NEUROSCIENCE

INTERNATIONAL BRAIN IBRO RESEARCH ORGANIZATION

REVIEW

K. D. Himberger et al. / Neuroscience xxx (2018) xxx–xxx

Principles of Temporal Processing Across the Cortical Hierarchy

Kevin D. Himberger, Hsiang-Yun Chien and Christopher J. Honey *

Department of Psychological & Brain Sciences, Johns Hopkins University, Baltimore, MD, United States

Abstract—The world is richly structured on multiple spatiotemporal scales. In order to represent spatial structure, machine-learning models repeat a set of basic operations at each layer of a hierarchical architecture. These iterated spatial operations – including pooling, normalization and pattern completion – enable these systems to recognize and predict spatial structure, while robust to changes in the spatial scale, contrast and noisiness of the input signal. Because our brains also process temporal information that is rich and occurs across multiple time scales, might the brain employ an analogous set of operations for temporal information processing? Here we define a candidate set of temporal operations, and we review evidence that they are implemented in the mammalian cerebral cortex in a hierarchical manner. We conclude that multiple consecutive stages of cortical processing can be understood to perform temporal pooling, temporal normalization and temporal pattern completion.

Key words: hierarchy, predictive coding, timescales, sequence processing, temporal integration.

OVERVIEW

The mammalian cerebral cortex is organized as a functional hierarchy. The lowest levels of this hierarchy are located in primary sensory and motor cortices and higher levels are reached at increasing synaptic distance from the periphery (Fuster, 1997; Mesulam, 1998; Huntenburg et al., 2017). When measuring neural responses to spatially structured stimuli, recordings further up the cortical hierarchy reveal larger spatial receptive fields and increasing selectivity for feature configurations (Hubel and Wiesel, 1962; Grill-Spector and Malach, 2004; also Rauschecker and Tian, 2000). In response to time-varying stimuli, recordings further up the hierarchy reveal longer temporal receptive windows and increasing selectivity for coherent temporal structures (Xu et al., 2005; Hasson et al., 2008; Lerner et al., 2011; Honey et al., 2012; de Heer et al., 2017). Overall, circuits at low levels are affected by local spatiotemporal properties such as the pitch of sounds and the orientation of image patches, while higher order circuits are most affected by more complex and spatiotemporally distributed properties, such as those that determine semantics, animacy, and object-scene relationships.

The hierarchical organization of the cerebral cortex has prompted the hypothesis that a common set of input–output operations is applied at each stage of processing (Creutzfeldt, 1977; Bastos et al., 2012). This hypothesis motivates many computational models of hierarchical image processing, inspired by the structure of the primate ventral visual stream (Fukushima and Miyake, 1982; Serre et al., 2007; Yamins and DiCarlo, 2016). Three canonical operations used in hierarchical image recognition models include: spatial pooling, spatial normalization, and spatial pattern completion. Pooling between stages is thought to make higher order representations robust to scaling and translation of the input (Fukushima and Miyake, 1982; Serre et al., 2007). Normalization at each stage is thought to enhance feature selectivity and robustness to input gain (i.e. amplitude) (LeCun et al., 2010; Carandini and Heeger, 2012). Pattern completion is thought to reduce the effects of noise and missing information, and to produce low-dimensional representations that can accurately re-generate high-dimensional input (Hopfield, 1982; Dayan et al., 1995; Yuille and Kersten, 2006).

If these spatial operations are effective for analyzing and representing static images in a multi-stage processing stream, might similar operations work well for processing temporal structure? This manuscript considers the evidence for an analogous set of basic temporal operations: temporal pooling, temporal normalization, and temporal pattern completion. In Secti on "Motivation and scope" we motivate this hierarchical perspective and set the scope of our proposal. In Sectio n "Principles of temporal information processing" we define each of our proposed temporal operations and we consider the evidence that these operations are implemented in the brain. Finally, in Section "Challenges,

---

*Corresponding author.
E-mail address: chris.honey@jhu.edu (C. J. Honey).

Abbreviations: ECoG, electrocorticographic; EEG, electroencephalographic; ERPs, event-related potentials; fMRI, functional magnetic resonance imaging; MMN, mismatch negativity.

https://doi.org/10.1016/j.neuroscience.2018.04.030

Please cite this article in press as: Himberger KD et al. Principles of Temporal Processing Across the Cortical Hierarchy. Neuroscience (2018), https://doi.org/10.1016/j.neuroscience.2018.04.030

open questions, and predictions", we discuss the functional implications of this temporal processing perspective, as well as limitations, open questions and predictions.

MOTIVATION AND SCOPE

Many theoretical frameworks assume that a common set of input–output mappings is applied across multiple stages of cortical processing (Hawkins and Blakeslee, 2007; Kiebel et al., 2008; Bastos et al., 2012; Hasson et al., 2015; Heeger, 2017). One motivation for this approach is from anatomy: if similar circuitry is observed across diverse cortical areas (Douglas and Martin, 2018), then perhaps similar functional operations are being implemented as well (Creutzfeldt, 1977). A second and more recent motivation is the success of neural networks for image recognition and analysis. These models from machine learning have bolstered the argument that hierarchically repeated computations are an effective way to recognize the properties of spatial images (e.g. Jarrett et al., 2009) and to form multi-level spatial representations (Hinton et al., 2006). More recently, time-varying linguistic stimuli have also been successfully analyzed and represented using multi-scale architectures (Chung et al., 2016; Bradbury et al., 2016; Gehring et al., 2017).

Hierarchical processing architectures may be especially effective for processing stimuli that possess a hierarchical structure. In the spatial domain, this can be seen in the way that edge-like features compose contours, which compose objects (Hubel and Wiesel, 1959, 1962; Kourtzi and Connor, 2011), with roughly corresponding representations in consecutive levels of the models (Zeiler and Fergus, 2014). Since information is also organized hierarchically in time (e.g. phonemes inside of words inside of prosodic curves; motions inside of actions inside of events; Gibson, 1982; Poeppel, 2003) then perhaps a hierarchical architecture is also effective in the time domain, and may be used in the brain.

In addition to the theoretical motivations, there are strong empirical reasons to seek a set of canonical temporal operations. Empirically, we observe that temporal properties affect information processing ubiquitously across the cortical hierarchy. In the auditory pathway, neuronal responses depend on stimulus history up to many seconds earlier, and this effect is already present in primary auditory cortex (Ulanovsky et al., 2004; Rosburg et al., 2008) and its inputs (Parras et al., 2017). In the visual cortex, even the earliest stages of processing depend on prior stimulus properties from hundreds of milliseconds (Gavornik and Bear, 2014) or seconds earlier (Homann et al., 2017). More generally, history-dependent responses – manifesting as sequential "mismatch effects" – have been demonstrated in diverse cortical systems in electroencephalographic (EEG) and electrocorticographic (ECoG) recordings in visual and auditory paradigms (Pazo-Alvarez et al., 2003; Näätänen et al., 2007; Bekinschtein et al., 2009; Kremláček et al., 2016).

There is further empirical evidence that temporal processing has a hierarchical character. Early sensory regions tend to be affected by more recent properties of the input stream, while higher order regions are affected by more complex features and by longer windows of prior stimulus context (Xu et al., 2005; Hasson et al., 2008; Bekinschtein et al. 2009; Wacongne et al., 2011; Lerner et al., 2011; Honey et al. 2012; Gauthier et al., 2012). Thus, the classic computational notion of sensory pathways as a hierarchy of growing receptive fields might be more accurately conceived as a hierarchy of spatiotemporal fields. A critical remaining question is then, what kinds of temporal operations are performed at each level of the hierarchy?

We define "temporal operations" broadly as any input–output relationship where temporal properties of the input can affect the output. Thus, we do not restrict our focus to aspects of behavior and perception related to "timing" per se (e.g. duration estimation (Grondin, 2010) or temporal sensitivity (Watson, 1986)) or to stimulus parameters that constrain motion processing (Exner, 1875; Wertheimer, 1912; Sekuler, 1996). Our goal is not to describe minimal delays or fusion thresholds, but instead to specify which classes of input–output operations might be ubiquitously and generically applied in the cerebral cortex. Because almost all information in the world has a temporal structure, we focus on temporal operations that are observable across many different sensory pathways, and across many stages of processing.

Because we seek operations that may be applied to temporal structure in many areas, we contrast and combine data across modalities (vision and audition) and across aspects of cognition (language, perception, and semantics). For example, the concept of temporal pooling can be understood to apply both in visual psychophysics experiments (Barlow, 1958) and in studies of temporally reversed acoustics (Saberi and Perrott, 1999). Thus, although many parameters of auditory and visual processing may differ (Kubovy and Van Valkenburg, 2001), we believe that there are fundamental anatomical and functional similarities that motivate a search for common operations (Von Melchner et al., 2000). Therefore, we focus on experiments for which (i) basic temporal properties are manipulated, and (ii) neural responses to these manipulations have been measured for more than one area or type of input. Such studies have the best chance to reveal information processing motifs that are repeated across levels and neural pathways. Often, electrophysiological and behavioral data provide some of the best information about precise timing and about mechanism, while methods with a broad spatial view, such as functional magnetic resonance imaging (fMRI) enable us to explore variations in response across regions (e.g. Lerner et al., 2011; Gauthier et al., 2012; Zhou et al., 2017).

The empirical and theoretical considerations sketched above indicate that history-dependent processes unfold ubiquitously in the neocortex. In the text that follows, we marshal evidence for three candidate operations – temporal pooling, temporal normalization, and temporal pattern completion – that may be applied at each stage

Please cite this article in press as: Himberger KD et al. Principles of Temporal Processing Across the Cortical Hierarchy. Neuroscience (2018), https://doi.org/10.1016/j.neuroscience.2018.04.030

of cortical processing. For each operation, we consider three questions: (i) how should this temporal operation be functionally defined? (ii) What is the evidence that this operation is implemented in cortical circuits? And (iii) is this temporal operation applied iteratively across multiple cortical levels? After discussing these temporal operations, we conclude by considering the functional implications of our perspective, as well as some limitations, open questions and predictions.

