted from a speech stream is perceptually restored when it is replaced by broadband noise, even when the disambiguating context occurs after the phoneme was presented. The model describes how the laminar circuits within a hierarchy of cortical processing stages may interact to generate a conscious speech percept that is embodied by a resonant wave of activation that occurs between acoustic features, acoustic item chunks, and list chunks. Chunk-mediated gating allows speech to be heard in the correct temporal order, even when what is heard depends upon future context. © *2011 Acoustical Society of America*.
[DOI: 10.1121/1.3589258]

PACS number(s): 43.71.An, 43.71.Rt, 43.66.Ba, 43.71.Sy [MAH] Pages: 440–460

I. INTRODUCTION

The present article further develops the hypothesis that conscious speech percepts are emergent properties that arise from resonant states of the brain (Boardman *et al.*, 1999; Grossberg, 1978b, 1986, 2003; Grossberg and Myers, 2000). 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. The top-down expectations carry out a matching process that selects those combinations of bottom-up features that are consistent with the learned prototype while inhibiting those that are not. In this way, an attentional focus concentrates processing on those feature clusters that are deemed important on the basis of past experience. The attended feature clusters, in turn, reactivate the cycle of bottom-up and top-down signal exchange. This reciprocal exchange of signals equilibrates in a resonant state that binds the attended features together into a coherent brain state. Such resonant states, rather than the activations that are due to bottom-up processing alone, are proposed to be the brain events that regulate fast and stable learning of speech and language and that give rise to conscious speech and language percepts. The feedback dynamics of these resonances enable the brain to incorporate both past and future contextual information, often acting over hundreds of milliseconds, into the processing of speech and language, without destroying the correct temporal order of consciously heard words. Such contextual disambiguation is necessary to understand speech and language during the multi-speaker noisy environments that are characteristic of real-life speech and language experiences.

A classical example of a percept in which future context disambiguates consciously heard speech is phonemic restoration (Samuel, 1981; Warren, 1970, 1984; Warren and Obusek, 1971; Warren and Sherman, 1974; Warren and Warren, 1970). The current model explains how a hierarchy of laminar cortical processing stages, gated by the basal ganglia, can explain this and related speech percepts wherein conscious percepts depend upon contextual information. Phonemic restoration was chosen to illustrate model dynamics because it emerges from fundamental processes of speech perception in a vivid way and has not been explained by alternative models.

To articulate the relevant conceptual issues, consider the following example of phonemic restoration. Suppose broadband noise replaces the phonemes /v/ and /b/ in the words delivery and deliberation, respectively. Despite the initially ambiguous initial portion of these words ("deli-"), if the broadband noise is immediately followed by "ery" or "eration," listeners hear the /v/ or /b/ as being fully intact and present in the signal. Such experiences show that top-down lexical influences contribute to the formation of conscious speech percepts.

To explain such percepts, we need to understand why the noise in "deli-noise-[ery/eration]" is not heard before the last portion of the word is even presented. This may be explained by the fact that, if the resonance has not developed fully before the last portion of the word is presented, then this portion can influence the expectations that determine the conscious percept. How then, does the expectation convert the noise in "deli-noise-[ery/eration]" into a percept of [/v/ -/b/]? This occurs due to the top-down matching process that

[a]Author to whom correspondence should be addressed. Electronic mail: steve@bu.edu

 0001-4966/2011/130(1)/440/21/$30.00

selects expected feature clusters for attentive processing while suppressing unexpected ones. In the "deli-noise-[ery/eration]" example, spectral components of the noise are suppressed that are not part of the expected consonant sound. It has elsewhere been mathematically proved that the properties of this top-down attentive matching process, called the ART Matching Rule, are necessary to enable fast learning without catastrophic forgetting (Carpenter and Grossberg, 1987). Thus, phonemic restoration illustrates attentive matching processes that enable speech and language to be learned quickly and stably.

This attentive selection process is not merely a process of symbolic inference. It directly influences phonetic percepts. For example, if a reduced set of spectral components is used in the noise, then a correspondingly degraded consonant sound is heard (Samuel, 1981).

A related question concerns how future events can influence past events without smearing over all the events that intervene. In particular, if the /v/ or /b/ in "delivery/deliberation" is replaced by silence, how is it that the silence is perceived as silence despite the fact the disambiguating cue would have influenced the percept were these phonemes to be replaced by noise? 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. The opposite concern is also of importance. How can sharp word boundaries be perceived even if the sound spectrum that represents the words exhibits no silent intervals between them? The current theory proposes that silence will be heard between words whenever there is a temporal break between the resonances that represent the individual words. In other words, just as conscious speech is a resonant wave, silence is a discontinuity in the rate at which this resonant wave evolves.

As noted above, the attentive resonance and matching processes that support phonemic restoration are necessary ones to enable speech and language to be learned quickly without forcing the non-selective forgetting of previously learned memories. Due to the critical role of resonances in mediating fast learning, the theory that tries to explain how such fast learning occurs is called Adaptive Resonance Theory, or ART (Grossberg, 1978b, 1986). Indeed, an analogous ART matching process seems to occur for the same reason in other perceptual modalities, notably vision (Bhatt *et al.*, 2007; Carpenter and Grossberg, 1987). Reviews of compatible perceptual and neurobiological data about the predicted link between attentive matching, resonance and learning can be found in Grossberg (2003a, 2003b), Grossberg and Versace (2008), and Raizada and Grossberg (2003).

Earlier ART models have been used to explain and simulate data about auditory streaming, speech perception and word recognition; e.g., Boardman *et al.*, 1999; Grossberg *et al.*, 1999; Grossberg *et al.*, 2004; Grossberg and Myers, 2000. However, these modeling advances did not succeed in explaining and simulating how contextual information could feed backward in time, as in the "deli-/v-b/-[ery/eration]" example, to select the correct completion of the /v/-/b/ phoneme *and* do so in a way that creates a percept of speech in its correct temporal order. The current model proposes a computationally precise and neurally testable answer to the following basic question: How does the future influence the past, yet enable speech and language to be consciously experienced in its correct temporal order?

In contrast, alternative models of speech and language have focused on recognizing the "correct" interpretation of an auditory stream, without explaining how conscious speech percepts and learning evolve autonomously in real time. The current model is contrasted with alternative models in Sec. VI.

The current model is called cARTWORD, for *conscious ARTWORD*, since its simulations provide examples of conscious speech percepts that build upon the ARTWORD model of Grossberg and Myers (2000). The ARTWORD model did not generate speech representations that map onto conscious speech percepts. In order to achieve the progress that is reported in this article, several innovations were embodied in cARTWORD:

First, it was necessary to model a hierarchy of neocortical processing stages from individual acoustic features to unitized representations of sequences, or lists, of such features, operating in real time.

Second, although our model is necessarily simplified, it predicts how all the neocortical regions in the model may be described as variations on a shared laminar cortical design. Indeed, it is well known that all granular neocortical areas share a common laminar circuit design of six primary layers of cells (see Raizada and Grossberg (2003) for a review). These laminar circuits are specialized to support such different processes as vision, visual object recognition, cognitive information processing, and speech and language. Earlier modeling work has shown how variations of this circuit design may be used to explain and predict challenging psychophysical and neurobiological data about vision and visual object recognition (e.g., Cao and Grossberg, 2005; Fang and Grossberg, 2009; Grossberg and Versace, 2008; Grossberg and Yazdanbakhsh, 2005) and cognitive information processing (Grossberg and Pearson, 2008). Here we show how variations of the *same* design can be used to process speech. This unity of processing allows a comparative analysis of how the brain is specialized for vision, speech, and cognition. Using a shared laminar design and its variations to support such different forms of biological intelligence also opens up the possibility of designing a unifying set of VLSI chips for biological intelligence.

Third, the model clarifies how top-down interactions between these laminar cortical circuits can coherently bind all of the processing stages into a *resonant wave* that represents the evolving conscious speech percept. A key theme here is that it is not sufficient to just choose the correct groupings, or list chunks, of speech features, as earlier articles have achieved; e.g., Grossberg and Myers (2000). In addition, one needs to explain how higher-order groupings can get resonantly bound to the correct lower-level feature representations, and in the correct temporal order, even when the selection of these feature representations depends on future context.

Fourth, given that there are multiple processing stages, each with their own top-down connections, the model needs to clarify why lower processing stages do not prematurely resonate before sufficient contextual information accumulates to generate correct representations of speech. The model clarifies how activation of higher-level groupings, or list chunks, can open processing *gates* that enable the entire hierarchy of processing stages—acoustic features, acoustic items, and list chunks—to resonate when sufficient context is available to choose the correct groupings of features for conscious perception. It is well-known that the basal ganglia carry out such gating operations at multiple levels of cortical processing (Bellmann *et al*., 2001; Brown *et al*., 1997; Damasio *et al*., 1980). The proposed interaction between list chunks and gates to release temporally evolving resonances may be interpreted in terms of known interactions of the prefrontal cortex with the basal ganglia (Hikosaka and Wurtz, 1989; Pasupathy and Miller, 2005). Brown *et al*. (1999, 2004) have modeled in greater detail how such frontal-basal ganglia interactions may regulate learning and performance of temporally organized behaviors, such as eye movements. Our current model simplifies this description but preserves its functional role.

Fifth, given that resonances are supported by positive feedback in multiple model cortical circuits, a mechanism is needed to distinguish between cases where top-down feedback matches bottom-up inputs versus cases where top-down feedback occurs with no bottom-up input support, or with mismatched bottom-up inputs. A resonance with a target cell should not develop if the cell receives no bottom-up input support. This property is realized by the fact that top-down matching circuits are *modulatory* on-center, off-surround networks. Due to the modulatory on-center, top-down feedback signals, in the absence of bottom-up inputs, can sensitize target cells but cannot fire them. When a bottom-up input does match such a top-down signal, the response of the cell can be amplified, consistent with resonance requirements. Such circuits have been proved capable of enabling fast learning without causing catastrophic forgetting (Carpenter and Grossberg, 1987, 1991). They have also succeeded in predicting behavioral and neurobiological data about top-down attention (e.g., see Grossberg and Versace, 2008, and Raizada and Grossberg, 2003, for reviews). Here it is shown how they can be coordinated within a cortical hierarchy.

Sixth, given that resonances are supported by positive feedback in multiple model cortical circuits, a mechanism is needed to prevent an established resonance from lasting indefinitely. Activity-dependent habituative gating processes limit the duration of a resonance in time, and enable a sequence of resonances to develop at appropriate times. Such activity-dependent habituative gating processes play a key role in explaining many data about the dynamic unfolding of brain processing, notably during cortical development (e.g., Grossberg and Seitz, 2007; Grossberg and Williamson, 2001) and auditory and visual perception (e.g., Francis *et al*., 1994; Grossberg and Myers, 2000). Here it is shown how habituative gating, acting within a hierarchy of cortical processing stages, can enable the correct temporal order of resonances to automatically develop through time. The model has been briefly reported in Kazerounian and Grossberg (2009a, 2009b, 2009c).

While attempts to distinguish between competing theories of speech perception have often relied on the ability of a model to explain the Ganong effect, the present paper focuses instead on phonemic restoration. This is because the phonemic restoration effect places more challenging constraints on any theoretical model. In the Ganong effect, an ambiguous phoneme can be heard differently depending on the lexical context in which it is used. For example, an acoustically ambiguous phoneme constructed by varying the voice onset timing on a /g/–/k/ continuum is likely to be heard as /k/ when followed by _iss, and as /g/ when followed by _ift (Connine and Clifton, 1987; Ganong, 1980). Without further constraints, the effect has been explained by models which argue that it is the result of top-down feedback from lexical representations, as well as models which argue it results from a decision process receiving feedforward information only (Gow *et al*., 2008; Magnuson *et al*., 2003; McClelland and Elman 1986; Norris *et al*., 2000; Pitt and Samuel, 1995). In addition to this, the phonemic restoration phenomenon requires that the speech system perceptually synthesize not simply an ambiguous phoneme but a phoneme which is fully absent and replaced by broadband noise. It furthermore requires that this synthesis occur for phonemes that are replaced by noise, but not those that are replaced by silence. Thus the current model can explain Ganong-type data using mechanisms that can explain the more complex phonemic restoration data. Indeed, prior simulations using a simpler ART model than the one developed here have explained how future vowel context can disambiguate how an ambiguous earlier consonant is perceived (Grossberg et al., 1999).

Some psychologists believe that the Ganong effect is a very different phenomenon than phonemic restoration. We do not agree. Both may be explained by a backward effect in time whereby a future context disambiguates an earlier sound, and both require an explanation of why the disambiguated percept seems to unfold from past to future, despite the influence of future context. Indeed, phonemic restoration may be viewed as a limiting case of the Ganong effect, in which a very ambiguous phoneme, or one so ambiguous that it is totally uninformative in itself (i.e., noise), is disambiguated by future context. Conversely, the Ganong effect may be viewed as a case of phonemic restoration in the limit where the spectral characteristics of the broadband noise are altered to match those of an ambiguous phoneme. This latter case is a variant of phonemic restoration that was studied by Samuel (1981a, 1981b) in which the perceived reconstruction is a subset of the formants that are available for selection.

Analogously to the Ganong effect, Miller and Liberman (1979) showed that varying the duration of a subsequent vowel /a/ can alter the percept of a preceding consonant from /b/ to /w/. Here too, changing a future context may alter the percept of a previous, ambiguous, consonant. Boardman *et al*. (1999) used a simpler ART model than the one developed here to simulate how this percept could arise. Their model showed how the working memory codes that support

the changed percept are created, but the model was not sophisticated enough to actually simulate the unfolding in real time of that percept from disambiguated consonant to vowel. The current model is the first one that we know that can explain all of these types of effects in a unified way.

II. PHONEMIC RESTORATION

A. Empirical background

The first studies of phonemic restoration (Warren, 1970; Warren and Warren, 1970) were shown to occur when a phoneme such as the first /s/ in "legislatures" is excised, and replaced with a broadband noise, such as a cough. When replaced by noise, the excised phoneme was restored and perceived by listeners as being present and intact in the stimulus. When the phoneme was removed and simply replaced by silence, however, the silence gap was perceived and no such restoration occurred. In these initial studies, perceptual restoration was assumed to occur by virtue of the fact that subjects reported all phonemes in the sentence as being intact. In addition to not being able to recognize which phoneme was replaced by noise, subjects were unable to localize the position of the noise with respect to the speech signal by a median value of 5 phonemes.

Warren and Sherman (1974) later showed that the phoneme to be restored could be determined by subsequent context due to acoustic input arriving after the deleted phoneme. This study considered two words, "delivery" and "deliberation," which are contextually neutral until the /v/ or /b/. Before presentation of /v/ or /b/, the initial portions of the two words, "deli" are virtually indistinguishable and do not contain sufficient coarticulatory information to predict whether /v/ or /b/ will follow. After presentation of "deli*" (where * denotes noise), this speech segment was then followed by either "ery" or "eration." As noted above, presentation of "ery" resulted in the perceptual restoration of the phoneme /v/, whereas presentation of "eration" in the restoration of the phoneme /b/. The critical question arising from this study regards how future acoustical events interact with past stimuli to form conscious percepts in a manner whereby the disambiguating cue ("y" or "ation" in "delivery" and "deliberation," respectively) can influence earlier stimuli and can do so without destructive interference of intervening portions such as "er."

