

Central auditory neurons display flexible feature recombination functions

Andrei S. Kozlov and Timothy Q. Gentner

J Neurophysiol 111:1183-1189, 2014. First published 18 December 2013; doi:10.1152/jn.00637.2013

You might find this additional info useful...

This article cites 29 articles, 11 of which can be accessed free at:

</content/111/6/1183.full.html#ref-list-1>

Updated information and services including high resolution figures, can be found at:

</content/111/6/1183.full.html>

Additional material and information about *Journal of Neurophysiology* can be found at:

<http://www.the-aps.org/publications/jn>

This information is current as of October 28, 2014.

Central auditory neurons display flexible feature recombination functions

Andrei S. Kozlov¹ and Timothy Q. Gentner^{1,2,3}

¹Department of Psychology, University of California San Diego, La Jolla, California; ²Neurosciences Program, University of California San Diego, La Jolla, California; ³Kavli Institute for Brain and Mind, La Jolla, California

Submitted 9 September 2013; accepted in final form 16 December 2013

Kozlov AS, Gentner TQ. Central auditory neurons display flexible feature recombination functions. *J Neurophysiol* 111: 1183–1189, 2014. First published December 18, 2013; doi:10.1152/jn.00637.2013.—Recognition of natural stimuli requires a combination of selectivity and invariance. Classical neurobiological models achieve selectivity and invariance, respectively, by assigning to each cortical neuron either a computation equivalent to the logical “AND” or a computation equivalent to the logical “OR.” One powerful OR-like operation is the MAX function, which computes the maximum over input activities. The MAX function is frequently employed in computer vision to achieve invariance and considered a key operation in visual cortex. Here we explore the computations for selectivity and invariance in the auditory system of a songbird, using natural stimuli. We ask two related questions: does the MAX operation exist in auditory system? Is it implemented by specialized “MAX” neurons, as assumed in vision? By analyzing responses of individual neurons to combinations of stimuli we systematically sample the space of implemented feature recombination functions. Although we frequently observe the MAX function, we show that the same neurons that implement it also readily implement other operations, including the AND-like response. We then show that sensory adaptation, a ubiquitous property of neural circuits, causes transitions between these operations in individual neurons, violating the fixed neuron-to-computation mapping posited in the state-of-the-art object-recognition models. These transitions, however, accord with predictions of neural-circuit models incorporating divisive normalization and variable polynomial nonlinearities at the spike threshold. Because these biophysical properties are not tied to a particular sensory modality but are generic, the flexible neuron-to-computation mapping demonstrated in this study in the auditory system is likely a general property.

adaptation; auditory system; MAX function; object-recognition models

ANY SUCCESSFUL PATTERN-RECOGNITION system, natural or artificial, must combine two key properties: selectivity and invariance. Selectivity is important because it allows the system to distinguish different stimuli; invariance is essential because the physical content of the signal can vary greatly, but not all variations are always informative. For instance, we can recognize our name spoken by a man or a woman, slowly or rapidly, in a quiet room or on a busy street. No artificial speech recognition algorithm today can match the performance of the human auditory system, in part because combining selectivity with invariance is in general a very hard problem (Rifkin et al. 2007; DiCarlo et al. 2012). How does the brain solve this problem? This question has been studied most extensively at the algorithmic level in natural and computer vision.

A powerful class of pattern-recognition algorithms inspired by advances in visual neuroscience posit that two basic oper-

ations, or feature recombination functions, implemented in individual neurons and equivalent to the logical “AND” and the logical “OR,” can provide both selectivity and invariance, respectively (Hubel and Wiesel 1962; Fukushima 1980; Riesenhuber and Poggio 1999; Serre et al. 2007; Walther and Koch 2007). Models in this class have been very successful because they reflect the hierarchical organization of the visual cortex, can learn from a limited number of examples and generalize well, and can match the performance of humans on rapid visual object-recognition tasks.