PRINCIPLES OF TEMPORAL INFORMATION PROCESSING

Pooling in space and time

What is spatial pooling?. Because neurons sum the input to their dendrites (Araya et al., 2006), input pooling is a component of almost every neuronal circuit model. Pooling can be thought of as a ‘‘sum” (or ‘‘max”) operation applied to multiple inputs: the output neuron will produce a response if any sufficient subset of its inputs is sufficiently excited (Fig. 1A). In hierarchical models of visual object recognition, applying input pooling at multiple stages produces high-level representations that are robust to changes in location and scale of input (Boureau et al., 2010).

Pooling in time. Is there a temporal analog of spatial pooling? We suggest that a neuron is pooling input over a time window, τ, if that unit will be active at time t whenever it receives sufficient summed input over the time window from $t - \tau$ to t. Thus, when dendrites integrate synaptic inputs over time, this implements a form of temporal pooling. Temporal pooling makes neural response robust to changes in the precise timing of inputs. This allows a neuron to respond in similar ways at time t to different temporal arrangements of the inputs that arrived in the window leading from $t - \tau$ up to t (Fig. 1B). Moreover, increasing the window of temporal pooling allows neurons to respond to inputs from further in the past.

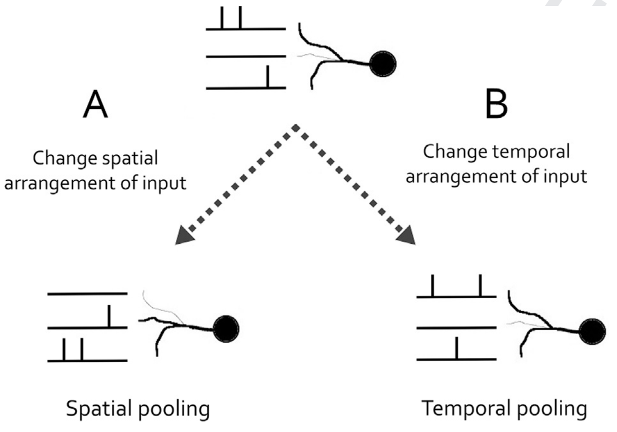

Fig. 1. Spatial and temporal pooling. (A) Spatial pooling: EPSPs on two dendrites provide sufficient excitation to generate an action potential at the soma. If the neuron is pooling information over space, then the same neuronal response can be generated, even when the spatial arrangement of the inputs is permuted. (B) Temporal Pooling: An action potential can be generated by a pair of EPSPs at the topmost dendrite combined with another EPSP at the bottom dendrite. If the neuron is pooling information over time, then the same response can be generated even if the inputs to the top dendrite are re-arranged in time. Lines represent dendritic arbors and circles represent the soma; blue highlight schematically indicates depolarization.

Temporal pooling is ubiquitous in the brain and operates over multiple timescales. Behavioral evidence indicates that different visual channels (color sensitive and colorblind) pool information over different timescales, perhaps as early as the retina (Stockman et al., 1991). More generally throughout the brain, dendritic integration of excitatory post-synaptic potentials (EPSPs), underlies temporal pooling on the order of tens of milliseconds. But pooling is not limited to short timescales: time-constants of receptor activation and inactivation, as well as other regulators of membrane excitability, produce temporal autocorrelation in single-neuron excitability that extends to seconds and hours (Gal et al., 2010; Gjorgjieva et al., 2016; Kukushkin and Carew, 2017). The excitability of neurons is also regulated by synaptic facilitation and depression (e.g. via residual presynaptic $Ca2+$; Mongillo et al., 2008) which can selectively bias subsequent neural responses. These changes in excitability can accumulate and linger over minutes of time, a process that may support working memory (Barak and Tsodyks, 2014; Stokes, 2015; Duarte et al., 2017).

Temporal pooling is apparent in the neural circuits supporting language processing. Fine-grained acoustics appears to be pooled temporally, because local speech segments (<100 ms) can be perturbed or even reversed with only minor effects on speech perception (Saberi and Perrott, 1999). More generally, speech perception is robust to changes in rate from about half speed to double speed, and neural responses within this range simply stretch or compress according to the input rate (Lerner et al., 2014). These speech phenomena are suggestive of processes that (i) are able to combine information over seconds of time and (ii) are robust to the precise arrival time of input on the scale of milliseconds and even seconds.

Recurrent circuit activity can lengthen the timescales of pooling. The presence of NMDA receptors, in addition to temporally smoothing the dynamics of individual neurons, can also stabilize reverberating excitation in neuronal populations, producing seconds of ramping activity in response to short transient input (Wang, 1999). In this way, circuits containing excitatory units equipped with NMDA receptors can act as a low-pass filter of their inputs. This low-pass filtering also implements some robustness to changes in timing and ordering: because the neurons that were active in the past continue to be active, this increases the likelihood that two neurons activated nearby in time will fire again coincidentally to excite a third neuron. Importantly, recurrent activity is not background noise, but propagates task-relevant activity from the past into the future: circuits with greater population level recurrence, usually found in higher order cortices (e.g. Chaudhuri et al., 2015), appear to integrate

Please cite this article in press as: Himberger KD et al. Principles of Temporal Processing Across the Cortical Hierarchy. Neuroscience (2018), https://doi.org/10.1016/j.neuroscience.2018.04.030

prior information to support behavior (Runyan et al., 2017).

Is temporal pooling applied consecutively in the cortical hierarchy?. There are at least two reasons to believe that temporal pooling is applied iteratively across multiple stages of cortical processing. Firstly, neurons exhibit many intrinsic processes for summing information over time, and these processes are not restricted to specialized cortical areas (Gjorgjieva et al., 2016; Kukushkin and Carew, 2017). Temporal pooling that arises from neuron-intrinsic factors likely applies iteratively at each stage along of hierarchical information processing. Secondly, the temporal autocorrelation (i.e. slowness) of neural activity increases as one moves from the sensory periphery toward higher order cortical areas. Hierarchical slowing of dynamics has been observed using fMRI (Stephens et al., 2013) and ECoG (Honey et al., 2012) as well as in single-unit recordings (Ogawa and Komatsu, 2010; Murray et al., 2014).

The gradual lengthening of temporal autocorrelation may arise because each stage of cortical processing introduces some constant amount of smoothing (via neuron-level or circuit-level time constants) and/or because the amount of local smoothing actually increases in higher order areas (e.g. via greater recurrent excitation in higher order areas) (Baria et al., 2013; Chaudhuri et al., 2015). Regardless of the mechanism, the functional outcome is that population neural dynamics become slower at consecutive stages of the cortical hierarchy, consistent with a temporal pooling operation being applied at each stage.

Normalization in space and time

If pooling was the only operation applied at each level of processing, each consecutive stage of processing would exhibit more activity and less spatial/featural selectivity. Therefore, computational models interleave the pooling operation with nonlinear operations that prevent overactivity and encourage more stimulus-selective responses (Serre et al., 2007; Jarrett et al., 2009).

What is spatial normalization?. Activity in a neural unit is normalized when its response is re-scaled according to the input received by other neurons with similar selectivity. A divisive form of normalization appears to be implemented across diverse brain regions and species (Carandini and Heeger, 2012). In the cortex, divisive normalization of a neuronal (or neural) response, R_j, takes the form:

$$R_j = \gamma\left(\frac{D_i^n}{\sigma^n + \sum_k D_k^n}\right) \tag{1}$$

where D_j is the input drive to the j-th neuron, R_j is the response of the j-th neuron, γ is a multiplicative scaling constant, σ is a shape constant, and the exponent n determines the sharpness of the nonlinearity.

Consider how the normalization in Eq. (1) operates in a set of neural units with varying levels of selectivity for a presented stimulus. Suppose that a "preferred" neural unit, p, receives the largest input drive, D_p. In that case, the numerator of Eq. (1) will be larger for p than for any other neural unit. If $n > 1$, then the relative difference between p and all other neural units will be magnified in favor of unit p. Therefore, even though unit p already had the advantage of the largest input drive, D_p, its output, R_p, after normalization will have an even greater advantage over the other units. Indeed, as n approaches infinity, the normalization operation instantiates a "winner-take-all" mechanism: all neural units will be normalized to zero, except the unit that receives the largest input drive.

To make this concrete, consider a square array of identical Gabor patches (Fig. 2A, top). The neuron that prefers the orientation and location of the central patch will have its response normalized, because of the similar surrounding patches. However, when the central patch is surrounded by patches of a differing orientation (Fig. 2A, bottom), the surrounding features would no longer add to the denominator of Eq. (1). In this case, the central patch experiences less divisive rescaling than the surrounding patches. This means that the neuron responding to the central patch will exhibit sharper selectivity for its preferred feature in this context. In addition, when the drive from each of the 9 locations is pooled at a subsequent stage of processing, the central item will have the largest signal, which can result in a visual "pop-out" effect (e.g. Itti et al., 1998).

In addition to enhancing the selectivity of responses, divisive normalization also provides robustness to changes in input gain (i.e. input amplitude). Consider what happens if the "gain" is turned up, so that the drive, D_i, to all neurons in the pool is doubled. As long as σ is not large, the numerator and denominator of Eq. (1) will be increased by similar amounts, and so the normalized response, R_i, will remain similar. Thus, normalization has been suggested to promote both (i) sharpening of response selectivity; (ii) spatial pop-out; and (iii) robustness to changes in gain. Finally, because the divisive normalization equation tends to produce a sub-linear increase in the output responses as a function of the input, it can prevent saturating the neural population and facilitates efficient neural coding.