The ability of future events to influence conscious percepts of earlier arriving inputs is not unique to this paradigm. For example, it has been shown that increasing the silence duration between the words "gray chip" may result in the percept "great chip." Moreover, at appropriate noise durations of the fricative /ʃ/, listeners reliably perceive "gray" as "great" and "chip" as "ship" even at the highest tested silence durations of 100 msec (Grossberg and Myers, 2000; Repp et al., 1978). A related phenomenon is the Auditory Continuity Illusion. This illusion occurs when a steady tone occurs both before and after a burst of broadband noise which, under the appropriate temporal and amplitude conditions, results in a percept in which the tone appears to continue through the noise. The "backward effect in time" in this illusion is made clear by the fact that, without a subsequent tone following the noise burst, the tone does not continue through the noise. In the absence of noise, with a tone played before and after a silence duration, the silence duration is perceived. All of these effects are proposed to be due to resonant processes.

One of the primary technical concerns in studies of phonemic restoration is whether subjects consciously perceive the excised phoneme or if they simply respond that they do as a result of post-perceptual decision-making biases. In order to address the methodological difficulties in differentiating these two possibilities, Samuel (1981a, 1981b) used concepts from signal detection theory in order to obtain quantitative measures of the degree to which subject responses were due to true perceptual restoration, and the degree to which they were based on response strategies. Because subjects report perceiving noise superimposed over a fully intact word, Samuel was able to apply signal detection theory to test the ability of subjects to distinguish between stimuli in which phonemes were replaced by noise (the real stimuli in restoration phenomena), and those in which noise was simply added over the phoneme (the reported percept). Testing for these cases provided a d' value (discriminability index) and a β value (bias measure) which, respectively, measure the perceptual similarity of the two categories and the bias in responding for one or another category. A low d' measure suggests that subjects do indeed consciously perceive a missing phoneme as being present in cases where restoration occurs, insofar as the stimulus corresponding to the reported percept and the stimulus with the excised phoneme which results in restoration are perceptually indistinguishable. Testing a large variety of cases, Samuel was able to show that, in many cases of phonemic restoration, not only was the discriminability low (indicating true perceptual restoration), but that in many of these cases there was also no post-perceptual decision bias (subjects were not making decisions simply on the basis of some response strategy).

The variations in d' and β measures that Samuel did find were the result of a number of factors which influence the nature and strength of bottom-up and top-down interactions. Top-down effects included the finding that, when words were primed before presentation of a noise-replaced version, discriminability (d') was lower (perceptual restoration was stronger) than in unprimed trials. Another top-down effect was shown to occur for longer words. Namely, phonemes embedded in longer words resulted in stronger perceptual restorations than phonemes restored as parts of shorter words. Bottom-up influences were also found, such that fricatives and stops resulted in stronger perceptual restorations than other phone classes. These findings fit naturally within the framework of the model presented here, in which bottom-up and top-down interactions are central to the neural dynamics involved in conscious perception, working memory storage, and stable long-term memory formation.

A subsequent study by Samuel (1997) has shown that restored phonemes, in addition to being perceived as being present, are similar in other ways to phonemes which are truly present in the acoustic signal. In numerous other studies, it has been shown, for example, that repeated presentation of English words containing either a /b/ or a /d/ can cause reliable adaptation shifts in the perceived boundaries in a /bI/-/dI/

continuum. Critically, Samuel was able to show that adaptation shifts were still observed in response to words in which the /b/ or /d/ was perceptually restored after they were removed from the acoustic stimulus and replaced by noise. This study provides evidence for feedback from lexical to prelexical levels, a point to which we will return in considering alternative models of speech perception. It also suggests that, because restored phonemes behave like actual phonemes despite their absence from a speech signal, phonemic restoration is a true perceptual phenomenon rather than the result of a post-perceptual decision bias.

Further evidence for phonemic restoration as a true perceptual phenomenon comes from experiments of Kashino (2006). The stimuli in this study contained multiple deletions of the speech signal "Do you understand what I am saying" such that the signal would alternate with silence intervals every 50 msec, 100 msec, or 200 msec. In these examples, stimuli with silence intervals not filled with broadband noise sound disjoint and are either difficult or impossible to understand. Filled with broadband noise, however, the speech signal immediately and without effort sounds more natural and continuous and becomes easily understandable. Because the earliest studies relied on just a single excised portion of the speech signal, the utterance remained relatively intact and understandable regardless of whether or not noise was presented in place of the silence, making response strategies susceptible to bias. In these stimuli, however, it would be very difficult to alter response strategies since a subject cannot force understanding of an otherwise unintelligible acoustic stream.

III. THE cARTWORD MODEL

A. Stages of processing

In the cARTWORD model (Fig. 1), lower processing levels are responsible for early auditory processing of peripheral inputs, such as acoustic features and phoneme-like items), whereas higher levels process increasingly compressed and context-sensitive global representations, such as lexical entries, or list chunks, comprised of a sequence of acoustic items from lower levels. The circuits at each level of this hierarchy are comprised of neurons across multiple cortical layers, wherein neurons in the deep layers (layers 6 and 4) carry out filtering and temporary storage of incoming features, while neurons in superficial cortical layers (layers 2/3) group these features into unitized representations.

More specifically, acoustic inputs are presented in real time to the neurons at lowest level of the model, which selectively encode particular acoustic features (Fig. 1, lower cortical area, layers 6 and 4). The pattern of activity across feature detectors within a prescribed time interval activates a compressed acoustic item representation, or *item chunk* (e.g., phoneme) (Fig. 1, lower cortical area, layers 2/3). As a sequence of item chunks becomes active, it is input to, and stored by, a cognitive working memory (Fig. 1, upper cortical area, layers 6 and 4). The working memory hereby transforms a sequence of sounds into an evolving spatial pattern of activity that encodes both the items that occurred, and their temporal ordering, with the most active stored items performed first (Bohland

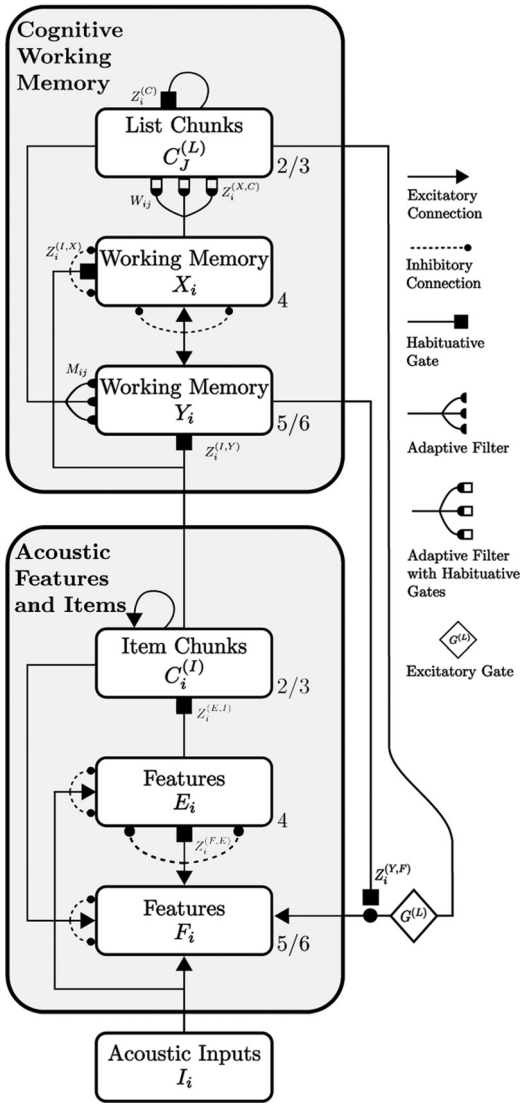

FIG. 1. Macrocircuit of the cARTWORD model. This macrocircuit shows a hierarchy of levels responsible for the processes involved in speech and language perception. Each level is organized into laminar cortical circuits, wherein deep layers (6 and 4) are responsible for processing and storing inputs, and superficial layers (2/3) are proposed to group distributed patterns across these deeper layers into unitized representations. The lowest level is responsible for processing acoustic features (cell activities F_i and E_i) and items [cell activities $C_i^{(I)}$], whereas the higher level is responsible for storing of sequences of acoustic items in working memory (activities Y_i and X_i), and representing these stored sequences of these items as unitized, context-sensitive representations by list chunks [activities $C_J^{(L)}$] in a masking field.

et al., 2010; Bullock and Rhodes, 2003; Bradski, Carpenter, and Grossberg, 1994; Grossberg, 1978a, 1978b; Grossberg and Pearson, 2008; Houghton, 1990; Lashley, 1951; Page and Norris, 1998a, 1998b). As the spatial pattern of stored items in working memory unfolds in time, it activates a network called a *masking field* (Cohen and Grossberg, 1986, 1987; Grossberg, 1978b; Grossberg and Myers, 2000). A masking field is a recurrent on-center off-surround network whose cells interact within and between multiple spatial scales, with the cells within larger scales capable of selectively representing item sequences of greater length (e.g., syllables or words), and of inhibiting cells that represent item sequences of lesser length. These cells are called *list chunks* (Fig. 1, upper cortical

area, layers 2/3) because each of them is a unitized, or chunked, context-sensitive representation of a particular temporal sequence, or list, of acoustic items. Active list chunks do at least two things: They readout previously learned top-down expectations that are matched against actively stored item chunks in the working memory, at the same time that they open processing gates which enable top-down feedback from the working memory to interact with the acoustic feature and item layers. By enabling both of these feedback loops to fire, the entire hierarchical system can begin to resonate between the levels of individual features, item chunks, working memory, and list chunks. In addition, the resonating list chunks are read out to subsequent processes, such as those which lead to the naming of words and other list chunks.

B. Acoustic features activate item chunks

These processes can be defined more precisely using the following mathematical notation. The cortical circuit responsible for processing lower-order speech representations contains auditory feature neurons in cortical layers 6 and 4 (activities F_i and E_i, respectively, in Fig. 1; see Eqs. (4)–(9) of Sec. IV) which become excited in response to a sequence of incoming acoustic inputs I_i. Auditory feature neurons such as these have been reported, for example, in single cell recordings of cat auditory cortex by He *et al.* (1997), who found selective tuning of cells to noise bursts of either long or short duration. These experimentally reported feature detectors could potentially respond selectively to affricates such as "ch," which contain a brief fricative burst, and fricatives such as "sh," which contain longer durations of fricative noise.

Excitatory activities E_i of layer 4 feature detectors can activate compressed auditory item chunks in layer 2/3 [activities $C_i^{(I)}$; see Eq. (10)]. The item chunk activities generate output signals to the next cortical area, at which sequences of items will be stored and unitized into list chunks.

It is important to note that, while the acoustic items themselves often correspond to "phoneme-like" units, they may not correspond exactly to phonemes as used in the International Phonetic Alphabet. This is due to the fact that these cells undergo a process of self-organized learning. They are learned item categories which may resemble aspects of what we now call phonemes for pragmatic purposes.

While the nature of early auditory processing is clearly important in explaining speech perception, a full description of this multiple-stage process is beyond the scope of this article. Several of these processes have been analyzed in previous ART-based models, such as how vowels and consonants may be differently pre-processed (Cohen and Grossberg, 1997), how acoustic sources may be segregated and auditory objects formed (Grossberg, 2003; Grossberg *et al.,* 2004), how speaker-invariant speech representations may be created (Ames and Grossberg, 2009), how rate-invariant speech codes may arise (Boardman *et al.,* 1999; Grossberg *et al.,* 1997), and how the grouping of sensory inputs into auditory objects may influence phonemic restoration (Grossberg, 2003; Grossberg *et al.,* 2004; Shinn-Cunningham and Wang, 2008). For computational simplicity, the acoustic features and items detailed here are developed just enough to simulate key properties of phonemic restoration.

C. Item chunks are stored in working memory

As sequences of item chunks are activated, they input to the next cortical processing stage. Here, incoming item inputs are stored in an Item and Order Working Memory (Bradski, Carpenter, and Grossberg, 1994; Grossberg, 1978a, 1978b), which has subsequently often been called a Competitive Queuing model (Houghton, 1990). This type of working memory model has gradually supplanted working memory models in which items move in a series of storage slots as more items occur (Anderson and Bower, 1974; Atkinson and Shiffrin, 1968). In contrast, within an Item and Order working memory, each item chunk activates a content-addressable item representation, and sequences of such activations are transformed into an evolving spatial pattern of activity across a network of content-addressable representations that selectively code the stored events. In such a network, the stored working memory pattern represents both the items and the temporal order in which they occurred, with no need for event representations to move. Moreover, Grossberg (1978a, 1978b) proved mathematically how to design such a working memory to obey what he called the LTM Invariance Principle. This Principle ensures that novel sequences of items may be stored and chunked through learning (e.g., MYSELF) in a way that does not destabilize memories of previously learned chunk subsequences (e.g., MY, SELF, ELF). This sort of memory stability is necessary to learn any language. Such a working memory may be realized by a recurrent on-center off-surround network. In our laminar cortical model, this working memory is embodied within layers 6 and 4 of the upper cortical area (Fig. 2); cf., Grossberg and Pearson (2008).

The LTM Invariance Principle constrains the kinds of patterns that can be stored in working memory. The correct

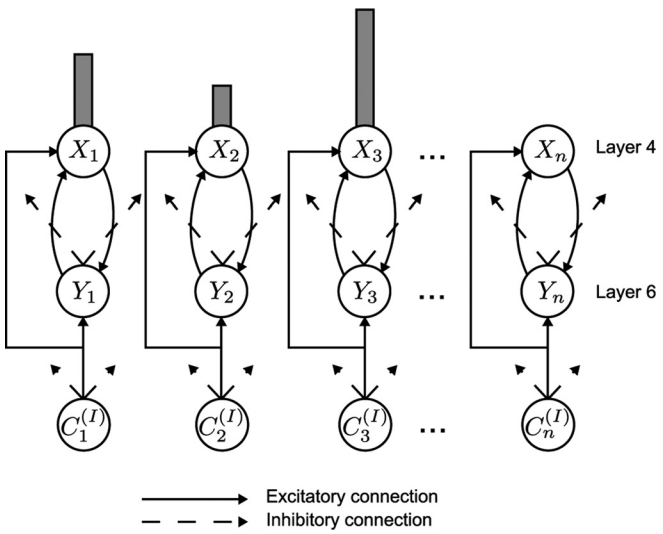

FIG. 2. For the sequence of acoustic items $C_1^{(I)}$, $C_2^{(I)}$, $C_3^{(I)}$, item and order are maintained in working memory by a primacy gradient of activity. The most active WM cell activity X_i, corresponds to the first item presented, the second most active corresponds to the second item presented, and so on.

temporal ordering is stored when early items achieve higher activation levels than later occurring items. This is called a *primacy gradient*. Grossberg (1978a, 1978b) showed how, in response to sufficiently short lists, either a primacy gradient or a *recency gradient*, in which the most recent items have the highest activity, can occur depending on the balance between the relative strengths of excitatory feedback to maintain stored activity, the strength of newly arriving inputs, and lateral inhibition between stored representations. In response to longer lists, a *bowed gradient* occurs, wherein early items exhibit a primacy gradient, and later items exhibit a recency gradient. In other words, only short lists can be recalled in the correct termporal order from short-term working memory to enable the stable learning and memory of list chunks whereby to remember these lists. Grossberg and Pearson (2008) review these issues.

Thus, in cARTWORD, an incoming temporal sequence of item chunks (from layer 2/3 acoustic item activities $C_i^{(I)}$ of the lower cortical area) is transformed into an evolving spatial pattern of activities that is stored across the neurons that form the working memory. These neurons occur in layers 6 and 4 of the working memory (activities Y_i and X_i; see Eq. (13)–(16); Figs. 1 and 2). The pattern of activation across layer 4 cell activities, X_i, then inputs to the list chunks [activities $C_J^{(L)}$; see Eqs. (17)−(20)], that occur in layer 2/3 in the upper cortical area.