A crucial feature of the model architecture is a strict division of labor between two types of neurons: those that perform the AND-like operations for selectivity and those that perform the OR-like operations for invariance. Operations similar to the logical “AND” are carried out by model components akin to simple cells tuned to respond supralinearly to an appropriate combination of multiple inputs. These operations build more refined representations from coarser ones. Operations similar to the logical “OR” are carried out by complex-like cells that respond sublinearly to a combination of multiple inputs. Instantiations of the logical “OR” (formally defined only for binary inputs), such as the MAX operation that computes the maximum over input activities (Riesenhuber and Poggio 1999; Serre et al. 2007; Kouh and Poggio 2007), are computationally powerful and used in a number of different state-of-the-art multistage visual architectures including HMAX nets (biologically inspired hierarchical networks using the MAX operation for invariance) (Serre et al. 2007), convolutional nets (neural networks using local receptive fields and shared weights among units in a given layer) (Boureau et al. 2010; Lee et al. 2009), and spatial pyramids, a type of support-vector machines (Yang et al. 2009). The power of the MAX-like operations is in causing a strong input, activated by a preferred stimulus of the neuron, to control the magnitude of the response regardless of weaker inputs, such as those activated by other stimuli present at the same time. Ignoring irrelevant sensory input, or distracters in complex visual or auditory scenes, is the essence of invariance and crucial to object recognition.

A number of studies using single-neuron recordings have quantified the operations for combining visual features in cortical areas V1, V4, and IT, and provided evidence for the MAX operation (Sato 1989; Gawne and Martin 2002; Lampl et al. 2004). These studies used the “summation index” (*Sml*) as a metric to quantify feature recombination functions (Sato 1989; Lampl et al. 2004). The *Sml* for a given neuron and pair of stimuli, “A” and “B,” captures the relationship between responses to these stimuli presented separately, R_A and R_B , and together (simultaneously) R_{A+B} , as $Sml = R_{A+B} - MAX(R_A, R_B) / MIN(R_A, R_B)$, where *MAX* takes the larger of the two responses and *MIN* the smaller. In the

Address for reprint requests and other correspondence: A. S. Kozlov, Dept. of Psychology, Univ. of California San Diego, La Jolla, CA 92093 (e-mail: akozlov@ucsd.edu).

visual cortex, *Sml* distributions are broad, include *Smls* >1 indicating supralinear (AND-like) summation (Lampl et al. 2004), and centered at zero indicating a MAX-like operation (Sato 1989; Lampl et al. 2004).

The biological instantiation of these basic neural computations, however, remains poorly understood. Although it has been suggested that different neural systems may share these operations, including the MAX function (Poggio and Bizzi 2004), they have not been investigated outside of the mammalian visual system. Therefore, it remains unclear whether they are specialized computations restricted either to vision or the circuitry of mammalian cortex. Likewise, other models, focused on implementation of these computations, predict that the same circuits may permit both AND-like and OR-like operations, instantiated as different parameter regimes of the same underlying function, e.g., the *softmax* (Kouh and Poggio 2008). Whether individual neurons operate in these different regimes, for example, when a given neuron is processing different stimuli, or the same stimuli under different conditions, or whether they are instead constrained by the cortical circuitry to a specific subset of operations, as postulated by the HMAX model, has not been studied experimentally. To address this question is the first aim of this study. To verify, using natural stimuli, whether the MAX-like operation exists in auditory system is the second aim of this study.

We investigated whether the central auditory system achieves a combination of selectivity and invariance using the same basic computations as the mammalian visual system. We obtained *Smls* from well-isolated single neurons in the caudomedial nidopallium (NCM) and caudal mesopallium (CM) of the European starling (*Sturnus vulgaris*), a songbird with richly patterned songs (Eens 1997). The NCM and CM are functionally, morphologically, and hodologically similar to secondary auditory cortices in mammals (Butler et al. 2011; Dugas-Ford et al. 2012) and contain neurons that respond selectively to natural songs on the time scales of tens to hundreds of milliseconds (Gentner and Margoliash 2003; Theunissen and Shavit 2006; Thompson and Gentner 2012). These nuclei are among the last purely sensory areas, situated at the highest level in the avian hierarchy of auditory processing and implicated in song recognition, analogous in this respect to the high-order visual areas V4 and IT in the visual ventral stream implicated in object recognition.