Normalization in time. Normalization in space occurs when the response of a neuron to feedforward drive is affected by the summed drive to other similar neurons. How should we think of normalization in time? Following Zhou et al. (2017), we suggest that normalization in time occurs when the response of a neuron at time t is divisively reduced by the total drive that it received in a recent time window. Under this definition, normalization in time has a close relationship with neuronal adaptation: the tendency of a neuron to decrease its response to a stimulus, when that stimulus was previously presented within some time window (Fig. 2B, top). Just as spatial normalization enables a neuron to re-scale its response according to spatial context, adaptation re-scales the responsiveness of a neuron according to its temporal context. In this way, temporal normalization can improve neuronal sensitivity to sequential patterns, and make them robust to

Please cite this article in press as: Himberger KD et al. Principles of Temporal Processing Across the Cortical Hierarchy. Neuroscience (2018), https://doi.org/10.1016/j.neuroscience.2018.04.030

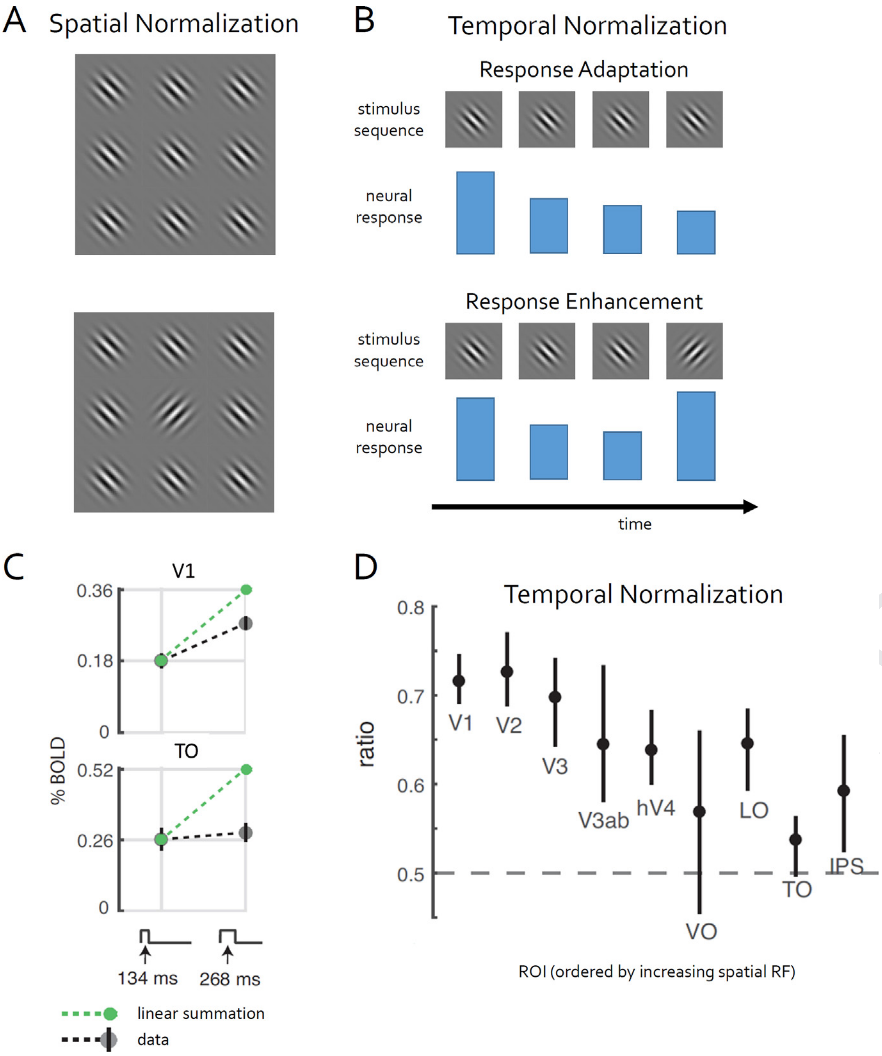

Fig. 2. Spatial and temporal normalization. (A) In the top array, the central Gabor patch is normalized in the same way as he other patches, as the features are shared. In the bottom array, the unique central patch "pops out"; it has a unique feature and there is not subject to the same divisive normalization as the surrounding patches. (B) Repeated presentation of the same stimulus (top) results in a reduced neural response, an example of adaptation. Repeated presentation of the same stimulus followed by a different stimulus (bottom) results in a response that is greater than the previous stimulus. (C) BOLD signal response compared across stimuli of 134-ms and 268-ms duration across area V1 and area TO of the visual pathway. If neural responses increase linearly in time with input duration (green dotted line), then the response amplitude should double when duration is doubled. The data indicate sublinear summation in V1 and even more sublinear summation in TO. (D) The temporal summation ratio measures the change in response when stimulus duration is doubled, relative to the change in response expected under a linear summation model. The temporal summation ratio is defined as the response to a stimulus duration at 2 *t* divided by twice the response to a stimulus duration of *t* (where t = 134 ms in this case). Smaller summation ratios indicate a small response increase with increasing duration; summation ratios below 1 indicate sublinear summation. Regions of interest in the ventral visual stream are organized left to right in approximately increasing order of spatial receptive field size. Later stages of visual processing exhibit response summation that is increasingly compressive (further below the predictions of a linear summation model). (C, D) are modified from Zhou et al. (2017).

changes in input gain (Abbott et al., 1997; Díaz-Quesada and Maravall, 2008).

Adaptation has been reported in diverse species, regions, and task settings (Krekelberg et al., 2006). The inferior colliculus and primary auditory cortex adapt to repeated tone stimuli (Ulanovsky et al., 2003; Todorovic et al., 2011; Malmierca et al., 2009, 2014), while higher order cortices such as the superior temporal gyrus and superior temporal sulcus adapt to repetition of more complex acoustic stimuli, such as speech or even multimodal stimuli (Altmann et al., 2008; van Atteveldt et al., 2010). Adaptation has also been observed throughout the visual system: the retina normalizes its response to local light using both spatial contrast and temporal contrast (Carandini and Heeger 2012), and multiple stages of visual cortex adapt their responses to repeated stimuli, as measured using fMRI (Kourtzi and Huberle, 2005; Henson, 2016), EEG (Stefanics et al., 2011), and single-unit recordings (Krekelberg et al., 2006).

Adaptation can also produce a sequential pop-out effect: infrequently presented stimuli are less adapted and therefore produce larger responses than frequent stimuli. In such a case, the relative enhancement of the infrequent stimulus is called a "fresh afferent" effect (May and Tiitinen, 2010) (Fig. 2B, bottom). The fresh afferent effect is thought to account for changes in the amplitude of early event-related potentials (ERPs) components such as the negative-going N1 (Rosburg et al., 2008).

There has been debate over whether later ERP components, such as the "mismatch negativity" (MMN) are also a result of adaptation and fresh afferent effects (Astikainen et al., 2006; May and Tiitinen, 2010; Kremlácek et al., 2016). The MMN component is elicited by a deviant stimulus amidst a stream of mostly repeated stimuli (Näätänen et al., 1978). It can be obtained using both auditory and visual stimulus streams (Pazo-Alvarez et al., 2003; Czigler, 2007). Because the magnitude of the MMN is greater when the deviant is less frequent and when inter-stimulus interval is shorter, it has been suggested that the MMN reflects an active comparison between the observed stimulus and a predicted stimulus (Näätänen et al., 2007). However, these phenomena may also be explained by increased adaptation to the repeated stimulus, exacerbated by increasing stimulus frequency or decreasing ISI (May and Tiitinen, 2010). Thus, just as spatial normalization contributes to "pop-out" of mismatching stimuli in spatial configurations (Itti et al.,

Please cite this article in press as: Himberger KD et al. Principles of Temporal Processing Across the Cortical Hierarchy. Neuroscience (2018), https://doi.org/10.1016/j.neuroscience.2018.04.030

1998), temporal normalization (adaptation) may also underlie many pop-out effects elicited for unusual elements within a sequence (May and Tiitinen, 2010).

Is normalization in time applied consecutively in the cortical hierarchy?. To determine if the neurons at a given level of cortical processing are adapting their responses to input, it is critical to know that the input is not itself becoming weaker. If neurons at stage N-1 are adapting their responses, then this could produce a weaker input to stage N; thus, the neurons at stage N could produce a weaker response to input, even if no adaptation is occurring at level N (Krekelberg et al., 2006). Therefore, in order to demonstrate that normalization in time (adaptation) is implemented in a hierarchical manner at multiple consecutive stages, it is most useful to focus on cases where a higher region adapts its response under conditions where the input to that region is not decreasing.

In the auditory domain, adaptation appears to operate over longer time windows in higher order regions. Early evidence for this claim was reported by Lü et al. (1992), who measured a slower recovery from adaptation for magnetoencephalography responses in higher stages of the auditory pathway. More recent work has confirmed this general pattern. Malinowska et al. (2017) found that early auditory regions adapted to pure tones, but not to complex speech stimuli, while association cortex adapted more strongly to both kinds of stimuli. The adaptation to speech in association cortex is unlikely to have been inherited from the earlier cortical stage, which exhibited little response reduction for this stimulus. Similarly, using electrophysiological recordings, Nieto-Diego and Malmierca (2016) reported greater adaptation in nonprimary auditory cortex than in primary auditory cortex. Together, these data suggest that adaptation in the auditory pathway occurs in many cortical & non-cortical regions and that its effects are larger, and operate over a longer window, in higher order regions.