The working memory stores item and order information using self-excitatory, or on-center, signals (the bi-directional excitatory pathways between cell activities X_i and Y_i in Eqs. (13) and (15)) which act to maintain a cell's activation after bottom-up input from an acoustic item category $C_i^{(I)}$ habituates. Competitive, or off-surround inhibitory, signals (the inhibitory pathways to X_i from all cells Y_k for $k \neq i$ in Eq. (15)), balance the self-excitatory signals with divisive normalization, which allows the network to preserve relative activations across items. Item and Order working memories are able to store order information in a primacy, recency, or bowed gradient by selecting parameters that balance between the strengths of newly arriving inputs [$2eC_i^{(I)}Z_i^{(I,Y)}$ in Eq. (13) and $2eC_i^{(I)}Z_i^{(I,X)}$ in Eq. (15)], the strength of excitatory feedback from already presented items [dX_i in Eq. (13) and eY_i in Eq. (15)], and the strength of the inhibitory off-surround ($\sum_{k\neq i}[2eC_k^{(I)}Z_k^{(I,X)} + eY_k]$ in Eq. (15)).

Items in working memory can be performed by activating a nonspecific rehearsal wave, predicted to be controlled by the basal ganglia, that allows all the cells within the working memory to begin emitting signals to the next processing stage. As the most active cell fires, it also activates a feedback signal that inhibits the corresponding working memory cell, thereby allowing the next cell to be selected, whence the process repeats. This inhibition-of-return readout mechanism was later called competitive queuing by Houghton (1990), referring to the competitive interactions which allow the most active item to be read out first and, once inhibited, allow the next most active item to read out, and so on. Supportive data from serial recall tasks have been reported, for example, by Farrell and Lewandowsky (2004) who wrote that "several competing theories of short-term memory can explain serial recall performance at a quantitative level. However, most theories to date have not been applied to the accompanying pattern of response latencies. Data from three experiments show that latency is a negative function of transposition displacement, such that list items that are reported too soon (ahead of their correct serial position) are recalled more slowly than items that are reported too late…these data rule out three of the four representational mechanisms. The data support the notion that serial order is represented by a primacy gradient that is accompanied by suppression of recalled items." In summary, multiple types of data support the use of an Item and Order working memory to temporarily store sequences of item chunks to set the stage for them to be unitized through learning into list chunks.

D. A masking field codes stored item chunk sequences as list chunks

A *masking field* is a recurrent shunting on-center off-surround network wherein unitized list chunks (activities $C_J^{(L)}$) selectively respond to sequences of item chunks of variable length that are stored in layer 4 cells (activities X_i) of working memory. The term list chunk is used, instead of lexical entry or word, because these representations can emerge through learning, and may represent phonemic, syllabic, or word representations.

Masking fields ensure that their list chunk cells are sensitive to, and respond selectively to, input sequences of variable length (Grossberg, 1978b; Grossberg and Myers, 2000). Each masking field cell responds optimally to a sequence of a prescribed length, so that a cell that is tuned to a sequence of length n cannot become strongly active in response to a subsequence of its inputs significantly less than length n. In other words, each cell accumulates evidence from its inputs until enough evidence has been sensed for the cell to fire. To accomplish this, the total input strength is normalized by the property of *conserved synaptic sites* (Cohen and Grossberg 1986, 1987), which states that cells which receive more input connections during development experience more activity-dependent growth during a period of endogenous activity of the input pathways. As a result, cells that receive more inputs grow larger and thereby dilute the effects of each input until a critical threshold is reached at which all inputs need to fire to activate the cell. This property is realized in Eq. (17), wherein the size of each bottom-up input $(70/|J|)X_iW_{ij}Z_i^{(X,L)}$ to activate list chunk activity $C_J^{(L)}$ varies inversely with the number $|J|$ of item chunks which input to that list chunk. Moreover, because masking field cells develop as a result of activity-dependent *self-similar* growth laws (Cohen and Grossberg 1986, 1987), as a cell grows larger, its inhibitory connections to other cells grow stronger too. As a result, a larger cell can inhibit cells that code subsequences of the inputs to which it is optimally tuned more than conversely. This property is realized by the asymmetric inhibitory coefficients of Eq. (17), in which the normalized inhibitory effect of a list chunk $C_K^{(L)}$ on another list chunk $C_J^{(L)}$ scales with the number of items $|K|$ which contact $C_K^{(L)}$, and the number of items $|K \cap J|$ which input to both chunks. Masking field list chunk cells that survive this asymmetric competition best represent the

sequence of item chunks that is currently stored in working memory.

E. Attentive matching, chunk-mediated gating, resonance, and conscious perception

After the masking field list chunks have been activated, both bottom-up and top-down interactions can occur. Bottom-up pathways [as filtered by weights W_{ij} in Eq. (17)] allow item or list chunk categories to be selectively activated, while top-down attentive pathways [via weights M_{ji} from the list chunk layer to the working memory activities Y_i of layer 6 in Eq. (13)] encode learned expectations whose prototypes can match, synchronize, and amplify the bottom-up distributed features to which attention is paid and can support resonant feedback and conscious speech percepts.

This resonant process occurs as follows: As the incoming sequence of acoustic item chunk activities $C_i^{(I)}$ is being stored in working memory, an evolving spatial pattern across layer 4 cell activities X_i begins to activate higher-order list chunk activities $C_J^{(L)}$ in layer 2/3 via the adaptive filter defined by the bottom-up weights W_{ij} [see Eq. (17)]. Once any of these list chunks receives sufficient bottom-up confirmatory evidence, active list chunks whose output signal functions $[C_J^{(L)} - \gamma_L]^+$ in Eq. (13) are positive, do two things:

First, they activate learned top-down expectations M_{ji} which select the components of bottom-up inputs, whether speech-like or broadband noise, that are consistent with their expected acoustic items as stored in working memory.

Second, they open a gate [term $G^{(L)}$ in Eq. (4)] which allows feedback from working memory (activities Y_i) to excite layer 6 acoustic feature cells (activities F_i). Acoustic feature activities are amplified in response to this feedback, then strongly activate acoustic item chunks [activities $C_i^{(I)}$ in Eq. (10)], thereby reactivating a positive feedback loop from those items back to their corresponding features (via the output signal $[C_i^{(I)} - \gamma_I]^+$ in Eq. (4)).

The resonant activations across the acoustic feature and item layers bind acoustic input into coherent perceptual groupings that are predicted to map onto a listener's conscious percepts. The time scale of these resonant dynamics enable backward effects to occur in time, as experienced during phonemic restoration. More specifically, while acoustic feature and item cells can rapidly respond to acoustic inputs, their storage and maintenance in working memory occurs on a slower time scale, as does the process by which competition across the masking field cells enables selection of the most predictive list chunk, thereby causing feedback from these layers to be delayed relative to the input. Because the acoustic feature and item resonance required to form a speech percept is coordinated, using the gate, with feedback from the cognitive working memory and masking field list chunks, the interactions between these multiple layers determine an emergent resonance time scale which reacts quickly enough to keep up with an incoming speech stream but slowly enough to allow future contextual information to influence it.

As noted above, another important factor in achieving sequential activation of resonant speech representations is the role of a habituative transmitter gating process whereby resonating circuits are desensitized in an activity-dependent way through time, thereby preventing perseveration of one item, and allowing the next item in a sequence to be experienced. See Eqs. (9), (11), (14), (16), and (18) for habituative gates operating at multiple stages of the model.

IV. RESULTS

A. Model simulations of phonemic restoration

The following conditions were simulated: First, presentation of a sequence of acoustic inputs under normal conditions activates the appropriate list chunk, and resonant activity across auditory features and items corresponds to the speech percept. Second, presentation of a sequence with a silence break in the input causes a corresponding break in the resonant activity between the auditory item and feature layers, leading to a conscious percept of silence. Third, when the silence interval is filled with broadband noise, the resonant wave of activity across the item and feature layers corresponds to a conscious percept of the full intact word with the missing phoneme now restored by properly timed resonant activity of the features and items corresponding to the missing phoneme. Last, it is demonstrated that such an excised phoneme can be restored on the basis of subsequent context alone. This is done by showing that, when multiple previously learned sequences have identical initial portions, input presentations which replace a medial portion of the word with noise give rise to perceptual restoration of the appropriate phoneme only after the final portion of the input sequence has been presented.

The acoustic features and items, as well as the learned sequences, were assumed to be generic, rather than language-specific auditory features, phonemes, and words. The model used five acoustic feature detectors, with five item categories, each coding for a single acoustic item from "1" to "5." The model was assumed to have seven learned list chunks which can be considered its lexicon. These include list chunks composed of the individual acoustic items "1" through "5," as well as list chunks for the learned sequences "1-2-3" and "1-4-5." List chunk representations such as "1-2" and "1-4" were left out because they could potentially bias the model toward restoration without necessarily reflecting the influence of future contextual information. Given that the primary focus of the model is to clarify the speech perception mechanisms that give rise to phonemic restorations, the structure chosen for the masking field reflects the most conservative allowable assumptions which can show that restoration necessarily occurs as a result of future context rather than past context information.

B. Normal condition

In the normal condition, a sequence of three acoustic feature cells was stimulated by back-to-back 50 msec square input pulses that selectively activated each of these cells. As the inputs arrive (the sequence of items "1," "2," and "3" are shown in blue, green, and red, respectively, in the bottom row of Fig. 3), they begin to excite acoustic feature cell activities, F_i and E_i (second and third rows from the bottom),

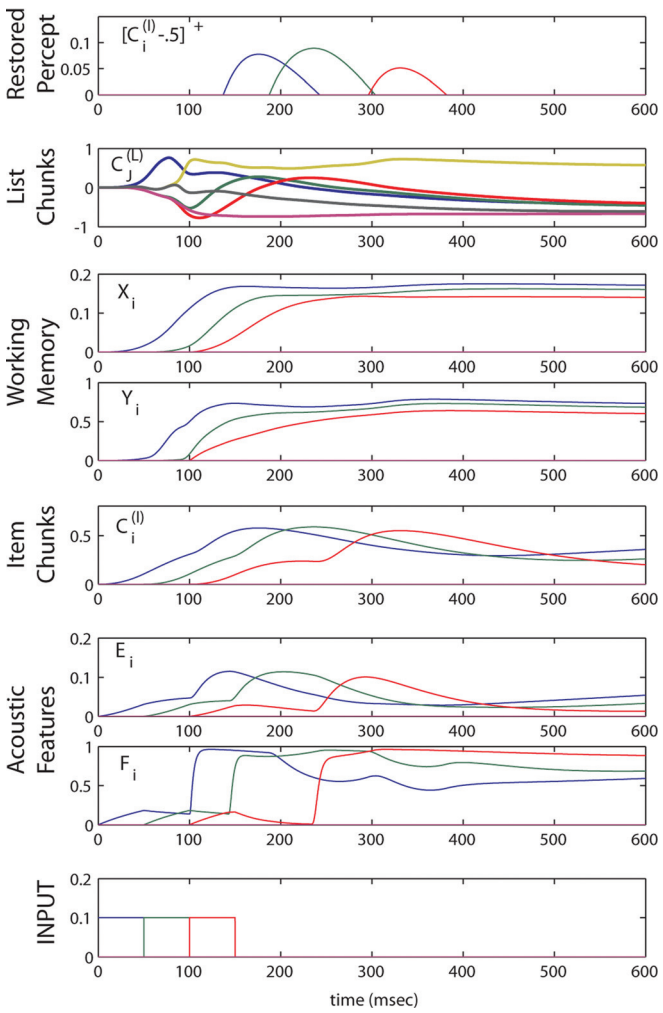

FIG. 3. Network dynamics in response to a sequence of three inputs "1-2-3" (shown in the bottom row as blue, green and red traces, respectively). The next two rows show the response of the acoustic feature layers F_i and E_i to this sequence of inputs. The fourth row from the bottom shows the activities of the acoustic item category cells $C_i^{(I)}$. These acoustic items are then stored in the cell activities Y_i and X_i (plots five and size from the bottom) in the cognitive working memory layers. The seventh plot from the bottom shows the response of list chunk activities $C_J^{(L)}$ in the masking field in response to the evolving pattern of activity in working memory. In this plot, the singleton list chunks coding for "1," "2," and "3" are shown in blue, green, and red, respectively, and the list chunks coding for "1-2-3" and "1-4-5" are shown in yellow and black, respectively. The top plot shows the super-threshold resonant activity of the acoustic item cells in response to the unbroken sequence of items "1-2-3" (shown in blue, green, and red, respectively). The resonant activity of these item cells reflects what is perceived under normal conditions, when a full sequence of acoustic input is presented to a listener.

which in turn excite their respective acoustic item category activities $C_i^{(I)}$ (fourth row from the bottom). The active items then begin to get stored by the cognitive working memory activities Y_i and X_i (fifth and sixth rows from the bottom). The working memory cells store both the items that were presented to the network and their temporal ordering. As these items are instated, the list chunk masking field activities $C_J^{(L)}$ (seventh row from the bottom) begin to respond to the evolving spatiotemporal pattern across the working memory. At first, the list chunk cell that codes for the single item list, "1" [the $C_J^{(L)}$ activity shown in blue], is most active, but it is quickly masked by the list chunk that codes for the sequence "1-2-3" [the $C_J^{(L)}$ activity shown in yellow] as inputs corresponding to the acoustic items "2" and "3" arrive.

The model's "conscious" speech code emerges from resonant feedback interactions that include the acoustic feature and item chunk layers, leading to a sequence of resonant activities corresponding to a percept of a unitized acoustic sequence. This can be seen in the top row, in which the super-threshold resonant activity of the acoustic item cells is shown. This plot shows the correct temporal order of resonant activation in the item/feature layers, as expected under conditions of normal presentation.

The bottom-up and top-down interactions which give rise to this resonant activity across the acoustic feature and item layers are shown in Fig. 4. As the masking field list chunks [activities $C_J^{(L)}$] compete in response to the pattern across working memory activities X_i, feedback from these layer 2/3 list chunks to layer 6 working memory activities Y_i (top left feedback loop of Fig. 4) selects and boosts the activities of those stored acoustic items which are consistent with previously learned expectations. Furthermore, once any list chunk receives sufficient bottom-up confirmation, gate opening via $G^{(L)}$ allows for the evolving pattern of activities Y_i in working memory to close a positive feedback loop with their respective acoustic features (shown in the loop on the right side of Fig. 4). This positive feedback loop subsequently causes the acoustic items [activities $C_i^{(I)}$] corresponding to the attended features to reach their resonant thresholds as well, allowing for positive feedback between acoustic items and features (shown in the bottom left feedback loop of Fig. 4). Resonant activations of acoustic features and items occur in a sequential manner as a result of habituation in synaptic pathways which prevents perservative supra-threshold activations from persisting due to a closed positive feedback loop.

C. Silence presentation

In the silence condition, in which a phoneme is excised from an acoustic stream and is not replaced by broadband noise, listeners report perceiving the break in the speech stream, and can often determine which phoneme was removed. The simulation inputs were presented to the acoustic features and items labeled "1" and "3" with a 50 msec silence duration corresponding to the excised phoneme "2" (bottom row of Fig. 5). The responses of the acoustic feature cells (activities F_i and E_i) as well as the acoustic item category cells [activities $C_i^{(I)}$] are shown in the next three rows from the bottom.