Using natural stimuli (see METHODS) we first determined the shape of the *Sml* distribution, and evaluated evidence for AND-like and OR-like operations, in particular the MAX function. We then asked whether individual neurons are specialized for a single computation, as assumed in object-recognition models, or whether selectivity (AND-like) and invariance (OR-like) computations can occur in the same neuron and, if so, what factors may cause a transition between these different computations.

METHODS

Under a protocol approved by the Institutional Animal Care and Use Committee of the University of California, San Diego, we performed experiments on adult European starlings (*Sturnus vulgaris*) of both sexes. Birds were anesthetized with urethane (7 ml/kg) and head-fixed in a stereotaxic apparatus using a small metal pin attached to the skull. Auditory stimuli were excerpts of starling songs and

ranged in duration from 0.3 to 2 s (corresponding to the time scale of notes or motifs in a starling song). We recorded starling songs from males in the presence of a female inside a sound attenuation box (Acoustic Systems, Austin, TX). All stimuli were played to the subjects placed inside a sound attenuation box at 60-dB mean level. Action potentials were recorded extracellularly using 16-channel and 32-channel electrode arrays (NeuroNexus Technologies, Ann Arbor, MI) inserted through a small craniotomy into the caudal mesopallium or NCM. Stimulus presentation, signal recording, and spike sorting were controlled through a PC using Spike2 software (CED, Cambridge, UK). Extracellular voltage waveforms were amplified with 16-channel amplifiers (model 3600, A-M Systems, Sequim, WA), filtered and sampled with a 50- μ s resolution, and saved for offline spike sorting. Single units were identified by principle components of the spike waveforms, only when no violations of the refractory period (assumed to equal 1 ms) occurred, and only from recordings with an excellent signal-to-noise ratio (large-amplitude extracellular action-potential waveforms). Peristimulus time histograms (see Figs. 1 and 3) were constructed by assorting identified spikes in 20-ms bins, and the number of spikes in each bin was normalized by the number of stimulus repetitions (from 10 to 100, usually 30). All analyses, except for spike sorting, were performed in Matlab (MathWorks, Natick, MA).

Stimuli were played separately and together (superimposed) in pair-wise combinations to determine the associated responses R_A , R_B , R_{A+B} , and *Smls* were calculated using the equation $Sml = R_{A+B} - MAX(R_A, R_B) / MIN(R_A, R_B)$, as described in the Introduction. This design parallels directly the simultaneous presentation of stimuli used in *Sml* determination in the visual studies, and it mimics an ethologically relevant situation of sound processing in noisy environments with overlapping sounds.

Throughout our analysis, we wanted to be as conservative as possible when assigning the MAX computation to a neuron. First, to avoid response saturation, we did not select song segments that (when presented alone) evoked maximal firing rates. Relative response duration is also important to control, because a longer response that contributes more total spikes will tend to dominate the combined response, R_{A+B} , in the *Sml* calculation. To eliminate any possibility of a longer response biasing the operation towards the MAX, we first grouped individual stimuli into pairs based on similarity in the duration of the associated neuronal responses. Within each stimulus pair (*A* and *B*), we measured R_A , R_B , and R_{A+B} by counting spikes over the same time interval determined by the shorter response R_A or R_B within each pair. Third, we made sure that any putative MAX response was not an artifact of having simply missed an additive response by not properly aligning the two individual stimuli in time. We therefore measured responses to many versions of the same two segments superimposed with different temporal alignments. Choosing an appropriate time increment by which to vary the temporal alignment of the stimuli was an important part of the design. On one hand, the increment needs to be small enough to allow peaks in subthreshold responses [unobserved excitatory postsynaptic potential (EPSPs)] to overlap with each other (e.g., a submillisecond time increment would guarantee such an overlap). On the other hand, the time increment cannot be so small that it requires a prohibitively large number of stimuli to cover the entire range of segment overlap. An appropriate time scale is provided by the intrinsic time scale of temporal summation, the time constant of a neuron, which is on the order of a few to a few hundred milliseconds (Bernander et al. 1991).