Higher order visual cortices also appear to exhibit larger adaptation effects over longer timescales, compared with early visual cortices. Uusitalo et al. (1996) observed slower recovery from adaptation in higher order visual regions, compared to early visual regions. Using fMRI, Kourtzi and Huberle (2005) investigated how different levels of the visual hierarchy processed stimuli at different spatial scales (global versus local structure) and temporal scales (100-ms versus 400-ms ISIs). Early visual areas adapted only to local features in space, whereas higher order areas demonstrated adaptation for global structure. On the temporal side, early visual areas (V1 & V2) adapted to the stimuli when the ISI was short (100 ms), but not when it was long (400 ms). In contrast, higher order regions, like V4, adapted equivalently to both ISIs. Finally, when studying category selective areas of the ventral stream, Weiner et al. (2010) observed stronger fMRI adaptation effects in higher regions of the cortical hierarchy.

A quantitative test of hierarchical temporal normalization across multiple visual regions was recently conducted by Zhou et al. (2017). The authors measured ECoG and BOLD response to sequentially presented large-field contrast patterns, while varying the ISI and stimulus duration between 0 and 533 ms. In early visual cortical areas (V1, V2, V3), the least adaptation was observed and the recovery from adaptation was most rapid; in higher visual areas (IPS, LO and TO), adaptation was greater and occurred for longer ISIs. By fitting a temporal analog of Equation 1 to their data, Zhou et al. (2017) separately quantified the effective temporal window over which normalization performed, as well as the ‘‘sublinearity” of summation of consecutive stimuli (Fig. 2C). The sublinearity was quantified as the ratio between twice the response to an individual stimulus (i.e. perfect linear summation) and the actual response to consecutive stimuli. They found that higher regions had longer integration windows (i.e. a longer window across which normalization was computed) as well as greater sublinearity (i.e. the most temporal compression) (Fig. 2D). Thus, it appears that processes at multiple consecutive stages of cortical processing exhibit adaptation, and later stages can adapt across longer durations than earlier ones (Tervaniemi et al., 1994; Astikainen et al., 2006; Wacongne et al., 2011).

Pattern completion in space and time

What is pattern completion?. Pattern completion can be variably construed. Here, we define pattern completion as the process by which a noisy, incomplete, or unusual input pattern is transformed into a more noiseless, more complete and more stereotyped output pattern. The output pattern is ‘‘stereotyped” because it is constrained to follow a set of ‘‘pre-learned” or ‘‘pre-defined” regularities. For example, when you see a partially obscured face, you are able to generatively infer some of the features of that face because of your existing knowledge about the regularities of faces (Yuille and Kersten, 2006). For concreteness, we will focus on the type of pattern completion that is performed by auto-associative networks (Hopfield, 1982); this is only one kind of pattern completion or generative inference but it is illustrative of many relevant computational properties.

Auto-associative networks receive an input vector (i.e. pattern) and generate an output vector that is an approximate ‘‘reconstruction” of that input. Auto-associative networks do not simply copy the input to the output: instead, the output is reconstructed in a multi-stage process. Consider a simple 3-layer network in which information flows from the input layer to a hidden layer, and the input is subsequently reconstructed at the output layer. For our toy example, the inputs to (and outputs from) the network are values corresponding to a 6×5 spatial image (Fig. 3A), and the auto-associative process is mediated by a hidden layer with only six elements. In this case, the input pattern is converted into a compressed (lower dimensional) form via the connection weights from the input layer to the hidden layer. The units in the hidden layer can be thought of as representing regularities: combinations of stereotyped features that were learned from the input. The output of

Please cite this article in press as: Himberger KD et al. Principles of Temporal Processing Across the Cortical Hierarchy. Neuroscience (2018), https://doi.org/10.1016/j.neuroscience.2018.04.030

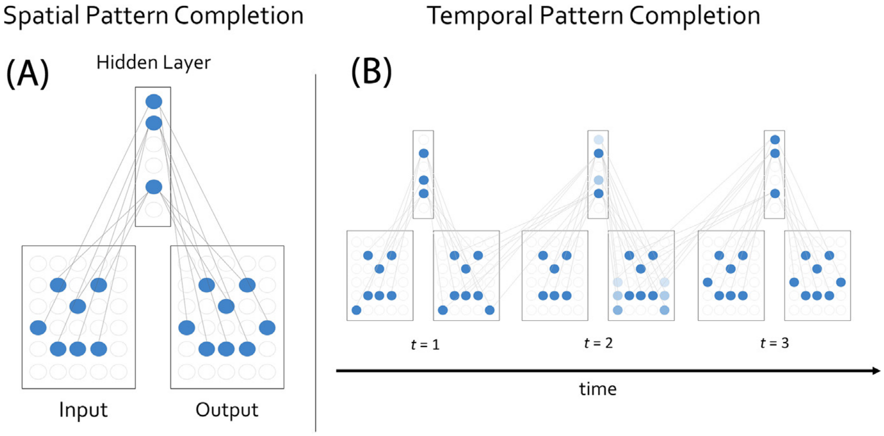

Fig. 3. Spatial and temporal pattern completion. (A) Spatial pattern completion: (Input) Pattern representing a partially smiling face image. (Hidden Layer) The input elicits activity in face-related units in the hidden layer. (Output) The reconstructed output image contains a more symmetric and stereotyped smiling face image. (B) Temporal pattern completion: At each time step, activity of the hidden layer is driven by the input image at the current step, as well as the output image from the previous step. At time step 1, a partial frowning face is reconstructed to a full frowning face. At time step 2, an expressionless face results in an ambiguous output representation for different facial expressions, and waits for the next input. At time step 3, a partial smile is presented, resulting in the output of a complete smiling face.

the network is then generated by "un-compressing" the stereotyped features from the hidden layer.

Suppose that a "face" pattern is stored in the network shown in Fig. 3A. If a partial face image is presented to the network, the partial image will excite face-related features in the hidden layer. These face-related features in the hidden layer will then be un-compressed as a "stereotyped" set of face elements, generated via connections from the hidden to the output layer. Thus, a partial face pattern (input) will be reconstructed as a more complete face pattern (output).

The hidden layer can accurately reconstruct input patterns that qualitatively match the previously learned patterns stored in its connections (e.g. the connections that generate a face). However, the hidden layer cannot reconstruct meaningful output from random noise or unfamiliar configurations. In this way, auto-encoders implement a kind of "fuzzy matching" which transforms incomplete or noisy inputs into more stereotyped representations that are more consistent with memories or priors. Because missing or distorted pieces of the input are corrected by this form of pattern completion, the notion of low-dimensional reconstruction has been proposed as a model of both memory (where a cue recovers related items, Hopfield, 1982) and perception (where an incomplete sensory input is converted to a complete representation, Buffart et al., 1981).

Spatial pattern completion can be perceptually experienced via the perceptual "fill-in" of visual textures. If a viewer closes one eye and fixates on the circled area in Fig. 4, then a homogeneous visual rectangular texture will be perceived after a few seconds. This perceptual phenomenon is paralleled by a "fill in" at the neuronal level: neurons in macaque V2 and V3 whose receptive fields covered the "empty" area, gradually increased firing in parallel with the perceptual change (de Weerd et al., 1995). In this case, the pattern that is completed is simply a continuation of the surrounding features.

Pattern completion in time. Pattern completion is a mapping from an input pattern [x_1, x_2, ... x_M] to a "reconstructed" output version [χ_1, χ_2, ... χ_M]. For spatial pattern completion, each of the M dimensions contains a feature located in space; for temporal pattern completion, each of the M dimensions contains a feature located in time. So, if the temporal input sequence is "ABC" and the network has stored an "ABCD" sequence, then information related to "D" may be activated in the network even before it appears. In this way, temporal pattern completion provides the functional benefit of prediction.

Consider a toy example in which a temporal auto-associator is presented with a series of images depicting facial dynamics (Fig. 3B). The memories stored in such an auto-associator may not only complete patterns in space, but also in time. Because of the connections between consecutive time-points in the model, the transition from a neutral expression to a halfway smile can be pattern-completed into the final stage of a complete smile.

Pattern completion may operate even in very early sensory cortical systems. Recording field potentials in primary visual cortex of mice, Gavornik and Bear (2014) tested whether sensory circuits would alter their responses to sequences of full-field gratings that were presented in regular patterns. They found that the responses to a repeated sequence of orientation stimuli

Fig. 4. A spatial "fill-in" effect. When viewing this figure, close one eye and fixate on the red circled location. The empty area of the image will eventually appear to be "filled in".

Please cite this article in press as: Himberger KD et al. Principles of Temporal Processing Across the Cortical Hierarchy. Neuroscience (2018), https://doi.org/10.1016/j.neuroscience.2018.04.030

(ABCD, ABCD, ABCD…) became greater over the course of hundreds of exposures, relative to mice that were exposed to a randomized sequence (DBAC, BCAD, ACDB…) (Fig. 5A). Critically, mice trained on the ABCD sequence exhibited a large response to an omission of element B in the sequence A_CD, but negligible response to a gap in the untrained sequence E_CD (Fig. 5B). A laminar analysis suggested that the response in the place of the missing "B" element was unlikely to arise from feedback from a higher visual area. This response suggests even early stages of sensory cortex can encode spatiotemporal structure (< 200 ms) and can "fill in" missing elements from a sequence, much like an auto-associator. This study does not definitively demonstrate temporal pattern completion, because the "fill in" response was not shown to encode information about the omitted "B" item. However, these results from early visual cortex are consistent with temporal pattern completion, and conclusively demonstrate sequence-specific context-sensitive processing at the earliest stages of cortical processing.