As the working memory cells begin to register activity from the acoustic item categories (activities Y_i and X_i, fifth and sixth rows from the bottom), the masking field list chunks (seventh row from the bottom) again respond to the evolving spatiotemporal pattern across the working memory. The masking field behavior shows that the list chunk that codes for the single item "1" [activity $C_J^{(L)}$ shown in blue] now competes more strongly with the list chunk for the whole sequence "1-2-3" (activity shown in yellow), as does the singleton list chunk which codes for the acoustic item

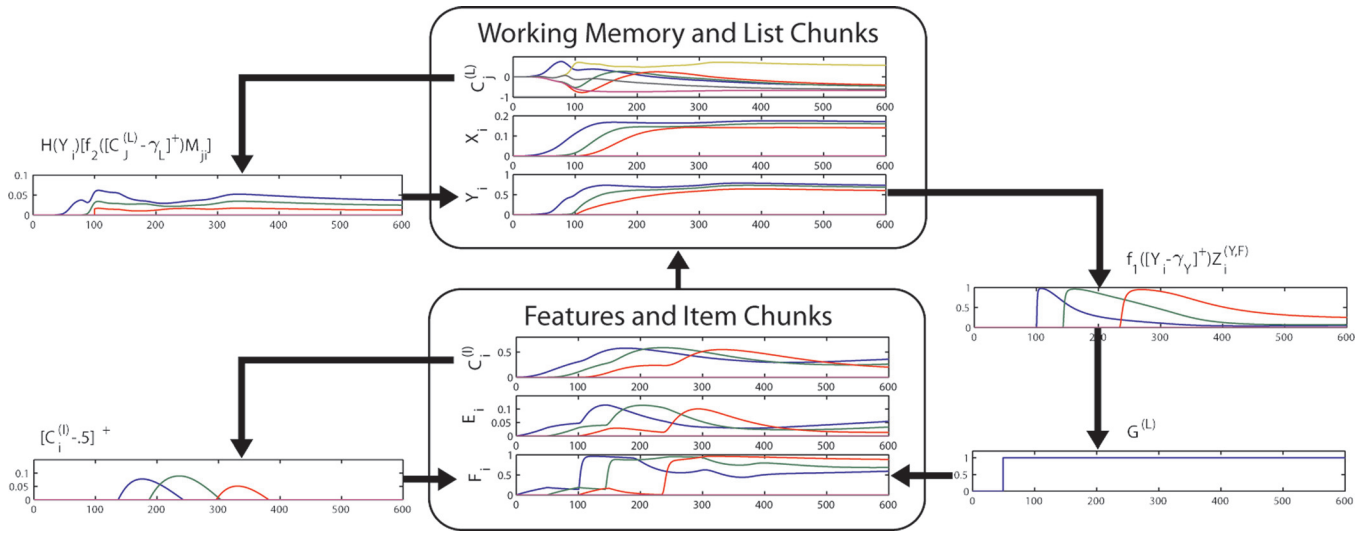

FIG. 4. This figure shows the bottom-up and top-down interactions which ultimately result in a wave of resonant activations across the acoustic feature and item layers. The activities in the acoustic features and items, as well as the working memory and masking field layers, are all shown, as are the top-down feedback signals from list chunk activities $C_J^{(L)}$ to layer 6 working memory activities Y_i (feedback path shown at the top-left), the top-down feedback signal from these working memory cell activities Y_i to F_i as gated by $G^{(L)}$ (feedback path shown to the right), and last the top-down feedback from acoustic item activities $C_i^{(I)}$ to their corresponding acoustic features F_i which drives the resonant activations across acoustic items and features (feedback path shown at the bottom-left).

"3" (activity shown in red), due to the fact that the list chunk coding for "1-2-3" never receives its expected bottom-up from the now excised acoustic item "2." Despite this increased competition, the "1-2-3" list chunk, by virtue of larger inhibitory masking coefficients due the larger size of the cell, is able to inhibit smaller list chunks, and as such becomes the most active cell.

Feedback from this layer 2/3 list chunk to the layer 6 working memory activities Y_i then selects and further excites the stored acoustic item activities corresponding to items "1" and "3." Because top-down feedback is modulatory, the working memory activities corresponding to the acoustic item "2" are not able to become excited in the absence of bottom-up input. As before, because competition across the masking field layer enables the selection, and sufficient activation, of a single list chunk, gate opening via $G^{(L)}$ enables feedback from supra-threshold working memory activities Y_i to excite their corresponding acoustic features items above their resonant thresholds. The resonant wave of activation (as shown by activities $[C_i^{(I)} - \gamma_I]^+$ in the top row of Fig. 5) contains a break between the supra-threshold activities of acoustic items corresponding to "1" and "3" which results from the intervening delay in the storage and, subsequently, the excitatory feedback, of these acoustic items in working memory. This break corresponds to the silence gap perceived by listeners in acoustic stimuli with an excised phoneme that is not replaced by noise.

D. NOISE presentation

1. Presentation of 1-*-3 yields 1-2-3 percept

To show that a phoneme can be restored in response to noise, the sequence "1-*-3" (where * denotes 50 msec of noise) is presented (bottom row of Fig. 6) and shown to give rise to resonant activity "1-2-3" (top row of Fig. 6), which includes the excised acoustic item "2." In order to simulate noise, at each time step of the numerical integration, a randomly chosen acoustic feature was stimulated. Given the integration rates of the various cells, this was equivalent to stimulating all the acoustic features at 1/5 the normal strength of input, because 5 acoustic feature cells were used in these simulations. Rows two and three from the bottom contain plots of feature cell activities F_i and E_i, and show that presentation of noise causes all the feature cells to become slightly active, and similarly cause all the acoustic item categories (fourth row from the bottom) to show a small response as well. Without some bottom-up input, these cells could not fire at all in response to top-down feedback, since the top-down excitatory feedback is modulatory. With it, the cells that get top-down feedback can be amplified, while those that do not can be inhibited by the off-surround of the top-down attentional network.

As the acoustic item cell activities $C_i^{(I)}$ become excited, they begin to register in working memory. After the first item "1" was unambiguously present in the input to working memory, all of the remaining working memory cells become active during the presentation of noise (fifth and sixth rows from the bottom). The activities $C_J^{(L)}$ of the "1-2-3" list chunk (shown in yellow in the seventh row from the bottom) and the "1-4-5" list chunk (shown in black) begin to increase in response to the ambiguous bottom-up input resulting from noise. Once the acoustic item "3" is presented to the network, beginning at 100 msec, the list chunk that codes for the sequence "1-2-3" is able to rapidly respond and overtake the list chunk coding for "1-4-5." Feedback from the list chunk coding for "1-2-3" is therefore able to select and boost its expected features in working memory activities Y_i of layer 6. As the pattern across working memory is thus corrected, gate opening via $G^{(L)}$ due to sufficient activation in the masking field layer allows corrected working memory

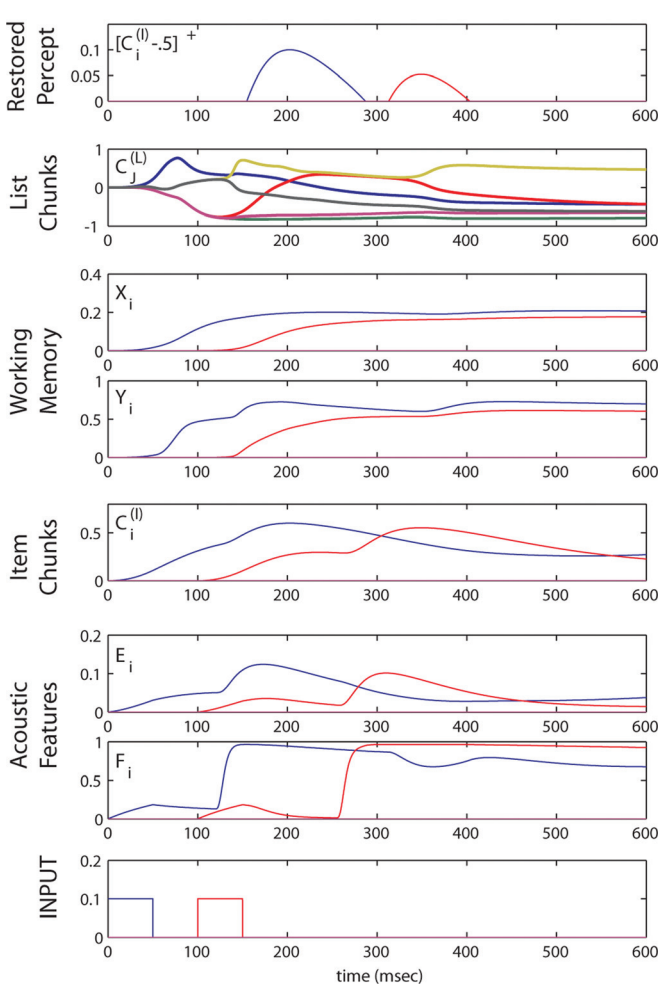

FIG. 5. Network dynamics in response to a sequence of three inputs presented "1-3" (bottom row, with "1" shown in blue and "3" in red), with a 50 msec silence duration interval. The plots in rows 2 and 3 from the bottom, show the response of the acoustic feature layers F_i and E_i. The fourth plot from the bottom shows the activities of the acoustic item category cells $C_i^{(I)}$. The activities of cells Y_i and X_i in the cognitive working memory layers (shown in the fifth and sixth plots from the bottom) respond to the incoming activity from the acoustic item layer. The seventh plot from the bottom shows the response of list chunk activities $C_J^{(L)}$ in the masking field in response to the evolving pattern of activity in working memory. As in Fig. 4, the singleton list chunks coding for "1," "2," and "3" are shown in blue, green, and red, respectively, and the list chunks coding for "1-2-3" and "1-4-5" are shown in yellow and black, respectively. The top plot shows the resonant activity across the acoustic item layer, and exhibits a break between the super-threshold activity of item cells "1" (blue trace) and "3" (red trace), corresponding to the silence perceived by listeners under these presentation conditions.

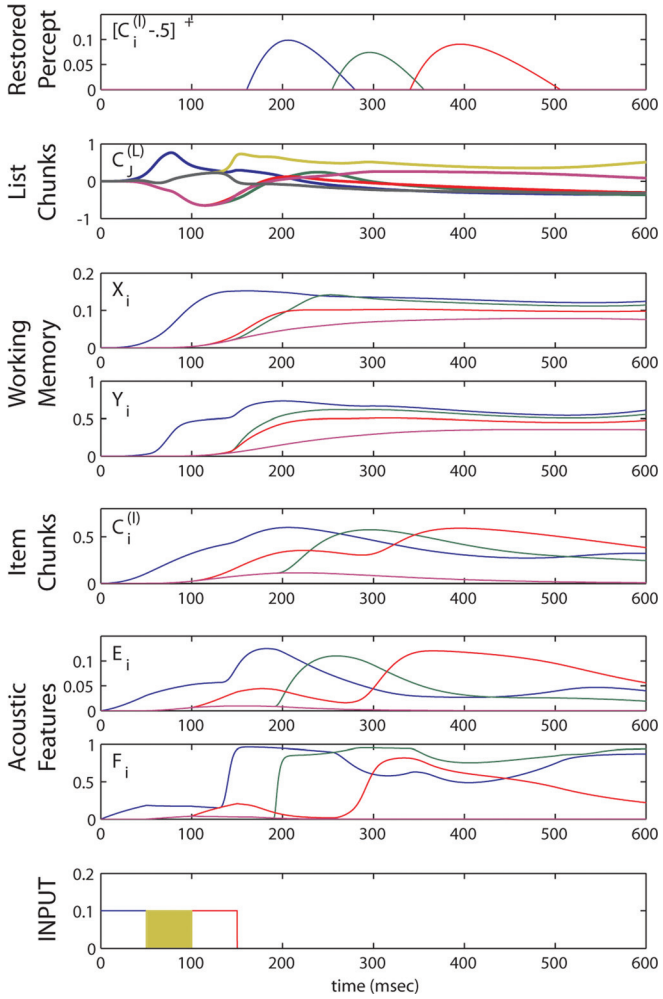

FIG. 6. Network dynamics in response to a sequence of three inputs presented "1- * -3" where "*" denotes noise as presented for 50 msec in place of any phoneme ("1" is shown in blue, "*" is shown as a filled yellow pulse, and "3" is shown in red). The bottom row shows presentation of the inputs, and the next two rows show the response of the acoustic feature layers F_i and E_i. The fourth plot from the bottom shows the activities of the acoustic item category cell activities $C_i^{(I)}$. The response of cells Y_i and X_i in the cognitive working memory layers in response to the incoming activity from the acoustic item cells, are shown in the fifth and sixth plots from the bottom. The seventh plot from the bottom shows the response of list chunk activities $C_J^{(L)}$ in the masking field in response to the evolving pattern of activity in working memory. As in Fig. 5, the singleton list chunks coding for "1," "2," and "3" are shown in blue, green, and red, respectively, and the list chunks coding for "1-2-3" and "1-4-5" are shown in yellow and black, respectively. Once the list chunk coding for "1-2-3" (the yellow trace) wins the competition with the "1-4-5" chunk (the black trace) upon unambiguous presentation of the acoustic item "3" at 100 msec, feedback from the chunk cells allows for the selection of and amplification of the components of noise consistent with its learned expectations, namely, the excised acoustic item "2" in working memory. Feedback from the working memory then drives acoustic features and items such that the resonant wave across these items (shown in the top plot) exhibits a continuous progression of resonant activity across "1," "2," then "3" (blue, green, and red traces, respectively), indicating that the excised item "2" has indeed been restored.

cell activities, having reached their output thresholds, to excite their corresponding acoustic feature cell activities F_i.

This excitatory feedback loop drives a resonant wave of activation across the attended features and their corresponding items, which show supra-threshold activations (top row of Fig. 6) corresponding to a listener's percept in cases when an excised phoneme has been replaced by broadband noise. Specifically, there is a smooth progression of resonant activity over the acoustic items and features, from "1" to "2" and then to "3," despite the fact that the stimulus for "2" had been removed from the input altogether. This simulation demonstrates how future contextual information (in this case, acoustic input "3" presented after the noise) is able to cause a wave of resonant activity in the correct forward sequential order while restoring the phoneme that was replaced by noise (in this case, acoustic item "2"). This is due to the fact that cells rapidly respond to bottom-up activations allowing for the selection of the correct list chunk (the

"1-2-3" list chunk shown in yellow, seventh row from the bottom) before the more slowly unfolding top-down feedback interactions drive acoustic features and items above their resonant thresholds in the correct order, modulated by activity-dependent habituation.

2. Presentation of 1-*-5 yields 1-4-5 percept

To test that subsequent context alone can determine which phoneme is restored, the next simulation presented the input sequence "1-*-5" to demonstrate that the correct item "4" is restored. Together, the simulations in Secs. IV D 1 and IV D 2 demonstrate how competing list chunks can sense noisy input sequences and restore correct missing phonemes "backward in time": Since these simulations used two competing chunks with the same initial portions, but different endings, the correct phoneme is restored entirely due to the future context that is provided by the final item.

In Fig. 7, as in Fig. 6, presentation of "1" (shown in blue in the bottom row), then noise, causes a similar response in the acoustic feature and item layers (second, third, and fourth rows from the bottom). When the input corresponding to "5" (shown in magenta, bottom row) is presented, this item causes the list chunk coding for the sequence "1-4-5" (shown in black, seventh row from the bottom) to become the most active, rather than the list chunk coding for "1-2-3" (shown in yellow) as in the previous simulation. As before, feedback from this layer to the cognitive working memory allows selection of expected features (the acoustic item "4" shown in cyan, rather than "2" as in Fig. 6) and subsequent feedback from the cognitive working memory to the acoustic network, causes the restoration of the acoustic item "4," which had been excised from the input. This can be seen in the top row, wherein resonant activities of the acoustic items follows a progression from item "1" to "4," and finally "5" (shown in blue, cyan, and magenta, respectively), as opposed to the simulation shown in Fig. 5, which exhibited a resonant wave across the items "1," "2," and "3."

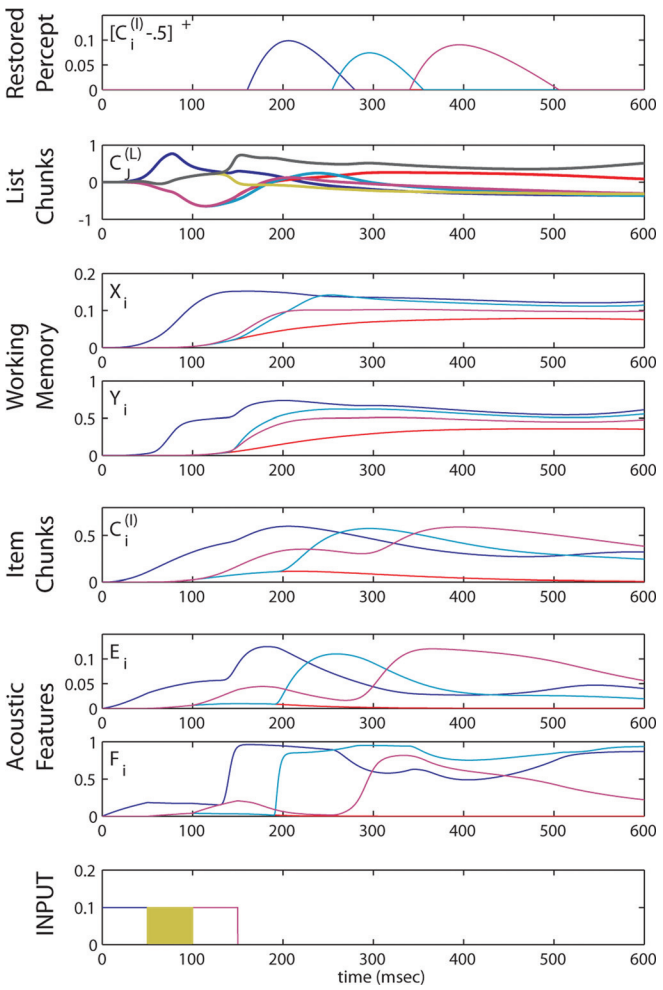

FIG. 7. This figure shows the network dynamics in response to the sequence "1- * -5", where * again denotes noise ("1" is shown in blue, "*" is shown in yellow, and "5" is shown in purple). The only difference between this simulation and that of Fig. 6 is the final item of the sequence, "5," which serves as future contextual information with respect to the excised phoneme, "4," which is to be restored. Rather than selection of the "1-2-3" list chunk (shown in yellow in the seventh plot from the bottom), presentation of the acoustic item "5" allows the "1-4-5" list chunk (shown in black) to win the competition across the masking field layer. Feedback from this chunk allows the selection and amplification of the components of noise consistent with its learned expectations, namely "4" (whose activity is shown in cyan in the working memory activities of Y_i and X_i). The feedback from working memory to acoustic features causes the super-threshold activity in the acoustic item layer (shown in the top plot) to exhibit a resonant wave in a continuous progression of activity across "1," "4," then "5" (blue, cyan, and magenta traces, respectively), indicating that the excised item "4" has indeed been restored. What is clear from this simulation is that the restoration occurs due to inputs arriving after the noise, just as the restoration cases with "delivery" and "deliberation."

V. MODEL EQUATIONS

The cARTWORD model is defined mathematically as a system of differential equations that describe how the activities of cells change in time. These equations also describe the habituation and recovery of the synaptic signaling strength of certain pathways which modulate the ability of cells to excite and inhibit one another. The cARTWORD model builds upon the ARTWORD model but goes considerably beyond it by embodying realistic laminar neocortical circuits, and a hierarchical cortical organization, gated by basal ganglia, that is capable of simulating the temporally evolving speech percepts that are heard during phonemic restoration and other context-sensitive percepts.

A. Cell membrane equations

The model is a network of interacting neurons whose cell dynamics obey membrane, or shunting, equations (Hodgkin and Huxley, 1952; Grossberg 1973). The single compartment voltage $V(t)$ of each cell obeys:

$$C_m \frac{d}{dt} V(t) = -[V(t) - E_{\text{leak}}]\gamma_{\text{leak}} - [V(t) - E_{\text{excit}}]\gamma_{\text{excit}}(t) - [V(t) - E_{\text{inhib}}]\gamma_{\text{inhib}}(t). \quad (1)$$

In (1), $V(t)$ is a variable voltage; C_m is a constant membrane capacitance; E_{excit}, E_{inhib}, and E_{leak} represent excitatory, inhibitory, and passive reversal potentials, respectively, that define shunting, or automatic gain control, properties of each cell; and the term γ_{leak} represents a constant leakage conductance, while the terms $\gamma_{\text{excit}}(t)$ and $\gamma_{\text{inhib}}(t)$ represent, respectively, the total excitatory and inhibitory inputs to the cell, as

determined by the architecture shown in Fig. 1. At equilibrium, the above equation can be rewritten as:

$$V = \frac{E_{\text{leak}}\gamma_{\text{leak}} + E_{\text{excit}}\gamma_{\text{excit}} + E_{\text{inhib}}\gamma_{\text{inhib}}}{\gamma_{\text{leak}} + \gamma_{\text{excit}} + \gamma_{\text{inhib}}}. \quad (2)$$

Increases in the excitatory and inhibitory conductances depolarize and hyperpolarize the membrane potential, respectively, which undergoes divisive normalization by all conductances (as shown in the denominator). When the reversal potential of the inhibitory channel is near the neuron's resting potential, the cell is said to undergo pure "shunting" inhibition (Borg-Graham, Monier and Fregnac, 1998). Equation (1) can be rewritten in the form:

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

where 0 is the passive equilibrium point, B (> 0) is the excitatory saturation point, and –C (≤ 0) is the inhibitory saturation point. Term P_i denotes the total excitatory input, and term Q_i is the total inhibitory input influencing cell activity x_i.

B. Acoustic feature/item layers

The acoustic feature/item network (lower cortical area of Fig. 1) consists of two deeper layers of interacting cells (layers 6 and 4) which contain the feature processing cells, and superficial layer 2/3 which contains the acoustic item category cells.

1. Acoustic feature network: Cognitively gated resonance

Acoustic feature processing occurs in recurrent on-center off-surround shunting networks (Grossberg, 1973, 1978b, 1980) in layers 6 and 4 that self-normalize their activities. Acoustic inputs I_i to the ith acoustic feature detector selectively activate the corresponding feature cell activities E_i in layer 4 and F_i in layer 6. The feature activities E_i then serve as bottom-up inputs to the acoustic item categories $C_i^{(I)}$ in layer 2/3.

a. Layer 6. Activity F_i of the ith layer 6 feature cell is described by the shunting recurrent on-center off-surround network:

$$\frac{dF_i}{dt} = -0.1F_i + (1 - F_i)\Big\{I_i + bE_i + [C_i^{(I)} - \gamma_I]^+ + 3G^{(L)}f_1([Y_i - \gamma_Y]^+)Z_i^{(Y,F)}\Big\} - 4F_i\left\{\sum_{k\neq i}[C_k^{(I)} - \gamma_I]^+\right\}. \quad (4)$$

Equation (4) contains a passive decay term ($-0.1F_i$). The total excitatory input $\{ I_i + bE_i + [C_i^{(I)} - \gamma_I]^+ + 3Gf_1([Y_i - \gamma_Y]^+) Z_i^{(Y,F)}\}$ is shunted by $(1 - F_i)$, thereby ensuring that activity remains bounded above by 1. Reading from left to right, the total excitatory input includes the bottom-up acoustic input I_i, which in these simulations are modeled as a series of 50 msec square pulses, each activating its corresponding acoustic feature activities F_i. The recurrent excitatory input bE_i from layer 4 feature cells (where $b = 0.2$) helps to maintain activity in layer 6 cells for a short while after acoustic inputs are removed. Without this feedback between layers 6 and 4, activities F_i would rapidly decay back to their resting potential. This feature persistence can last until top-down feedback from the list chunks via working memory occurs, so that feature-based resonances can develop.

Positive feedback from layer 2/3 auditory item category activities $[C_i^{(I)} - \gamma_I]^+$ is balanced against an inhibitory off-surround from the item category layer $\left[\sum_{k\neq i}[C_k^{(I)} - \gamma_I]^+\right]$. This top-down on-center off-surround network helps to boost the activities of matched features when resonance begins. The signal functions $[C_i^{(I)} - \gamma_I]^+$ and $[Y_i - \gamma_Y]^+$ in both feedback loops are linear above a threshold ("threshold-linear") with $[C_i^{(I)} - \gamma_I]^+ = \max[C_i^{(I)} - \gamma_I, 0]$ and $[Y_i - \gamma_Y]^+ = \max(Y_i - \gamma_Y, 0)$, where the output signal thresholds $\gamma_I = 0.5$ and $\gamma_Y = 0.5$.

In order to trigger sufficient activity for resonance to occur, feedback from the cognitive working memory cell activities $f_1([Y_i - \gamma_Y]^+)$ via the sigmoidal signal function:

$$f_1(w) = \frac{w^2}{0.01^2 + w^2} \quad (5)$$

is necessary. The feedback $f_1([Y_i - \gamma_Y]^+)$ is multiplicatively gated by a term $G^{(L)}$, as well as by a habituative synaptic strength, $Z_i^{(Y,F)}$. The gating term $G^{(L)}$ equals 1 when any list chunk cell activity $C_J^{(L)}$ in the masking field exceeds its output threshold (set to 0.2), and equals 0 otherwise. This gating property can be defined mathematically as follows:

$$G^{(L)} = H\left\{\sum_j \max\left[C_J^{(L)} - \gamma_L, 0\right]\right\}, \quad (6)$$

where the heaviside function $H(w) = 1$ if $w > 0$ and 0 otherwise, and threshold $\gamma_L = 0.2$. Thus, at least one list chunk needs to sufficiently match the sequence of acoustic items before feedback from list chunks can amplify the wave of activation evolving across the feature and item layers, thereby leading to a conscious resonance. The gating term $G^{(L)}$ simplifies the process whereby prefrontal cortical working memories interact with the basal ganglia to open gates that enable thalamocortical circuits to resonate and thereby express plans, thoughts, and actions. More detailed models of this gating process are found in Brown *et al.* (2004) and Grossberg and Pearson (2008).

Function $Z_i^{(Y,F)}$ in Eq. (4) describes the habituative synaptic strength, or habituative transmitter gate, of the pathway from working memory cell activity Y_i to feature activity F_i. It prevents perseveration of top-down feedback from acoustic item categories stored in working memory and thereby helps to coordinate the sequence of resonant activations by allowing a new item/feature resonance to take place after previously occurring one has habituated. This habituative synaptic strength, is defined as follows:

$$\frac{dZ_i^{(Y,F)}}{dt} = \varepsilon[1 - Z_i^{(Y,F)}] - Z_i^{(Y,F)}\{\lambda[Y_i - \gamma_Y]^+ + \mu([Y_i - \gamma_Y]^+)^2\}. \quad (7)$$

Equation (7) implies that the synaptic strength from Y_i to F_i recovers at a rate of ε until it reaches its maximal level, 1, due to the recovery term $\varepsilon[1 - Z_i^{(Y,F)}]$. As a signal $[Y_i - \gamma_Y]^+$ is sent along a pathway from the pre-synaptic to post-synaptic cell, its synaptic strength weakens at a rate determined by the strength of the signal and the parameters λ and μ, which specify linear and quadratic rates of activity-dependent habituation (Gaudiano and Grossberg, 1991; Grossberg and Myers, 2000). These linear and quadratic terms allow the gated signal $f_1([Y_i - \gamma_Y]^+)Z_i^{(Y,F)}$ emitted from the cell to exhibit a non-monotonic response, such that as signal $[Y_i - \gamma_Y]^+$ in (4) increases, the gated signal increases as well, until, at high enough $[Y_i - \gamma_Y]^+$ levels, it decreases. With only a linear term, the gated signal at equilibrium would be a monotonically increasing function of the input activity $[Y_i - \gamma_Y]^+$, and would require an external "supervisor" to manually shut off signals maintaining high activation levels. The parameters for all habituative gating equations were set to $\varepsilon = 0.01$, $\lambda = 0.1$, and $\mu = 3$.

The superscripts in the habituative strength $Z_i^{(Y,F)}$ in Eq. (7) denote the pathway along which synaptic strength habituates; that is, $Z_i^{(Y,F)}$ is the synaptic strength along the pathway from Y_i to F_i. Similarly, $Z_j^{(L)}$, in Eq. (19) below, is the synaptic strength of the self-excitatory path from $C_j^{(L)}$ to itself.

The inhibitory off-surround $\sum_{k\neq i}[C_k^{(I)} - \gamma_I]^+$ of Eq. (4) is derived from all supra-threshold acoustic item category activities, $[C_k^{(I)} - \gamma_I]^+$ for $k \neq i$ and is shunted by term $-4F_i$, thereby keeping the activity of the cell non-negative. Because inhibitory feedback from layer 2/3 activities $[C_k^{(I)} - \gamma_I]^+$ arrives only from supra-threshold acoustic items, this off-surround prevents the simultaneous resonant activation of multiple acoustic feature cells. This is due to the fact that, when any cell activity $[C_k^{(I)} - \gamma_I]^+$ reaches threshold, it strongly inhibits off-surround layer 6 feature cell activities F_i, until habituative synaptic strength in the bottom-up pathway [due to $Z_i^{(F,E)}$ in Eq. (8) and $Z_i^{(E,I)}$ in Eq. (10)] causes the currently supra-threshold cell to fall below threshold, thereby allowing the resonant activation of the next most active acoustic item category.

b. Layer 4. Activity E_i of the ith layer 4 feature cell is described by the recurrent shunting on-center off-surround network:

$$\frac{dE_i}{dt} = -0.1E_i + (1 - E_i)\left[eI_i + eF_iZ_i^{(F,E)}\right] - eE_i\left[\sum_{k\neq i}(I_k + F_k)\right]. \tag{8}$$

Equation (8) contains a passive decay term $(-0.1E_i)$, as well as a shunted excitatory input $\left[eI_i + eF_iZ_i^{(F,E)}\right]$ and inhibitory input $\left[\sum_{k\neq i}(I_k + F_k)\right]$. All excitatory and inhibitory inputs in Eq. (8) are scaled by the parameter $e = 0.05$. The two excitatory inputs are bottom-up acoustic inputs eI_i and recurrent excitatory feedback $eF_iZ_i^{(F,E)}$ from layer 6 cell activities that represent the same feature. Excitatory input from layer 6 cells is gated by each cell's habituative synaptic strength $Z_i^{(F,E)}$ to temporally limit activity persistence due to the positive feedback between layers 6 and 4:

$$\frac{dZ_i^{(F,E)}}{dt} = \varepsilon\left[1 - Z_i^{(F,E)}\right] - Z_i^{(F,E)}\left\{\lambda[F_i - \gamma_F]^+ + \mu([F_i - \gamma_F]^+)^2\right\}. \tag{9}$$

The signal function $[F_i - \gamma_F]^+ = \max(F_i - \gamma_F, 0)$ in Eq. (9) is threshold-linear, where $\gamma_F = 0.65$, and thus requires cell activity F_i to reach threshold 0.65 before the synaptic strength from F_i to E_i begins to habituate. As before, this habituative gate helps to prevent perseveration of resonant activity of acoustic features and item categories, since feature cell activities F_i which reach threshold quickly lose their ability to continue exciting cell activities E_i.