If pattern completion is occurring throughout the brain, then what happens when the observed pattern differs from what was expected? Findings from the ERP literature exhibit a typical pattern: continuations of expected patterns produce a reduced neural response, while violations produce an increased neural response. For example, a sequence of descending tones followed by one repeated tone will produce an MMN (Tervaniemi et al., 1994). This result cannot be explained by adaptation: if adaptation was operating then the repeated stimulus should produce a smaller response. Instead, it seems that the cortex "expects" the continuation of the descending pattern, and the violation of this expectation leads to the mismatch signal. As noted earlier, it can be difficult to determine whether a "mismatch effect", such as an MMN, reflects "fresh afferent" activity (due to adaptation; May and Tiitinen, 2010) or an active pattern completion process (Parras et al. 2017). But there are many cases – for example when an MMN is elicited by the omission of a stimulus (Nordby et al., 1994) – where the mismatch ERP seems unlikely to be explained by fresh afferents, and requires some endogenous process such as a generative pattern completion mechanism (Garrido et al., 2009; Bendixen et al., 2012; Kremlácek et al., 2016; Barascud et al., 2016; Parras et al., 2017).

Further evidence that some ERPs reflect an active pattern completion process comes from studies in which prior experience and knowledge alter the mismatch or surprise effect. For example, the MMN to language-deviant syllables is larger to native speakers of a language, compared to non-native speakers (Näätänen et al., 1997), and the MMN to silence is larger in musicians than non-musicians (Rüsseler et al., 2001). Similarly, a P600 response elicited by violation of a grammatical rule can be eliminated when a participant has not learned the rule (Friederici et al., 2002). The evidence is still indirect in these cases, but the involvement of experience-dependent "expectations" seems a more natural mechanism for these effects rather than a simple recovery from adaptation.

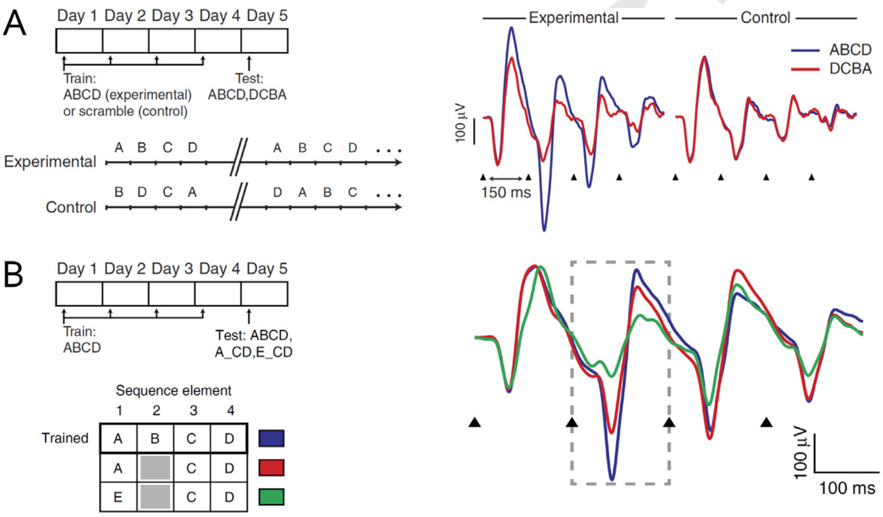

Fig. 5. Learned sequence-dependent responses in primary visual cortex. (A) Mice were trained over four days either with a repeated sequence (ABCD) of visual full-field gratings (experimental group) or with scrambled combinations of those visual elements (control group). In field potential recordings on the fifth day, increased response amplitudes were recorded in V1 in response to the trained sequence (ABCD), for the experimental group only and not the control group. This enhanced response is an expression of sequence learning, detectable in primary visual cortex. (B) Mice were exposed over four days to many repetitions of the visual sequence ABCD and were then tested on the fifth day with one of 3 sequences: ABCD, A_CD, or E_CD (where "_" indicates the omission of any stimulus). The response to the omitted stimulus following "A" was larger than the response to an omitted stimulus following "E". A sequence-dependent response is thus generated even in the absence of bottom-up drive, consistent with a predictive "fill in" process. Panels D and E modified from Gavornik and Bear (2014).

Sequence-dependent responses have also been reported using invasive electrophysiology and imaging. In these cases, pattern violations often produce an elevated neural response and are again interpreted as "prediction errors" or "surprise". For example, when Kaposvari et al. (2016) recorded from neurons in IT cortex of rhesus monkeys viewing sequences of images (e.g. A, B, C), they observed increased firing in response to images that violated the learned sequence (e.g. an elevated response to E in the unlearned sequence A, B, E). Similarly, Homann et al. (2017) used calcium imaging of mouse V1 layer 2/3 neurons to measure the responses to repeating full-field visual sequences of various lengths. As short sequences of images (e.g. ABCD) were repeated, most neurons gradually adapted their responses to the entire sequence to zero. Then, at the transition to a novel sequence (e.g. when switching from repetitions of ABCD to repetitions of EFGH), there was a large population-level "novelty" response. Critically, the decay and novelty processes could not be predicted by simple adaptation processes operating with a fixed time constant. Still, the results of Kaposvari et al. (2016) and Homann et al. (2017) might be explained by a more complex form of adaptation applied to neurons that respond to entire sequences of input, rather than to single inputs. Thus, to establish that prediction is operating in the sense described in Fig. 5 (with "fill in"), one would still need to test the neural response to an omitted sequence element.

Finally, temporal pattern completion has also been studied extensively in the domain of language. Although many of these studies do not directly demonstrate the completion of a missing element, they demonstrate that the processing of linguistic input at multiple levels is shaped by whether it is consistent with the preceding temporal pattern or structure. For example, deviant words in sentences are associated with an ERP known as the N400 (latency ~400 ± 200 ms) (Kutas and Federmeier, 2011), which is thought to continuously index how much the prior context eases the processing of new information (van Berkum 2009). A later P600 component (latency ~600 ± 250 ms; Sassenhagen et al., 2014) can be elicited by grammatical errors in a sentence. Moreover, it seems that linguistic pattern completion can operate at multiple levels in parallel. For example, the expectation of upcoming words can be determined by lexical frequency, but the context generated by a sentence can "supersede lexical frequency effects on the N400" (Kutas and Federmeier, 2011).

Is pattern completion in time applied consecutively in the cortical hierarchy?. Temporal pattern completions ("predictions") concerning low-level stimulus features appear to be generated in lower cortical regions, while predictions regarding higher level features are observed in higher cortical regions. Much of the evidence for this claim comes from mismatch paradigms which contrast violations of "local" and "global" patterns (Fig. 6). Electrophysiological and neuroimaging studies indicate that the local violations elicit mismatch signals in earlier regions, while violations of global patterns were more likely to elicit mismatch responses on associative and frontal cortices (Bekinschtein et al. 2009; Wacongne et al., 2011). A similar local–global violation distinction was observed in fMRI studies with rhesus macaques in fMRI (Uhrig et al., 2014). Similarly, all stimulus changes in a sequence induced ECoG high-frequency activity in sensory regions (a known proxy for neural activity), but only *globally* unpredictable stimuli elicited high-frequency activity in frontal regions (Dürschmid et al., 2016). The evidence here should not be overstated to indicate a strict hierarchy, as "higher" order regions such as the inferior frontal gyrus are also likely involved in low-level mismatch. Moreover, some of these local–global phenomena may be attributed to adaptation operating in both sensory and higher order cortices (e.g. Symonds et al. 2017). Thus, there remains uncertainty about the content of the context-sensitive processing that is occurring. Nonetheless, the current evidence is consistent with higher order cortices implementing predictive function over longer timescales.

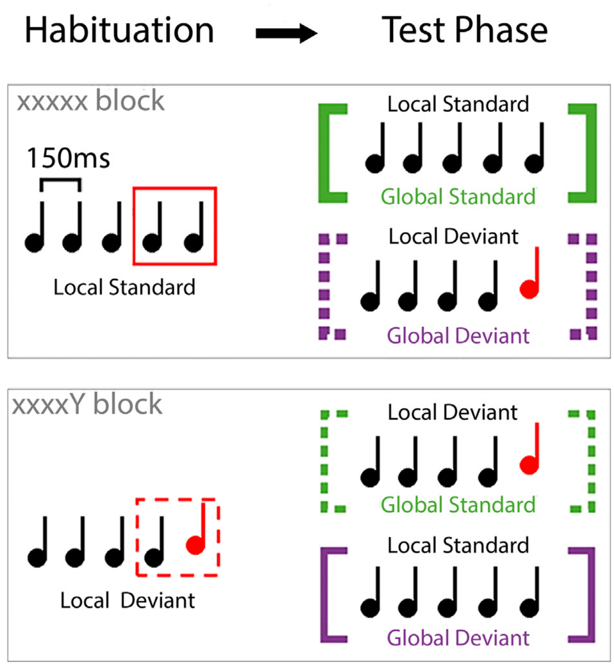

Fig. 6. Violations of local and global temporal structure. In local–global mismatch paradigms, participants are pre-exposed to one of two patterns: either repeating five element sequences of a single tones (xxxxx) or repeating five element sequences in which four tones are followed by a local deviant tone (xxxxY). In the subsequent test phase, both xxxxx and xxxxY sequences are presented, but now what is surprising (locally and globally) is different depending on what was presented in the pre-exposure phase. This paradigm enables the measurement of the difference between a local deviation (mismatch with local context) and a global deviation (mismatch with global context) while keeping the stimuli constant. Figure adapted from Wacongne et al. (2011).