Off-surround inhibitory inputs $\sum_{k\neq i}(I_k + F_k)$ in (5) come from all the other bottom-up acoustic inputs I_k and the output signals of layer 6 cell activities F_k for $k \neq i$, as shown in Fig. 1 by the inhibitory connections arriving at layer 4 feature cells from the acoustic input as well as from layer 5/6 feature cells.

The off-surround input is shunted by the term $-eE_i$ which keeps the activity of the cell non-negative

c. Layer 2/3. Activity $C_i^{(I)}$ of the ith acoustic item category, or item chunk, cell is described by:

$$\frac{dC_i^{(I)}}{dt} = -0.1C_i^{(I)} + (1 - C_i^{(I)})\left\{2E_iZ_i^{(E,I)} + \psi f_2([C_i^{(I)} - \gamma_I]^+)\right\}. \tag{10}$$

Equation (10) contains a passive decay term $(-0.1C_i^{(I)})$ and a shunted excitatory term $\{2E_iZ_i^{(E,I)} + \psi f_2([C_i^{(I)} - \gamma_I]^+)\}$ whose bottom up excitatory input $2E_i$ from layer 4 cell activities is gated by the habituative synaptic strength $Z_i^{(E,I)}$:

$$\frac{dZ_i^{(E,I)}}{dt} = \varepsilon[1 - Z_i^{(E,I)}] - Z_i^{(E,I)}[\lambda E_i + \mu(E_i)^2]. \tag{11}$$

As in layer 4 cells, habituative input to layer 2/3 activities helps to prevent perseveration of supra-threshold resonant activations, by causing bottom-up gated signals from cell activities E_i to collapse at sufficiently high levels. Layer 2/3 cell activities $C_i^{(I)}$, also receive self-excitatory feedback activity $\psi f_2([C_i^{(I)} - \gamma_I]^+)$, thereby allowing cells which have reached threshold to maintain their supra-threshold activations for longer than they would be capable of if they only received bottom-up inputs from layer 4. The sigmoidal signal function in (10) is given by:

$$f_2(w) = \frac{w^2}{1^2 + w^2}, \tag{12}$$

The self-excitatory feedback term in (10) is scaled by the parameter $\Psi = 0.125$.

C. Cognitive working memory and list chunk network

The cognitive working memory (upper cortical area of Fig. 1) consists of two layers of interacting cells (layers 6 and 4) which together comprise the Item and Order Working Memory, and a third layer (layer 2/3) which contains the masking field list chunk network.

1. Item and order working memory

The sequence of auditory item chunk activities $C_i^{(I)}$ is stored as a primacy gradient of activation in the Item and Order Working Memory, which consists of a shunting recurrent on-center off-surround network between layers 6 and 4 of the cognitive working memory.

The ith auditory item chunk activity $C_i^{(I)}$ inputs to the ith layer 6 cell activity Y_i as well as the ith layer 4 cell activity X_i of the cognitive working memory. Layer 4 cells, in turn, excite masking field list chunk activities $C_J^{(L)}$ in layer 2/3.

a. Layer 6. Activity Y_i of the ith layer 6 cell obeys the shunting equation:

$$\frac{dY_i}{dt} = -0.1Y_i + (1 - Y_i)\Big[2eC_i^{(I)}Z_i^{(I,Y)} + dX_i + \eta H(Y_i)\{f_2([C_J^{(L)} - \gamma_L]^+)M_{ji}\}\Big]. \quad (13)$$

This equation contains a passive decay $(-0.1Y_i)$ term, and shunted excitatory input terms $[2eC_i^{(I)}Z_i^{(I,Y)} + dX_i + \eta H(Y_i)(M_{ji}f_2(C_J^{(L)} - \gamma_L^+))]$. The bottom-up excitatory input from auditory item chunk activities $2eC_i^{(I)}$ is gated by its habituative synaptic strength $Z_i^{(I,Y)}$, where:

$$\frac{dZ_i^{(I,Y)}}{dt} = \varepsilon\left[1 - Z_i^{(I,X)}\right] - Z_i^{(I,X)}\left\{\lambda C_i^{(I)} + \mu[C_i^{(I)}]^2\right\}. \quad (14)$$

This habituative gate limits the duration of item chunk inputs to working memory, and thereby also prevents them from strongly altering the spatial pattern of activations in working memory once acoustic features and item categories have reached their resonant thresholds. Input to layer 6 cells is also received from top-down intra-cortical feedback from the ith layer 4 cell activities dX_i, where $d = 0.7$. As discussed in Sec. III C, the on-center inputs dX_i in Eq. (13) and eY_i in Eq. (15), and off-surround inputs eY_k for $k \neq i$ in Eq. (15) allow these layers to achieve short-term memory storage of inputs presented to these layers, while meeting the constraints of the LTM Invariance principle.

Last, layer 6 cells receive top-down feedback from layer 2/3 list chunk activities $\eta H(Y_i)\{f_2([C_J^{(L)} - \gamma_L]^+)M_{ji}\}$ via the signal f_2 given in Eq. (12). Layer 2/3 feedback is multiplicatively gated by top-down adaptive weights, M_{ji}, which enable long-term memory traces of list chunks to be readout into working memory (cf. Grossberg and Pearson, 2008). This feedback is further gated by the heaviside function $H(Y_i)$, which ensures that top-down feedback from list chunks is modulatory. As a result, a list chunk cannot activate a working memory cell Y_i in the absence of its prior bottom-up activation. This feedback term allows previously learned expectations from list chunk cells to influence bottom-up activations being stored in working memory, such that once a given list chunk cell activity $C_J^{(L)}$ exceeds the threshold γ_L, it begins to influence active working memory activities Y_i via the adaptive filter defined by M_{ji}. The scaling parameter $\eta = 8$. For the purposes of these simulations, top-down weights M_{ji} were set equal to weights W_{ij} in the bottom-up adaptive filter [Eq. (17)], as would result from outstar and instar learning laws, wherein weights track post-synaptic and pre-synaptic activities, respectively (Grossberg, 1968, 1978b, 1980).

b. Layer 4. Activity X_i of the ith layer 6 cell obeys the shunting recurrent on-center off-surround equation:

$$\frac{dX_i}{dt} = -0.1X_i + (1 - X_i)\left[2eC_i^{(I)}Z_i^{(I,X)} + eY_i\right] - 1.25X_i\left\{\sum_{k\neq i}[2eC_k^{(I)}Z_k^{(I,X)} + eY_k]\right\}. \quad (15)$$

Equation (15) contains a passive decay term $(-0.1X_i)$, a shunted on-center excitatory term $[2eC_i^{(I)}Z_i^{(I,X)} + eY_i]$, and a shunted off-surround inhibitory term $\sum_{k\neq i}[2eC_k^{(I)}Z_k^{(I,X)} + eY_k]$. Bottom-up excitatory inputs arrive at the ith cell from the ith auditory item chunk category activity $2eC_i^{(I)}$, which is gated by $Z_i^{(I,X)}$, where:

$$\frac{dZ_i^{(I,X)}}{dt} = \varepsilon[1 - Z_i^{(I,X)}] - Z_i^{(I,X)}\{\lambda C_i^{(I)} + \mu[C_i^{(I)}]^2\}. \quad (16)$$

As in Eq. (13), this habituative gate prevents perseveration of acoustic item chunk inputs to the working memory. Bottom-up input X_i also arrives from the ith layer 6 cell activities eY_i that is part of the cognitive working memory feedback loop with layer 4.

Off-surround inhibitory inputs to X_i come from all other bottom-up inputs $2eC_k^{(I)}Z_k^{(I,X)}$ and layer 6 cell activities eY_k for all $k \neq i$. The parameters $e = 0.05$ and $2e = 0.1$ describe the relative strengths of bottom-up auditory item inputs and feedback inputs. A primacy gradient is achieved across the working memory layers by the relative strengths of the bottom-up and recurrent excitatory input parameters and the strength of the off-surround inputs in Eq. (15).

2. List chunk network

Item sequences stored in the cognitive working memory are categorized by list chunk cells in a masking field network within layer 2/3. The activity of a list chunk cell $C_J^{(L)}$ that codes the sequence J is defined by the shunting recurrent on-center off-surround network:

$$0.5\frac{dC_J^{(L)}}{dt} = -0.1C_J^{(L)} + [1 - C_J^{(L)}]\left\{\frac{70}{|J|}X_iW_{ij}Z_i^{(X,L)} + |J|f_2(C_J^{(L)}Z_j^{(L)})\right\} - [C_J^{(L)} + 1] \times \left\{\frac{\sum_K g(C_K^{(L)})|K|(1 + |K \cap J|)}{\sum_K |K|(1 + |K \cap J|)}\right\}. \quad (17)$$

Equation (17) contains a passive decay term $(-0.1C_j^{(L)})$, on-center shunted excitatory inputs $\{(70/|J|)X_iW_{ij}Z_i^{(X,L)} + |J|f_2(C_J^{(L)})Z_j^{(L)}\}$ and off-surround shunted inputs

$$\left\{\frac{\sum_k g(C_K^{(L)})|K|(1+|K\cap J|)}{\sum_k |K|(1+|K\cap J|)}\right\}.$$

Excitatory bottom-up inputs from the ith layer 4 cell activities $(70/|J|)X_iW_{ij}Z_i^{(X,L)}$, are filtered by bottom-up weights, or long-term memory traces, W_{ij}, which allow a list chunk to be selectively activated due to learning (not simulated here; see Cohen and Grossberg, 1987) and are normalized by a factor of $1/|J|$, which is inversely proportional to the number of inputs $|J|$ converging on list chunk, $C_J^{(L)}$, from the sequence J that is stored in working memory. As discussed in Sec. III D, the scaling of bottom-up inputs to list chunk cell size by $1/|J|$ normalizes the maximum total input to the cell using the property of conservation of synaptic sites. This property helps a masking field to maintain selectivity in response to sequences of different length, by preventing cells which code for lists of length n from becoming active in response to sequences much smaller than n.

Weights W_{ij} were set as follows: $W_{11}=0.1$, $W_{22}=0.1$, $W_{33}=0.1$, $W_{44}=0.1$, $W_{55}=0.1$, $W_{16}=0.15$, $W_{26}=0.1$, $W_{36}=0.05$, $W_{17}=0.15$, $W_{47}=0.1$, $W_{57}=0.05$, with all other values set to 0. These weights reflect a primacy gradient and are normalized such that each chunk receives the same total bottom-up weight, properties that would arise naturally from a normalized instar learning law whose weights track primacy gradient activitie across an Item and Order Working Memory (Grossberg, 1978b; Grossberg and Pearson, 2008).

The bottom-up input is also gated by a habituative synaptic strength $Z_i^{(X,L)}$ that is defined as:

$$\frac{dZ_i^{(X,L)}}{dt} = \varepsilon\left[1 - Z_i^{(X,L)}\right] - Z_i^{(X,L)}\left[\lambda X_i + \mu(X_i)^2\right]. \quad (18)$$

The other excitatory input term, $|J|f_2(C_J^{(L)})Z_j^{(L)}$, which results from activity dependent self-similar growth, describes the self-excitatory feedback activity of a list chunk onto itself. This self-excitatory feedback term is proportional to the number J of cortical inputs received by the list chunk, and further helps a masking field to achieve selectivity by providing a competitive advantage to cells which receive stored inputs from longer lists. The self-excitatory feedback signal function f_2 is defined in Eq. (12) above. The feedback is gated by the habituative transmitter $Z_j^{(L)}$, where:

$$\frac{dZ_j^{(L)}}{dt} = \varepsilon\left(1 - Z_j^{(L)}\right) - Z_j^{(L)}\left\{\lambda C_J^{(L)} + \mu\left[C_J^{(L)}\right]^2\right\}. \quad (19)$$

The inhibitory inputs to a list chunk $C_J^{(L)}$ are shunted by $[C_J^{(L)}+1]$, ensuring that activity remains above -1. In the inhibitory input

$$\left\{\frac{\sum_k g(C_K^{(L)})|K|(1+|K\cap J|)}{\sum_k |K|(1+|K\cap J|)}\right\},$$

J and K denote the sequences that activate $C_J^{(L)}$ and $C_K^{(L)}$, respectively, terms $|J|$ and $|K|$ denote the numbers of items in these sequences, and the term $|K\cap J|$ denotes the number of items that the two cells share. Thus, the inhibitory input to a cell $C_J^{(L)}$ from a neighboring cell $C_K^{(L)}$, is proportional to the signal $g(C_k^{(L)})$, where the sigmoid signal function g is defined by:

$$g(w) = \frac{w^2}{0.2^2 + w^2}, \quad (20)$$

the number of inputs $|K|$ that converge on $C_K^{(L)}$, and the number of inputs $|K\cap J|$ shared by $C_K^{(L)}$ and $C_J^{(L)}$. These inhibitory coefficients, which also describe self-similar competitive growth between list chunks, further provide masking field selectivity by allowing larger cells to more strongly inhibit smaller cells, with inhibition proportional to the number of items contacting a given list chunk. Shunting inhibition in the denominator of the inhibitory term results in divisive normalization such that the maximum total strength of inhibitory connections to each list chunk is equal to 1.

VI. DISCUSSION

Although various models of human speech perception have used phonemic restoration as a motivating factor in their creation, and even as evidence of their validity, we are not aware of any that have attempted to explicitly explain and simulate why and how phonemic restoration occurs (Elman and McClelland, 1986; Norris, 1994; Norris *et al.*, 2000). Indeed, although it is never simulated, phonemic restoration is claimed to be one of the primary motivations of the TRACE model. "We start with [the Ganong] phenomenon because it, and the related phonemic restoration effect, were among the primary reasons why we felt that the interactive-activation approach would be appropriate for speech perception as well as visual word recognition and reading," (Elman and McClelland, 1986, p. 24).

There are also some speech models that draw on knowledge of human speech perception in order to deal with how speech can be recognized when portions of speech are occluded by noise, or absent from the signal altogether. Some of these models have addressed the question of phonemic restoration (Masuda-Katsuse and Kawahara, 1999; Srinivasan and Wang, 2005), and produce a spectral representation of the speech signal with the appropriately restored phoneme. They do not, however, explain how phonemic restoration may arise in humans. Restoration in Masuda-Katsuse and Kawahara (1999), for example, uses a Kalman filter to track and predict the spectral envelopes of segmented speech streams, which then produce the output spectrum of the restored phoneme. The model does not use top-down lexical information. Conversely, while Srinivasan and Wang (2005) make use of lexical information to restore

a masked, but not missing, phoneme, it does not operate in real-time. Instead, multiple processing stages make use of Kalman filtering, Hidden Markov Modeling, and Dynamic Time Warping in order to predict, track, and reconstruct a phoneme occluded by noise, necessitating that these processing steps occur off-line in multiple passes over the input.

Prominent alternative models of human speech perception include the TRACE model (McClelland and Elman, 1986) and the MERGE model (Norris *et al.,* 2000). The cARTWORD model has conceptual and explanatory advantages over these models. Fundamental conceptual problems of these alternative models are summarized below. The explanatory advantages of cARTWORD include its use of neurobiological circuits whose variations have been used to successfully explain many kinds of data and that are based on a well-known laminar cortical organization. Explanatory advantages also include cARTWORD's ability to explain contextual effects that can operate over long time intervals, including the effects of future context, by using its resonance mechanisms. Another key advantage of cARTWORD is its ability to represent consciously perceived percepts and to explain how these percepts are related to mechanisms of attention, resonance, and learning. Although the model's perceptual representations are simplified, the working memory, list chunking, and resonance mechanisms of cARTWORD can be naturally extended to explain much more complex percepts as the complexity of feature preprocessing is extended, much as has already been done in models of visual perception.

A. TRACE

The TRACE model is based on the Parallel Distributed Processing framework of Rumelhart *et al.* (1986) and is closely related to the interactive activation model (IAM) of letter perception. The model uses a network of fully connected, yet simple, processing units. The activity of each of these units is governed by an activation function, and results in a spreading of activation to all the other units to which it is connected. TRACE embodies such general properties, however, in a way, that is, inconsistent with basic properties of human speech perception, or indeed of any real-time physical model. These include:

(1). Not a real time model. TRACE does not operate in real time. Indeed, it does not include a plausible representation of time that can be used in any physical process. Rather than treat time as an independent variable, it is treated as a structural variable used to create a series of "time slices" that are sequentially activated to represent a sequence of events. As a result, the model massively duplicates feature detector, phoneme, and word units, as well as their connectivity patterns. Every word and phoneme representation thus has a copy at every time slice, in striking contrast to the content-addressable unique representations in cARTWORD that may be activated at certain times. Aside from preventing the model from being able to recognize variable-rate speech data, it makes learning difficult since it is not clear how learning at a representation in one time slice should interact with a corresponding representation in a different time slice.

(2). Silence is not context-sensitive. Silence in the model is explicitly built in and is represented by a unit, or node that is activated in the absence of input. There are, however, many examples wherein perceived silence is context-sensitive and does not correspond to silent breaks in acoustic inputs. cARTWORD and its antecedents ARTWORD (Grossberg and Myers, 2000) and ARTPHONE (Grossberg et al., 1997) simulated such data as temporal breaks in the resonant wave that embodies conscious speech.

(3). Driving top-down feedback and unstable learning. TRACE does not implement the ART Matching Rule. The proposed alternative is that when "higher levels insist that a particular phoneme is present, then the unit for that phoneme can be activated… then the learning mechanism can 'retune' the detector." However, it has been mathematically proved that such a driving top-down feedback mechanism leads to unstable learning and memory (Carpenter and Grossberg, 1987; Grossberg, 1988). Indeed, behavioral, neurophysiological, and anatomical data support the proposal that top-down attention is modulatory, not driving, except when volition may alter top-down signals to induce visual imagery, fantasy, or internal planning (Grossberg, 2000, 2003; Raizada and Grossberg, 2003). Due to this driving property of TRACE top-down processing, over and beyond the lack of resonance as a mediating mechanism, TRACE cannot simulate phonemic restoration data. Specifically, it cannot explain how silence in a restoration condition remains silent and how a reduced set of spectral components in a noise input leads to a correspondingly degraded consonant sound (Grossberg et al., 1997; Samuel, 1981a, b).

A reviewer kindly sent us a simulation, and illustrative figure, to advance the claim that TRACE can simulate phonemic restoration, despite its incorrect form of top-down feedback. The simulation used a java implementation of TRACE known as jTRACE. Figure 8 depicts a simulation of jTRACE using parameters provided by the reviewer, as well a reconstruction of that simulation. This simulation depends upon another assumption of TRACE; namely, that silence activates a "silence node" that strongly inhibits all phonemic nodes. Such an assumption has no biological support. Moreover, it is incompatible with several types of data that ART can explain. For example, this hypothesis prevents TRACE from explaining percepts in which a physical silence is heard as sustained sound. Data of the latter kind have been simulated in earlier ART articles (e.g., Grossberg *et al.*, 1997) on which the current model builds. In contrast, ART predicts that silence is a temporal discontinuity in the resonant wave that represents conscious speech. Moreover, activity of a silence node would inhibit all primed activity during silent intervals in priming experiments, thereby undermining TRACE's ability to simulate RT data, among others, in such experiments. ART can accommodate priming data as well (e.g., Grossberg and Stone, 1986).

In the reviewer's simulation, jTRACE is presented with the word *luxury*, whose input representation is given by -

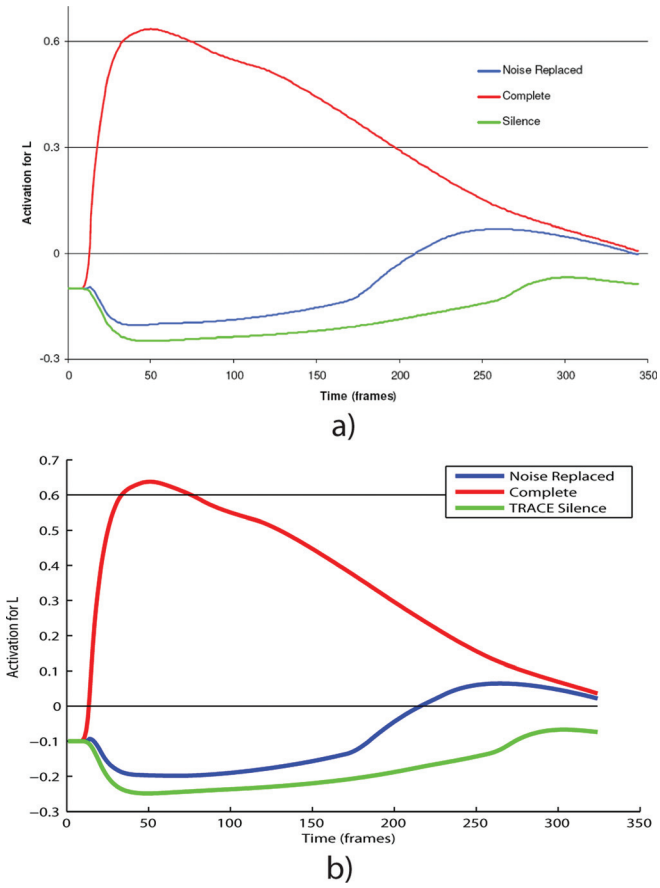

FIG. 8. This figure shows the activities of the /l/ phoneme in the jTRACE model, when presented with the word luxury under three conditions. In the normal condition, all the corresponding feature level inputs to the phonemes are present, whereas in the silence condition the feature inputs corresponding to /l/ are replaced with the feature inputs corresponding to the silence phoneme node. In the noise condition, a noise vector was created by setting all values of all features to 4. The plot shown in (a) was provided by one of the reviewers, and the plot shown in (b) is a recreation to be certain that subsequent plots are accurate.

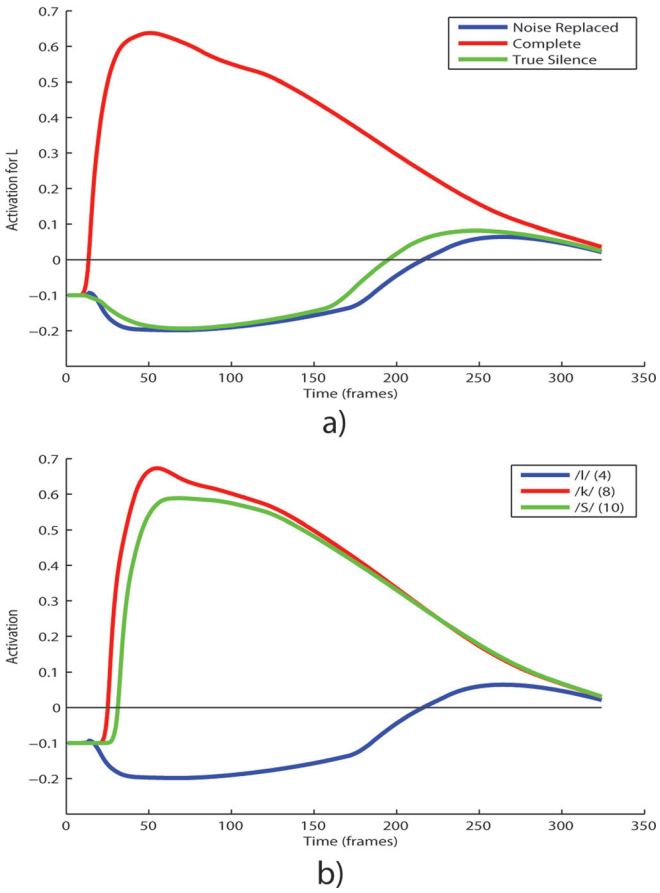

FIG. 9. (a) Recreates Fig. 8 for the normal and noise conditions. However, in the silence condition, rather than replace the feature level inputs corresponding to /l/ with the feature vector corresponding to the silence phoneme, the silence replacing /l/ in this figure was simply the absence of acoustic input. When silence is represented in this biologically relevant way, the activation of the /l/ phoneme node is earlier and higher than in the case when noise is presented. The main reason for this is that there is no competition from an artificial silence node, whose strong activation during silence presentations yields strong lateral inhibition with every other phoneme. Another reason is that competition due to activation by noise attenuates rather than facilitates activation of the /l/ phoneme node when compared to silence. This is in direct contrast to phonemic restoration data, wherein a phoneme is perceived when replaced by noise, but not by silence. The plot shown in (b) is identical to the simulation from Fig. 8, except the correctly time-aligned phoneme activations for /k/ and /S/ are shown as well (/r/ and /^/are left out for simplicity). This figures shows that the Traces for /k/ and /S/ become active well before /l/ becomes positively excited, suggesting that the percepts described by the TRACE model do not mirror the fluent and sequential percepts formed when listening to a speech stream.

l^kS^ri-, where "-" represents a silence. The figure shows the activity of the /l/ phoneme whose center is at time slice 4, under three conditions. In the normal condition, all inputs are presented to the network, preceded and followed by activation of the silence node. In the silence condition, the /l/ is removed from the input and replaced by activation of the silence node. In the noise condition, a constructed noise vector replaces features for /l/ in the input. While it may at first seem that this simulation can account for phonemic restoration, since the /l/ phoneme becomes active in the noise condition but not the silence condition, serious problems appear upon closer inspection.

The first problem is that this simulation relies on the explicit activation of a silence node. Because silence is treated like any other phoneme and because it is arbitrarily given a feature representation that is orthogonal to every other phoneme, its representations in all the time slices become strongly activated in response to the absence of bottom-up input and strongly inhibit all other phonemes. If, instead, the TRACE model were to properly treat silence as the absence of acoustic input, the /l/ phoneme instead becomes more activated, and at an earlier time frame, when replaced by silence rather than noise. This is shown in Fig. 9(a). While parameter changes may allow for the /l/ phoneme to become more active in noise than in silence, in the absence of an artificial means of representing silence, TRACE perceptually restores a phoneme even when that phoneme is replaced by silence; i.e., the absence of input. This problem can be traced to how TRACE defines top-down inputs as driving, rather than modulatory.

The earlier activation of the /l/ phoneme during a silent interval than during a noise presentation arises from another problem of the TRACE model regarding the representation of time. Specifically, because time is represented as a series of frames during which reduplicated nodes process input, not only does the time course of activations lose

meaning from both a behavioral and neurobiological perspective, but so too do the existence of the reduplicated phoneme/word nodes themselves. More specifically, because it is argued that the TRACE, or node activity, is the percept, and interactive activation is the process of perception (Elman and McClelland, 1986), the TRACEs for all the reduplicated phonemes and words must explicitly be ignored, discarded, or shifted to the appropriate time alignment, in order to avoid the implication that they are all being perceived throughout the full duration of any stimulus presentation. These problems are made clearer if we consider the time course of activations in the case of noise replacing a phoneme. Figure 9(b) shows a simulation in which, even allowing for the use of a silence phoneme node, the /l/ phoneme node (centered at time 4), becomes active *only after* the phoneme nodes for /k/ and /S/ (we ignore here /r/ and /^/ for simplicity, since /r/ shares overlapping input features with /l/, and /^/ is a duplicated phoneme).

There are a couple of possible explanations for this property, yet none of them corresponds to properties of phonemic restoration. If we accept that the Trace is the percept itself, then we would expect the percept (Trace) of /l/ to become active before the percepts (Traces) of the phonemes subsequent to it, such as /k/ and /S/. Alternatively, consider the possibility that the Trace corresponds to a response probability as calculated by the Luce Choice Rule, which is used in McClelland and Elman (1986). Although these response probability curves are not shown here, they are roughly equal to the activation traces shown. Then one could make the argument that it is only after the lexical entry for *luxury* is recognized as such that the /l/ can be perceived as an /l/ rather than as noise. That is to say, only after enough evidence has accumulated for the lexical item, will a listener report perceiving /l/ rather than noise. The trouble with this argument is that, as evidence for the lexical item accumulates as a winning lexical entry more strongly inhibits its competitors, the response probabilities for perceiving the other phonemes in their respective positions would increase as well. This does not happen, however: Both their activations and their response probabilities are decreasing by the time /l/ begins to get activated. These are fundamental problems of the TRACE model which result from the fact that time is represented in an *ad hoc* manner, that bottom-up and top-down interactions are not plausible given the structure of the model itself, and that silence is represented as an explicit phoneme category.

B. MERGE

Norris *et al.* (2000) developed the MERGE model to argue that feedback from lexical to pre-lexical levels is not necessary in explaining speech perception. The model is a competition activation network, with excitatory connections between layers, and inhibitory connections within each layer. It consists of an input layer, which sends excitatory inputs to a word layer as well as a phoneme decision layer, and there are additionally feedback connections from the word layer to the phoneme decision layer. The MERGE model builds on the SHORTLIST model (Norris, 1994), which attempted to address some of the shortcomings of the TRACE model. However, MERGE also relies on activation functions that are not biologically plausible, and does not include learning laws. In fact, the MERGE model proposes that connections from lexical and pre-lexical levels to a decision layer should be built "on the fly" in a task-dependent manner, a proposal greatly at odds with how the brain works. Furthermore, as with the TRACE model, it simulates only a decision process by which a perceived word may be chosen, but does not describe what is actually perceived. As such, the MERGE model provides no explanation for why broadband noise is required in the perceptual restoration of a missing phoneme. Nor can it explain the grouping processes which give rise to perceived silence in the case where a phoneme is replaced by a silent interval. Finally, it is well known that top-down feedback processes are ubiquitous in the brain and are even more numerous than the bottom-up processes that they modulate (Felleman and Van Essen, 1991; Goldman-Rakic, 1987; Rempel-Clower and Barbas 2000). cARTWORD and other ART models clarify how these top-down processes control attentional and learning processes that are necessary for fast and stable language learning and context-sensitive conscious perception.

VII. CONCLUSION

The cARTWORD model describes in quantitative terms how a hierarchy of interacting laminar cortical circuits may give rise to conscious speech percepts and how bottom-up and top-down interactions serve to filter, store, chunk, modulate, and complete acoustic inputs into a coherent speech code. To do this, cARTWORD simulates how a temporal sequence of feature patterns is unitized into item representations. The item representations send matching feedback to the feature patterns as they are sequentially stored in working memory. In this way, a temporal series of speech sounds is stored as an evolving spatial pattern of activity through time. As this spatial pattern changes, it selects unitized sequence, or list, chunks that compete among one another to select the list chunk, or chunks, that best represents the currently active stored item sequence. These list chunks, in turn, can send top-down matching signals to the stored working memory items while they activate a gating network. As the gates open, the entire hierarchy of networks can enter a synchronous resonance. Activity-dependent habituative transmitter gates, or synaptic strengths, enable individual feature-item resonances to become activated in their correct order without perseveration. All of these mechanisms, notably the top-down attentive matching mechanisms, are part of an emerging cognitive theory that helps to explain how speech and language may be rapidly and stably learned. Using these resonant mechanisms, the cARTWORD model simulates key properties of data in which context acting over hundreds of milliseconds can influence what speech percept is heard. The phonemic restoration effect illustrates such contextual dynamics, including why a silence duration replacing an excised phoneme gives rise to a break in perceived speech, why broadband noise enables restoration of an excised phoneme, and how stimuli subsequent to the excised phoneme

can determine which phonemes are earlier perceived. These demonstrations contrast with competing models of speech perception, which have not shown how representations of such conscious speech percepts may arise.