The pattern of results from these local–global paradigms has also been interpreted as evidence for a hierarchy of ERP components. In this hierarchy of ERPs, the MMN occurs at the lowest level and earliest latency; the P3 component is intermediate; and the "Contingent Negative Variation" (CNV), generated in the anterior frontal cortex, occupies the highest level (Chennu et al., 2013). These local–global studies have also been interpreted in terms of formal predictive coding models (e.g. Mumford, 1992). Model comparison applied to fMRI data indicated that "prediction errors" were passed forward along the sensory processing hierarchy and expectations were transmitted back from higher order regions (Summerfield and Koechlin, 2008). Altogether, the evidence marshalled in support of predictive coding models can also be read as evidence for temporal pattern completion at many levels of the cortical hierarchy. In Section "Challenges and clarifications regarding hierarchical temporal operations" we consider the similarities and differences between predictive coding models and our own hierarchical perspective.

CHALLENGES, OPEN QUESTIONS, AND PREDICTIONS

The previous sections provided evidence for two claims: (i) the cerebral cortex applies pooling, normalization,

Please cite this article in press as: Himberger KD et al. Principles of Temporal Processing Across the Cortical Hierarchy. Neuroscience (2018), https://doi.org/10.1016/j.neuroscience.2018.04.030

and pattern completion to sequences of information from the world; and (ii) each of these basic operations is applied consecutively at multiple stages of the cortical hierarchy. This is far from a complete description of how time-varying information is processed, but this set of three principles can usefully organize our understanding. Hierarchical temporal pooling makes prior information available to later processing, and it contributes robustness to temporal jitter and rate. Hierarchical temporal normalization contributes gain control, selectivity for particular temporal sequences, and better use of the dynamic range available for neuronal coding (Abbott et al., 1997). Hierarchical temporal pattern completion provides robustness to noise and missing information; more generally, the combination of bottom-up and top-down information flow enhances adaptive behavior in a changing environment (Friston and Kiebel, 2009; Heeger, 2017).

Below we consider some challenges to this hierarchical approach to temporal information processing, and we sketch some open questions and predictions.

Challenges and clarifications regarding hierarchical temporal operations

(i) Are pooling, normalization and pattern completion actually implemented in the cerebral cortex? And does this actually occur locally within each cortical region? The hippocampus (Schapiro et al., 2016), basal ganglia (Graybiel, 1998) and inferior frontal cortex (Fadiga et al., 2009) exhibit some specializations for processing sequential structure. These "sequence-specialized circuits" also interact more directly with higher order cortical areas than with sensory cortical areas. So one could reasonably questions whether history-dependent responses in cortical regions are inherited from interactions with other circuits specialized for sequence processing. However, there are two reasons why sequential effects in neocortical processing cannot all be inherited. First, invasive recordings from early sensory regions reveal short-latency history-dependent effects (Ulanovsky et al., 2004; Gavornik and Bear, 2014; Homann et al., 2017; see also Leonard et al., 2015) some of which certainly depend on local circuitry (Natan et al., 2015). Non-invasive recordings also reveal history-dependent effects at short latencies (Boutros et al., 1995; Rosburg et al., 2008). Secondly, lesions to specialized area do not abolish many aspects of sequence processing: hippocampal amnesics and Parkinson's patients are only moderately impaired on complex language tasks and they continue to enjoy music (e.g. Mackay et al., 1998). Moreover, the cortical lesions that elicit classical aphasias do not appear to be tightly localized (Blank et al., 2016). Thus, specialized sequence-processing circuits likely coexist with a ubiquitous and local neocortical capability for sequence processing.

(ii) Why do we suggest that higher order regions pool and normalize their responses over longer timescales than early sensory regions? After all, even the earliest stages of cortical processing already exhibit adaptation on timescales, from hundreds of milliseconds to tens of seconds (Ulanovsky et al., 2004). In response to this point, we must clarify that our claims about the lengthening of timescales are at the population level. Every cortical region is proposed to contain a mixture of timescales (see also Bernacchia et al., 2011). Thus, the existence of individual neurons with long timescales in early areas is possible, but we propose that the overall mixture shifts toward longer timescales in higher order regions. As a result, a larger proportion of neurons in higher order regions will be instantiate pooling, normalization and pattern completion over longer time windows.

(iii) How does our framework differ from hierarchical predictive coding (HPC) models (Lee and Mumford, 2003; Friston and Kiebel, 2009)? We share a basic assumption of HPC models that anatomical and functional hierarchy in the cerebral cortex arises from repeated computational motifs. However, there are at least two basic differences between the present framework and HPC. First, HPC models propose that prediction is the fundamental principle underlying coding and learning of information at all levels. While we agree that prediction is a fundamental cortical computation, especially important for driving unsupervised learning (Honey et al., 2017; Singer et al., 2017), we think it is only part of the story. Other computational motifs such as pooling and normalization are also needed to enhance signal-to-noise and robustness of online representation and analysis. A second difference with HPC models is that we do not specify how prediction violations are signaled. HPC models differ widely (see reviews in Heilbron and Chait, 2017; Stephan et al., 2017) but they share the defining characteristic that error signals are fed forwards and predictions are fed back. Our framework is neutral on this point, and can therefore accommodate other configurations, such as the control-theoretic model of Heeger (2017), in which predictions are fed forward and errors are fed back.

Open questions

(1) Chunking is an important operation which we have not considered in space or in time. In the spatial domain, the conjunction of color, shape, boundary and other properties into objects is a fundamental component of visual cognition (Scholl, 2001). Similarly, in the temporal domain, one might propose a temporal grouping operation for "events" that are separated by "event boundaries" (Zacks et al., 2001). Future work might consider whether object construction and event assignment are also basic operations that could be implemented in a hierarchical fashion (Baldassano et al., 2017).

(2) If both spatial and temporal operations are applied across multiple stages of the cortical hierarchy, might the parameters of spatial and temporal operation co-vary across regions? For example, suppose that divisive normalization of spatial features is implemented in a winner-take-all fashion in Region X, with the exponent n in Equation 1 very large. In that case, perhaps temporal normalization in Region X would also be winner-take-all: this would predict large adaptation effects from even a single prior presentation of a stimulus. Zhou et al. (2017) have already demonstrated that spatial and temporal operations can be fit using similar quantitative models, and so this question appears tractable.

Predictions arising from the framework

We have sketched a hierarchical framework of temporal operations and their functional benefits. Clear predictions cannot be made until this framework is instantiated to a quantitative model. Still, our framework broadly predicts that when temporal normalization or pattern completion is disrupted, this should have functional consequences that are detectable in behavior. For example, disruption of normalization may disrupt the ability to recognize the same melody being presented under different patterns of loudness. Similarly, disruption of pattern completion should affect sequential inference in noise. Indeed, if pattern completion depends on recurrent processing that occurs after an initial feedforward sweep of activity, then experiments using masking or very rapid presentation may be able to manipulate the strength of temporal pattern completion (Bruner and Potter, 1964; Potter et al., 2014).

Our framework also predicts that timescales of pooling, adaptation and pattern completion should covary. So if region A pools information over a longer timescale than region B, then region A should also exhibit temporal normalization (i.e. adaptation) and pattern completion over longer time windows than region B.

A final prediction of our framework is that different cortical areas use common circuit mechanisms to implement pooling, normalization and pattern completion. The similarities of circuitry across cortical stages suggest that, if these temporal operations are indeed performed hierarchically, then it would be parsimonious for them to be implemented in similar circuit-level mechanism. O'Reilly et al. (2017) have recently proposed a multi-level scheme that combines spatial and temporal operations, including pooling, normalization and pattern completions.

CONCLUSION

Our brains employ a hierarchical architecture for extracting useful information from spatiotemporal input. Computational neuroscience and machine learning research have yielded a set of spatial operations – including pooling, normalization and pattern completion – that provide functional advantages when implemented hierarchically in models of image analysis. We have proposed that analogous operations are implemented hierarchically for temporal information processing in the cerebral cortex.

More targeted modeling work will be required to test whether these operations truly yield functional advantages when implemented hierarchically (Jung et al., 2015; Chung et al., 2016) and to test whether quantitatively parameterized models of temporal pooling, normalization and pattern completion are plausible matches to neural data (Zhou et al., 2017). This is only an initial step, but we are optimistic that principles developed to understand multi-stage spatial computation will generalize to our understanding of spatiotemporal information processing in the cerebral cortex.

CONFLICT OF INTERESTS

The authors declare that no competing financial interests exist

ACKNOWLEDGMENT

The authors gratefully acknowledge support from the National Institutes of Mental Health (R01 MH111439-01, subaward CJH) and a Sloan Foundation (Research Fellowship to CJH).

REFERENCES

Abbott LF, Varela JA, Sen K, Nelson SB (1997) Synaptic depression and cortical gain control. Science 275:220–224.

Altmann CF, Nakata H, Noguchi Y, Inui K, Hoshiyama M, Kaneoke Y, Kakigi R (2008) Temporal dynamics of adaptation to natural sounds in the human auditory cortex. Cereb Cortex 18 (6):1350–1360.

Araya R, Eisenthal KB, Yuste R (2006) Dendritic spines linearize the summation of excitatory potentials. Proc Natl Acad Sci 103 (49):18799–18804.

Astikainen P, Ruusuvirta T, Wikgren J, Penttonen M (2006) Memory-based detection of rare sound feature combinations.... NeuroReport 17(14):1561–1564.

Baldassano C, Chen J, Zadbood A, Pillow JW, Hasson U, Norman KA (2017) Discovering event structure in continuous narrative perception and memory. Neuron 95:709–721.

Barak O, Tsodyks M (2014) Working models of working memory. Curr Opin Neurobiol 25:20–24.

Baria AT, Mansour A, Huang L, Baliki MN, Cecchi GA, Mesulam MM, Apkarian AV (2013) Linking human brain local activity fluctuations to structural and functional network architectures. NeuroImage 73:144–155.