As these concepts become increasingly well developed and used to explain ever more complex speech and language data, they may have an increasing influence on the design of speech recognition systems in technology, especially in multi-speaker noisy environments, where the coherent completion and noise suppression properties of resonant dynamics are most valuable.

ACKNOWLEDGMENTS

Supported in part by CELEST, an NSF Science of Learning Center (SBE-0354378) and by the DARPA SyNAPSE program (HR0011-09-C-0001).

Ames, H., and Grossberg, S. (**2009**). "Speaker normalization using cortical strip maps: A neural model for steady state vowel categorization," J. Acoust. Soc. Am. **124**, 3918–3936.

Anderson, J., and Bower, G. (**1974**). "A prepositional theory of recognition memory," Mem. Cognit. **2**, 406–412.

Aslin, R., Woodward, J., LaMendola, P., and Bever, T. (**1996**). "Models of word segmentation in fluent maternal speech to infants," in *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*, edited by J. Morgan and K. Demuth, (Lawrence Erlbaum Associates, Mahwah), pp. 117–134.

Atkinson, R., and Shiffrin, R. (**1968**). "Human memory: A proposed system and its control processes," in *The psychology of learning and motivation*, edited by K. Spence and J. Spence, (Academic Press, New York), pp. 89–195.

Averbeck, B., Chafee, M., Crowe, D., and Georgopoulos, A. (**2003a**). "Neural activity in prefrontal cortex during copying geometrical shapes. i. Single cells encode shape, sequence and metric parameters," Exp. Brain Res. **150**, 127–141.

Averbeck, B., Chafee, M., Crowe, D,. and Georgopoulos, A. (**2003b**). "Neural activity in prefrontal cortex during copying geometrical shapes. ii. Decoding shape segments from neural ensembles," Exp. Brain Res. **150**, 142–153.

Bellmann, A., Meuli, R., and Clarke, S. (**2001**). "Two types of auditory neglect," Brain **124**(4), 676–687.

Bhatt, R., Carpenter, G., and Grossberg, S. (**2007**). "Texture segregation by visual cortex: Perceptual grouping, attention, and learning," Vis. Res. **47**, 3173–3211.

Boardman, I., Grossberg, S., Myers, C., and Cohen, M. (**1999**). "Neural dynamics of perceptual order and context effects for variable-rate speech syllables," Percep. Psychophys. **6**, 1477–1500.

Bohland, J. W., Bullock, D., and Guenther, F. H. (**2010**). "Neural representations and mechanisms for the performance of simple speech sequences," J. Cog. Neurosci. **22**(7), 1504–1529.

Borg-Graham, L. J., Monier, C., and Fregnac, Y. (**1998**). "Visual input evokes transient and strong shunting inhibition in visual cortical neurons," Nature **393**(6683), 369–373.

Bradski, G., Carpenter, G., and Grossberg, S. (**1994**). "Store working memory networks for storage and recall of arbitrary temporal sequences," Bio. Cybern. **71**, 469–480.

Brown, J., Bullock, D., and Grossberg, S. (**1999**). "How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues," J. Neurosci. **19**, 10,502–10,511.

Brown, J. W., Bullock, D., and Grossberg, S. (**2004**). "How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades," Neural Networks **17**, 471–510.

Brown, L., Schneider, S., and Lidsky, T. (**1997**). "Sensory and cognitive functions of the basal ganglia," Curr. Op. Neurobio. **7**(2), 157–163.

Bullock, D., and Rhodes, B. J. (**2003**). "Competitive queuing for planning and serial performance," in *The Handbook of Brain Theory and Neural Networks*, edited by M. A. Arbib (MIT Press, Cambridge, MA), pp. 241–244.

Cao, Y., and Grossberg, S. (**2005**). "A laminar cortical model of stereopsis and 3D surface perception: Closure and da Vinci stereopsis," Spat. Vis. **18**, 515–578.

Carpenter, G. A., and Grossberg, S. (**1987**). "A massively parallel architecture for a self-organizing neural pattern recognition machine," Comp. Vis., Graph., Image Proc. **37**, 54–115.

Cohen, M., and Grossberg, S. (**1986**). "Neural dynamics of speech and language coding: Developmental programs, perceptual grouping, and competition for short-term memory," Hum. Neurobio. **5**, 1–22.

Cohen, M., and Grossberg, S. (**1987**). "Masking fields: A massively parallel neural architecture for learning, recognizing, and predicting multiple groupings of patterned data," App. Opt. **26**, 1866–1891.

Cohen, M. A., and Grossberg, S. (**1997**). "Parallel auditory filtering by sustained and transient channels separates coarticulated vowels and consonants," IEEE Trans. Speech Aud. Proc. **5**, 301–318.

Connine, C. M., and Clifton, C. Jr. (**1987**). "Interactive use of lexical information in speech perception," J. of Exp. Psych.: Human Percept. and Perf. **13**, 291–299.

Cowan, N. (**2001**). "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," Behav. Brain Sci. **24**, 87–185.

Damasio, A. R., Damasio, H., and Chui, H. C. (**1980**). "Neglect following damage to frontal lobe or basal ganglia," Neuropsychologia **18**(2), 123–132.

Fang, L., and Grossberg, S. (**2009**). "From stereogram to surface: How the brain sees the world in depth," Spat. Vis. **22**, 45–82.

Felleman, D. J., and Van Essen, D. C. (**1991**). "Distributed hierarchical processing in primate cerebral cortex," Cereb. Cortex **1**, 1–47.

Farrell, S., and Lewandowsky, S. (**2004**). "Modeling transposition latencies: Constraints for theories of serial order memory," J. Mem. Lang. **51**, 115–135.

Francis, G., Grossberg, S., and Mingolla, E. (**1994**). "Cortical dynamics of feature binding and reset: Control of visual persistence," Vis. Resear. **34**, 1089–1104.

Ganong, W. F. (**1980**). "Phonetic categorization in auditory word perception," J. Exp. Psy: Human Percept. and Perf. **6**, 110–125.

Gaudiano, P., and Grossberg, S. (**1991**). "Vector associative maps: Unsupervised real-time error-based learning and control of movement trajectories," Neural Networks **4**, 493–504.

Goldman-Rakic, P. S. (**1987**). "Circuitry of primate prefrontal cortex and regulation of behavior by representational memory," in *Handbook of Physiology*, edited by F. Plum, and V. Mountcastle (American Physiological Society, Bethesda), Vol. **5** pp. 373–417.

Gow, D. W., Segawa, J. A., Ahlfors, S. P., and Lin, F.-H. (**2008**). Lexical influences on speech perception: A Granger causality analysis of MEG and EEG source estimates, NeuroImage **43**(3), 614–623.

Grossberg, S. (**1968**). "Some physiological and biochemical consequences of psychological postulates," Proc. Nat. Acad. Sci. **60**, 758–765.

Grossberg, S. (**1969**). "On the production and release of chemical transmitters and related topics in cellular control," J. Theor. Bio. **22**, 325–364.

Grossberg, S. (**1973**). "Contour enhancement, short–term memory, and constancies in reverberating neural networks," Stud. App. Math. **52**, 213–257.

Grossberg, S. (**1978a**). "Behavioral contrast in short-term memory: Serial binary memory models or parallel continuous memory models?," J. Math. Psych. **3**, 199–219.

Grossberg, S. (**1978b**). "A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans.," in *Progress in Theoretical Biology*, edited by R. Rosen and F. Snell, (Academic Press, NY), Vol.**5**, pp. 233–374.

Grossberg, S. (**1980**). "How does a brain build a cognitive code?," Pysch. Rev. **87**, 1–51.

Grossberg, S. (**1984**). "Unitization, automaticity, temporal order, and word recognition," Cogn. 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*, edited by E. C. Schwab and H. C. Nusbaum (Academic Press, New York), pp. 187–294.

Grossberg, S. (**1987**). "Competitive learning: From interactive activation to adaptive resonance," Cogn. Sci. **11**, 23–63.

Grossberg, S. (**1988**). "Nonlinear neural networks: Principles, mechanisms, and architectures," Neural Networks **1**, 17–61.

Grossberg, S. (**2000**). "How hallucinations may arise from brain mechanisms of learning, attention, and volition," J. Intl. Neuropsych. Soc. **6**, 579–588.

Grossberg, S. (**2003**). "Resonant neural dynamics of speech perception," J. Phonetics **31**, 423–445.

Grossberg, S., Boardman, I., and Cohen, M. (**1999**). "Neural dynamics of variable-rate speech categorization," J. Exp. Psych: Hum. Percept. Perf. **23**, 418–503.
Grossberg, S., Govindarajan, K., Wyse, L., and Cohen, M. (**2004**). "ARTSTREAM: A neural network model of auditory scene analysis and source segregation," Neural Networks **17**, 511–536.
Grossberg, S., and Myers, C. (**2000**). "The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects," Psych. Rev. **107**, 735–767.
Grossberg, S., and Pearson, L. (**2008**). "Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: Toward a unified theory of how the cerebral cortex works," Psych. Rev. **115**, 677–732.
Grossberg, S., and Seitz, A. (**2003**). "Laminar development of receptive fields, maps, and columns in visual cortex: The coordinating role of the subplate," Cerebral Cortex **13**, 852–863.
Grossberg, S., and Stone, G. O. (**1986**). "Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance," Psych. Rev. **93**, 46–74.
Grossberg, S., and Versace, M. (**2008**). "Spikes, synchrony, and attentive learning by laminar thalamocortical circuits," Brain Res. **1218**, 278–312.
Grossberg, G., and Williamson, J. R. (**2001**). "A neural model of how horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual groupings and learning," Cerebral Cortex **11**, 37–58.
Grossberg, S., and Yazdanbakhsh, A. (**2005**). "Laminar cortical dynamics of 3D surface perception: Stratification, transparency, and neon color spreading," Vis. Res. **45**, 1725–1743.
He, J., Hashikawa, T., Ojima, H., and Kinouchi, Y. (**1997**). "Temporal integration and duration tuning in the dorsal zone of cat auditory cortex," J. Neurosci. **17**, 2615–2625.
Hikosaka, O., and Wurtz R. H. (**1989**). "The basal ganglia," in *The Neurobiology of Saccadic Eye Movements*, edited by R. H. Wurtz, M. E. Goldberg (Elsevier, Amsterdam), pp. 257–281.
Hodgkin, A., and Huxley, A. (**1952**). "A quantitative description of membrane current and its application to conduction and excitation in nerve," J. Physiol. **117**, 500–544.
Houghton, G. (**1990**). "The problem of serial order: A neural network model of sequence learning and recall," in *Current Research in Natural Language Generation*, edited by R. Dale, C. Mellish, and M. Zock. (Academic Press, London), pp. 287–319.
Lashley, K. S. (**1951**). "The problem of serial order in behavior," in *Cerebral Mechanisms in Behavior*, edited by L. A. Jeffries (Wiley, New York), pp. 506–528.
Kashino, M. (**2006**). "Phonemic restoration: The brain creates missing speech sounds," Acoust. Sci. Techn. **27**, 318–321.
Kazerounian, S., and Grossberg, S. (**2009a**). "Neural dynamics of speech perception: Phonemic restoration in noise using subsequent context," J. Acoust. Soc. Am. **125**, 2658(A).
Kazerounian, S., and Grossberg, S. (**2009b**). "Neural dynamics of phonemic restoration: How the brain uses context backwards in time," in *Proceedings of 13th International Conference on Cognitive and Neural Systems (ICCNS)*, Boston MA, May, p. 114.
Kazerounian, S., and Grossberg, S. (**2009c**). "Laminar cortical dynamics of conscious speech perception: Phonemic restoration in noise using subsequent context," Soc. Neurosci. Abstracts, Chicago, IL, pp. 1678–1679.
Masuda-Katsuse, I., and Kawahara, H. (**1999**). "Dynamic sound stream formation based on continuity of spectral change," Speech Comm. **27**, 235–259.
Magnuson, J. S., McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (**2003**). "Lexical effects on compensation for coarticulation: The ghost of Christmash past," Cognitive Science **27**, 285–298.
McClelland, J., and Elman, J. (**1986**). "The trace model of speech perception," Cog. Psych. **18**, 1–86.
Miller, G. (**1956**). "The magical number seven, plus or minus two: Some limits on our capacity for processing information," Psych. Rev. **63**, 81–97.
Miller, J. L., and Liberman, A. M. (**1979**). "Some effect of later-occuring information on the perception of stop consonant and semivowel," Percept. Psychophys. **25**, 457–465.
Norris, D. (**1994**). "Shortlist: A connectionist model of continuous speech recognition," Cognition **52**, 189–234.
Norris, D., McQueen, J., and Cutler, A. (**2000**). "Merging information in speech recognition: Feedback is never necessary," Behav. Brain Sci. **23**, 299–370.
Page, M. P., and Norris, D. (**1998a**). "The primacy model: A new model of immediate serial recall," Psychol. Rev. **105**(5): 761–781.
Page, M. P., and Norris, D. (**1998b**). "Modeling immediate serial recall with a localist implementation of the primacy model," in *Localist Connectionist Approaches to Human Cognition*, edited by J. Grainger, and A. M. Jacobs (Erlbaum, Mahwah, NJ), pp. 227–255.
Pasupathy, A., Miller, E. K. (**2005**). "Different timecourses of learning-related activity in the prefrontal cortex and striatum," Nature. **433**, 873–876.
Pitt, M.A. and Samuel, A.G. (**1995**). "Lexical and sublexical feedback in auditory word recognition," Cognitive Psychology **29**, 149–188.
Raizada, R., and Grossberg, S. (**2003**). "Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system," Cereb. Cort. **13**, 100–113.
Repp, B., Liberman, A., Eccardt, T., and Pesetsky, D. (**1978**). "Perceptual integration of acoustic cues for stop, fricative, and affricate manner," J. Exp. Psychol.: Human Percept. Perf. **4**, 621–637.
Rempel-Clower, N. L., and Barbas, H. (**2000**). "The laminar pattern of connections between prefrontal and anterior temporal cortices in the rhesus monkey is related to cortical structure and function," Cerebral Cortex, **10**, 851–865.
Rumelhart, D., Hinton, G., and McClelland, J. (**1986**). "A general framework for parallel distributed processing," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, (MIT Press, Cambridge), Vol. **1**: Foundations, pp. 45–76.
Samuel, A. (**1981a**). "Phonemic restoration: Insights from a new methodoloty," J. Exp. Psychol.: Human Percept. Perf. **4**, 474–494.
Samuel, A. (**1981b**). "The role of bottom-up confirmation in the phonemic restoration illusion," J. Exp. Psychol.: Human Percept. Perf. **7**, 1124–1131.
Samuel, A. (**1997**). "Lexical activation produces potent phonemic percepts," Cog. Psych. **32**, 97–127.
Shinn-Cunningham, B., and Wang, D. (**2008**). "Influences of auditory object formation on phonemic restoration," J. Acoust. Soc. Am. **121**, 295–301.
Srinivasan, S., and Wang, D. (**2005**). "A schema-based model for phonemic restoration," Speech Comm. **45**, 63–87.
Warren, R. (**1970**). "Perceptual restoration of missing speech sounds," Science **167**, 392–393.
Warren, R., and Obusek, C. (**1971**). "Speech perception and phonenemic restorations," Percept. Psychophy. **9**, 358–362.
Warren, R., and Sherman, A. (**1974**). "Phonemic restorations based on subsequent context," Percept. Psychophys. **16**, 150–156.
Warren, R., and Warren, R. (**1970**). "Auditory illusions and confusions," Sci. Am. **223**, 30–36.
Werbos, P. (**1974**). "Beyond regression: New tools for prediction and analysis in the behavioral sciences," Ph.D thesis, Harvard University, Cambridge, MA.