Barascud N, Pearce MT, Griffiths TD, Friston KJ, Chait M (2016) Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. Proc Natl Acad Sci 113(5): E616–E625.

Barlow HB (1958) Temporal and spatial summation in human vision at different background intensities. J Physiol 141(2):337–350.

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for predictive coding. Neuron 76 (4):695–711.

Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L (2009) Neural signature of the conscious processing of auditory regularities. Proc Natl Acad Sci 106(5):1672–1677.

Bendixen A, SanMiguel I, Schröger E (2012) Early electrophysiological indicators for predictive processing in audition: a review. Int J Psychophysiol 83(2):120–131.

Bernacchia A, Seo H, Lee D, Wang XJ (2011) A reservoir of time constants for memory traces in cortical neurons. Nat Neurosci 14 (3):366–372.

Blank I, Balewski Z, Mahowald K, Fedorenko E (2016) Syntactic processing is distributed across the language system. NeuroImage 127:307–323.

Boureau YL, Ponce J, LeCun Y (2010) A theoretical analysis of feature pooling in visual recognition. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp. 111–118.

Boutros NN, Torello MW, Barker BA, Tueting PA, Wu S-C, Nasrallah HA (1995) The P50 evoked potential component and mismatch detection in normal volunteers: implications for the study of sensory gating. Psychiatry Res 57(1):83–88.

Bradbury J, Merity S, Xiong C, Socher R (2016) Quasi-recurrent neural networks. arXiv preprint arXiv:1611.01576.

Bruner JS, Potter MC (1964) Interference in visual recognition. Science 144(3617):424–425.

Buffart H, Leeuwenberg E, Restle F (1981) Coding theory of visual pattern completion. J Exp Psychol Hum Percept Perform 7 (2):241.

Carandini M, Heeger DJ (2012) Normalization as a canonical neural computation. Nat Rev Neurosci 13(1):51–62.

Chaudhuri R, Knoblauch K, Gariel MA, Kennedy H, Wang XJ (2015) A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. Neuron 88(2):419–431.

Chennu S, Noreika V, Gueorguiev D, Blenkmann A, Kochen S, Ibáñez A, Ellipsis Bekinschtein TA (2013) Expectation and attention in hierarchical auditory prediction. J Neurosci 33 (27):11194–11205.

Chung J, Ahn S, Bengio Y (2016) Hierarchical multiscale recurrent neural networks. arXiv preprint arXiv:1609.01704.

Creutzfeldt OD (1977) Generality of the functional structure of the neocortex. Naturwissenschaften 64(10):507–517.

Czigler I (2007) Visual mismatch negativity. J Psychophysiol 21(3–4):224–230.

Dayan P, Hinton GE, Neal RM, Zemel RS (1995) The Helmholtz machine. Neural Comput 7(5):889–904.

de Heer WA, Huth AG, Griffiths TL, Gallant JL, Theunissen FE (2017) The hierarchical cortical organization of human speech processing. J Neurosci 37(27):6539–6557.

De Weerd P, Gattass R, Desimone R, Ungerleider LG (1995) Responses of cells in monkey visual cortex during perceptual filling-in of an artificial scotoma. Nature 377(6551):731–734.

Díaz-Quesada M, Maravall M (2008) Intrinsic mechanisms for adaptive gain rescaling in barrel cortex. J Neurosci 28 (3):696–710.

Douglas RJ, Martin KA (2018) Canonical cortical circuits. In: Shepherd GM, Grillner S, editors. Handbook of brain microcircuits. Oxford University Press. p. 23–34.

Duarte R, Seeholzer A, Zilles K, Morrison A (2017) Synaptic patterning and the timescales of cortical dynamics. Curr Opin Neurobiol 43:156–165.

Dürschmid S, Edwards E, Reichert C, Dewar C, Hinrichs H, Heinze H-J, Ellipsis Knight RT (2016) Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. Proc Natl Acad Sci 113(24):6755–6760.

Exner S (1875) Experimentelle Untersuchung der einfachsten psychischen Processe. Archiv für die gesamte Physiologie des Menschen und der Tiere 11(1):581–602.

Fadiga L, Craighero L, D'Ausilio A (2009) Broca's area in language, action, and music. Ann N Y Acad Sci 1169(1):448–458.

Friederici AD, Steinhauer K, Pfeifer E (2002) Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. Proc Natl Acad Sci 99(1):529–534.

Friston K, Kiebel S (2009) Predictive coding under the free-energy principle. Philos Trans R Soc B Biol Sci 364(1521):1211–1221.

Fukushima K, Miyake S (1982) Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition. In: Competition and cooperation in neural nets, pp. 267–285 Berlin, Heidelberg: Springer.

Fuster JM (1997) Network memory. Trends Neurosci 20 (10):451–459.

Gal A, Eytan D, Wallach A, Sandler M, Schiller J, Marom S (2010) Dynamics of excitability over extended timescales in cultured cortical neurons. J Neurosci 30(48):16332–16342.

Garrido MI, Kilner JM, Stephan KE, Friston KJ (2009) The mismatch negativity: a review of underlying mechanisms. Clin Neurophysiol 120(3):453–463.

Gauthier B, Eger E, Hesselmann G, Giraud AL, Kleinschmidt A (2012) Temporal tuning properties along the human ventral visual stream. J Neurosci 32(41):14433–14441.

Gavornik JP, Bear MF (2014) Learned spatiotemporal sequence recognition and prediction in primary visual cortex. Nat Neurosci 17(5):732–737.

Gehring J, Auli M, Grangier D, Yarats D, Dauphin YN (2017) Convolutional sequence to sequence learning. arXiv preprint arXiv:1705.03122.

Gibson J. Jerome, Reed E, Jones R (1982) Reasons for realism: selected essays of James J. Gibson. Hillsdale, N.J.: L. Erlbaum.

Gjorgjieva J, Drion G, Marder E (2016) Computational implications of biophysical diversity and multiple timescales in neurons and synapses for circuit performance. Curr Opin Neurobiol 37:44–52.

Graybiel AM (1998) The basal ganglia and chunking of action repertoires. Neurobiol Learn Mem 70(1):119–136.

Grill-Spector K, Malach R (2004) The human visual cortex. Annu Rev Neurosci 27:649–677.

Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N (2008) A hierarchy of temporal receptive windows in human cortex. J Neurosci 28 (10):2539–2550.

Hasson U, Chen J, Honey CJ (2015) Hierarchical process memory: memory as an integral component of information processing. Trends Cogn Sci 19(6):304–313.

Hawkins J, Blakeslee S (2007) On intelligence: How a new understanding of the brain will lead to the creation of truly intelligent machines. Macmillan.

Heeger DJ (2017) Theory of cortical function. Proc Natl Acad Sci 114 (8):1773–1782.

Heilbron M, Chait M (2017) Great expectations: is there evidence for predictive coding in auditory cortex? Neuroscience.

Henson RN (2016) Repetition suppression to faces in the fusiform face area: a personal and dynamic journey. Cortex 80:174–184.

Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. Neural Comput 18(7):1527–1554.

Homann J, Koy SA, Glidden A, Tank D, Berry MJ (2017) Predictive coding of novel versus familiar stimuli in the primary visual cortex. biorxiv.

Honey CJ, Thesen T, Donner TH, Silbert LJ, Carlson CE, Devinsky O, et al. (2012) Slow cortical dynamics and the accumulation of information over long timescales. Neuron 76(2):423–434.

Honey CJ, Newman EL, Schapiro AC (2017) Switching between internal and external modes: a multi-scale learning principle. Network Neurosci. https://doi.org/10.1162/netn_a_00024.

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci 79 (8):2554–2558.

Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. J Physiol 148(3):574–591.

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160 (1):106–154.

Huntenburg JM, Bazin PL, Margulies DS (2017) Large-scale gradients in human cortical organization. Trends Cogn Sci.

Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259.

Jarrett K, Kavukcuoglu K, LeCun Y (2009) What is the best multi-stage architecture for object recognition?. In: Computer Vision, 2009 IEEE 12th International Conference on, pp. 2146–2153 IEEE.

Jung M, Hwang J, Tani J (2015) Self-organization of spatio-temporal hierarchy via learning of dynamic visual image patterns on action sequences. PLoS ONE 10(7):e0131214.

Kaposvari P, Kumar S, Vogels R (2016) Statistical learning signals in macaque inferior temporal cortex. Cereb Cortex:1–17.

Kiebel SJ, Daunizeau J, Friston KJ (2008) A hierarchy of time-scales and the brain. PLoS Comput Biol 4(11):e1000209.

Krekelberg B, Boynton GM, van Wezel RJA (2006) Adaptation: from single cells to BOLD signals. Trends Neurosci 29(5):250–256.

Kourtzi Z, Huberle E (2005) Spatiotemporal characteristics of form analysis in the human visual cortex revealed by rapid event-related fMRI adaptation. NeuroImage 28(2):440–452.

Kourtzi Z, Connor CE (2011) Neural representations for object perception: structure, category, and adaptive coding. Annu Rev Neurosci 34:45–67.

Kremláček J, Kreegipuu K, Tales A, Astikainen P, Põldver N, Näätänen R, Stefanics G (2016) Visual mismatch negativity (vMMN): a review and meta-analysis of studies in psychiatric and neurological disorders. Cortex 80:76–112.

Kubovy M, Van Valkenburg D (2001) Auditory and visual objects. Cognition 80(1):97–126.

Kukushkin NV, Carew TJ (2017) Memory takes time. Neuron 95 (2):259–279.

Kutas M, Federmeier KD (2011) Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). Annu Rev Psychol 62:621–647.

LeCun Y, Kavukcuoglu K, Farabet C (2010) Convolutional networks and applications in vision. In: Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on, pp. 253–256 IEEE.

Lee TS, Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. J Opt Soc Am A Opt Image Sci Vis 20 (7):1434–1448.

Leonard MK, Bouchard KE, Tang C, Chang EF (2015) Dynamic encoding of speech sequence probability in human temporal cortex. J Neurosci 35(18):7203–7214.

Lerner Y, Honey CJ, Silbert LJ, Hasson U (2011) Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. J Neurosci 31(8):2906–2915.

Lerner Y, Honey CJ, Katkov M, Hasson U (2014) Temporal scaling of neural responses to compressed and dilated natural speech. J Neurophysiol 111(12):2433–2444.

Lü ZL, Williamson SJ, Kaufman L (1992) Human auditory primary and association cortex have differing lifetimes for activation traces. Brain Res 572(1):236–241.

MacKay DG, Stewart R, Burke DM (1998) HM revisited: relations between language comprehension, memory, and the hippocampal system. J Cogn Neurosci 10(3):377–394.

Malinowska U, Crone NE, Lenz FA, Cervenka M, Boatman-Reich D (2017) Multi-regional adaptation in human auditory association cortex. Front Hum Neurosci 11.

Malmierca MS, Cristaudo S, Pérez-González D, Covey E (2009) Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. J Neurosci 29(17):5483–5493.

Malmierca MS, Sanchez-Vives MV, Escera C, Bendixen A (2014) Neuronal adaptation, novelty detection and regularity encoding in audition. Front Syst Neurosci 8.

May PJC, Tiitinen H (2010) Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. Psychophysiology 47(1):66–122.

Mesulam MM (1998) From sensation to cognition. Brain 121 (6):1013–1052.

Mongillo G, Barak O, Tsodyks M (2008) Synaptic theory of working memory. Science 319(5869):1543–1546.

Mumford D (1992) On the computational architecture of the neocortex. Biol Cybern 66(3):241–251.

Murray JD, Bernacchia A, Freedman DJ, Romo R, Wallis JD, Cai X, Wang XJ (2014) A hierarchy of intrinsic timescales across primate cortex. Nat Neurosci 17(12):1661–1663.

Näätänen R, Gaillard AWK, Mäntysalo S (1978) Early selective-attention effect on evoked potential reinterpreted. Acta Psychol 42 (4):313–329.

Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Ellipsis Alho K (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 385:432–434.

Näätänen R, Paavilainen P, Rinne T, Alho K (2007) The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin Neurophysiol 118(12):2544–2590.

Natan RG, Briguglio JJ, Mwilambwe-Tshilobo L, Jones SI, Aizenberg M, Goldberg EM, Geffen MN (2015) Complementary control of sensory adaptation by two types of cortical interneurons. Elife 4: e09868.

Nieto-Diego J, Malmierca MS (2016) Topographic distribution of stimulus-specific adaptation across auditory cortical fields in the anesthetized rat. PLoS Biol 14(3):e1002397.

Nordby H, Hammerborg D, Roth WT, Hugdahl K (1994) ERPs for infrequent omissions and inclusions of stimulus elements. Psychophysiology 31(6):544–552.

Ogawa T, Komatsu H (2010) Differential temporal storage capacity in the baseline activity of neurons in macaque frontal eye field and area V4. J Neurophysiol 103(5):2433–2445.

Parras GG, Nieto-Diego J, Carbajal V, Valdés-Baizabal C, Escera C, Malmierca MS (2017) Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. Nat Commun 8:2148.

Pazo-Alvarez P, Cadaveira F, Amenedo E (2003) MMN in the visual modality: a review. Biol Psychol 63(3):199–236.

Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. Speech Commun 41(1):245–255.

Potter MC, Wyble B, Hagmann CE, McCourt ES (2014) Detecting meaning in RSVP at 13 ms per picture. Atten Percept Psychophys 76(2):270–279.

Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of ‘‘what” and ‘‘where” in auditory cortex. Proc Natl Acad Sci 97(22):11800–11806.

Rosburg T, Boutros NN, Ford JM (2008) Reduced auditory evoked potential component N100 in schizophrenia—A critical review. Psychiatry Res 161(3):259–274.

Runyan CA, Piasini E, Panzeri S, Harvey CD (2017) Distinct timescales of population coding across cortex. Nature.

Rüsseler J, Altenmüller E, Nager W, Kohlmetz C, Münte TF (2001) Event-related brain potentials to sound omissions differ in musicians and non-musicians. Neurosci Lett 308(1):33–36.

Saberi K, Perrott DR (1999) Cognitive restoration of reversed speech. Nature 398(6730):760.

Sassenhagen J, Schlesewsky M, Bornkessel-Schlesewsky I (2014) The P600-as-P3 hypothesis revisited: single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. Brain Lang 137:29–39.

Schapiro AC, Turk-Browne NB, Norman KA, Botvinick MM (2016) Statistical learning of temporal community structure in the hippocampus. Hippocampus 26(1):3–8.

Scholl BJ (2001) Objects and attention: the state of the art. Cognition 80(1):1–46.

Sekuler R (1996) Motion perception: a modern view of Wertheimer’s 1912 monograph. Perception 25(10):1243–1258.

Serre T, Oliva A, Poggio T (2007) A feedforward architecture accounts for rapid categorization. Proc Natl Acad Sci 104 (15):6424–6429.

Singer Y, Teramoto Y, Willmore B, Schnupp JW, King AJ, Harper NS (2017) Sensory cortex is optimised for prediction of future input. bioRxiv 224758.

Stefanics G, Kimura M, Czigler I (2011) Visual mismatch negativity reveals automatic detection of sequential regularity violation. Front Hum Neurosci 5.

Stephens GJ, Honey CJ, Hasson U (2013) A place for time: the spatiotemporal structure of neural dynamics during natural audition. J Neurophysiol 110(9):2019–2026.

Please cite this article in press as: Himberger KD et al. Principles of Temporal Processing Across the Cortical Hierarchy. Neuroscience (2018), https://doi.org/10.1016/j.neuroscience.2018.04.030

Stockman A, MacLeod DI, DePriest DD (1991) The temporal properties of the human short-wave photoreceptors and their associated pathways. Vision Res 31(2):189–208.

Stokes MG (2015) 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. Trends Cogn Sci 19 (7):394–405.

Summerfield C, Koechlin E (2008) A neural representation of prior information during perceptual inference. Neuron 59(2):336–347.

Symonds RM, Lee WW, Kohn A, Schwartz O, Witkowski S, Sussman ES (2017) Distinguishing neural adaptation and predictive coding hypotheses in auditory change detection. Brain Topogr 30 (1):136–148.

Tervaniemi M, Saarinen J, Paavilainen P, Danilova N, Näätänen R (1994) Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. Biol Psychol 38 (2):157–167.

Todorovic A, van Ede F, Maris E, de Lange FP (2011) Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. J Neurosci 31(25):9118–9123.

Uhrig L, Dehaene S, Jarraya B (2014) A hierarchy of responses to auditory regularities in the macaque brain. J Neurosci 34 (4):1127–1132.

Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. Nat Neurosci 6(4).

Ulanovsky N, Las L, Farkas D, Nelken I (2004) Multiple time scales of adaptation in auditory cortex neurons. J Neurosci 24 (46):10440–10453.

Uusitalo MA, Williamson SJ, Seppä MT (1996) Dynamical organisation of the human visual system revealed by lifetimes of activation traces. Neurosci Lett 213(3):149–152.

Van Atteveldt NM, Blau VC, Blomert L, Goebel R (2010) FMR-adaptation indicates selectivity to audiovisual content congruency in distributed clusters in human superior temporal cortex. BMC Neurosci 11:11.

Van Berkum JJA (2009) The neuropragmatics of ''simple'' utterance comprehension: an ERP review. In: Sauerland U, Yatsushiro K, editors. Semantics and pragmatics: From experiment to theory. Basingstoke: Palgrave Macmillan. p. 276–316.

Von Melchner L, Pallas SL, Sur M (2000) Visual behaviour mediated by retinal projections directed to the auditory pathway. Nature 404 (6780):871–876.

Wacongne C, Labyt E, van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S (2011) Evidence for a hierarchy of predictions and prediction errors in human cortex. Proc Natl Acad Sci 108(51):20754–20759.

Wang XJ (1999) Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. J Neurosci 19 (21):9587–9603.

Watson AB (1986) Temporal sensitivity. In: Boff K, Kaufman L, Thomas JP, editors. Handbook of perception and human performance. New York, NY: Wiley. p. 1–43.

Weiner KS, Sayres R, Vinberg J, Grill-Spector K (2010) FMRI-adaptation and category selectivity in human ventral temporal cortex: regional differences across time scales. J Neurophysiol 103(6):3349–3365.

Wertheimer M (1912) Experimentelle Studium uber das Sehen von Bewegung. Zeitschrift fur Psychologie 61(3):161–265.

Xu J, Kemeny S, Park G, Frattali C, Braun A (2005) Language in context: emergent features of word, sentence, and narrative comprehension. NeuroImage 25(3):1002–1015.

Yamins DL, DiCarlo JJ (2016) Using goal-driven deep learning models to understand sensory cortex. Nat Neurosci 19 (3):356–365.

Yuille A, Kersten D (2006) Vision as Bayesian inference: analysis by synthesis? Trends Cogn Sci 10(7):301–308.

Zacks JM, Tversky B, Iyer G (2001) Perceiving, remembering, and communicating structure in events. J Exp Psychol Gen 130(1):29.

Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. In: European conference on computer vision, pp. 818–833 Cham: Springer.

Zhou J, Benson NC, Kay K, Winawer J (2017) Compressive temporal summation in human visual cortex. J Neurosci:1724–1817.

(Received 5 August 2017, Accepted 19 April 2018)
(Available online xxxx)