RESULTS

We determined *Sml* values in well-isolated single auditory neurons, taking several measures to eliminate possible biases in *Sml* determination, as described in METHODS. We presented multiple repetitions (typically 30) of continuous, 1-min long natural starling songs and selected two or more short segments

of those songs (ranging in duration between 0.3 and 2 s) and grouped them into pairs. For each pair (*A* and *B*), we measured R_A , R_B , and R_{A+B} and calculated the *Sml* for that pair of stimuli in that neuron. To determine R_{A+B} , we chose a temporal overlap between two stimuli that would maximize the summation of putative evoked EPSPs by sampling all possible overlaps with a time increment (typically 20 ms; range: 5–50 ms) comparable to the time constant of a neuron *in vivo* (Bernander et al. 1991) and used the maximal response R_{A+B} averaged across all repetitions at the associated overlap (see METHODS).

Our protocol required that neurons be presented with the same stimuli hundreds or thousands of times, which, if done rapidly, will cause long-lasting response desensitization in NCM (Chew et al. 1995). To avoid desensitization, we randomly interspersed the target stimuli between longer epochs of conspecific songs and periods of silence. Collectively, these controls required very long experiments. Because we selected units only with an excellent signal-to-noise ratio, holding times in excess of 12 h were routinely achieved, allowing us to sort spikes and prepare stimuli for each neuron, to obtain multiple *Smls* (i.e., for different stimulus pairs) and to minimize input saturation when determining R_{A+B} , by assessing stimulus-specific adaptation, as described below.

Similar to the results in visual cortex (Lampl et al. 2004), we observed AND-like responses and OR-like responses (Fig. 1, *A* and *B*). The *Sml* distribution across our sample of recorded neurons was broad with a peak at zero (Fig. 2*A*) resembling the *Sml* distribution in visual cortex (Lampl et al. 2004). Many

Smls (54%) were >1 including nine infinite *Smls* (Fig. 2*A*, inset), corresponding to AND-like supralinear summation. The nine *Sml* values in the central bin suggest a MAX-like operation (mean *Sml* = 0.03, SE = 0.04, $n = 7$ cells). To verify this model, we performed a normalized residual analysis to compare the MAX model to other alternatives (linear sum and arithmetic average). The same analysis was used in Lampl et al. (2004) to select between the MAX model and its alternatives. The residual error of the MAX model in our study was one-tenth of the arithmetic-average model error and one-hundredth of the linear-sum model error (Fig. 2*B*).

For *Smls* between 0 and 1, if both stimuli in a pair activate shared inputs, the value could be biased toward 0 (the MAX function) because of input saturation. To avoid input saturation, we tested input independence in the majority of cases (17 of 25) where $Sml < 1$ by searching for pairs of stimuli that produced stimulus-specific adaptation (Fig. 3), which is common in sensory systems and can be used to verify input independence (Movshon and Lennie 1979; Scholl et al. 2010). In particular, for three of the seven neurons that displayed the MAX response in Fig. 2, and for five of the seven different neurons that displayed a broad range of summation rules including the MAX response in Fig. 4 (see below), an adaptation to one stimulus, which could completely eliminate the response to that stimulus (Fig. 3*B*), preserved (or even potentiated) the response to the other stimulus in a pair, and vice versa. This key control could not be performed for all the *Smls* because not every stimulus produced adaptation in every neu-

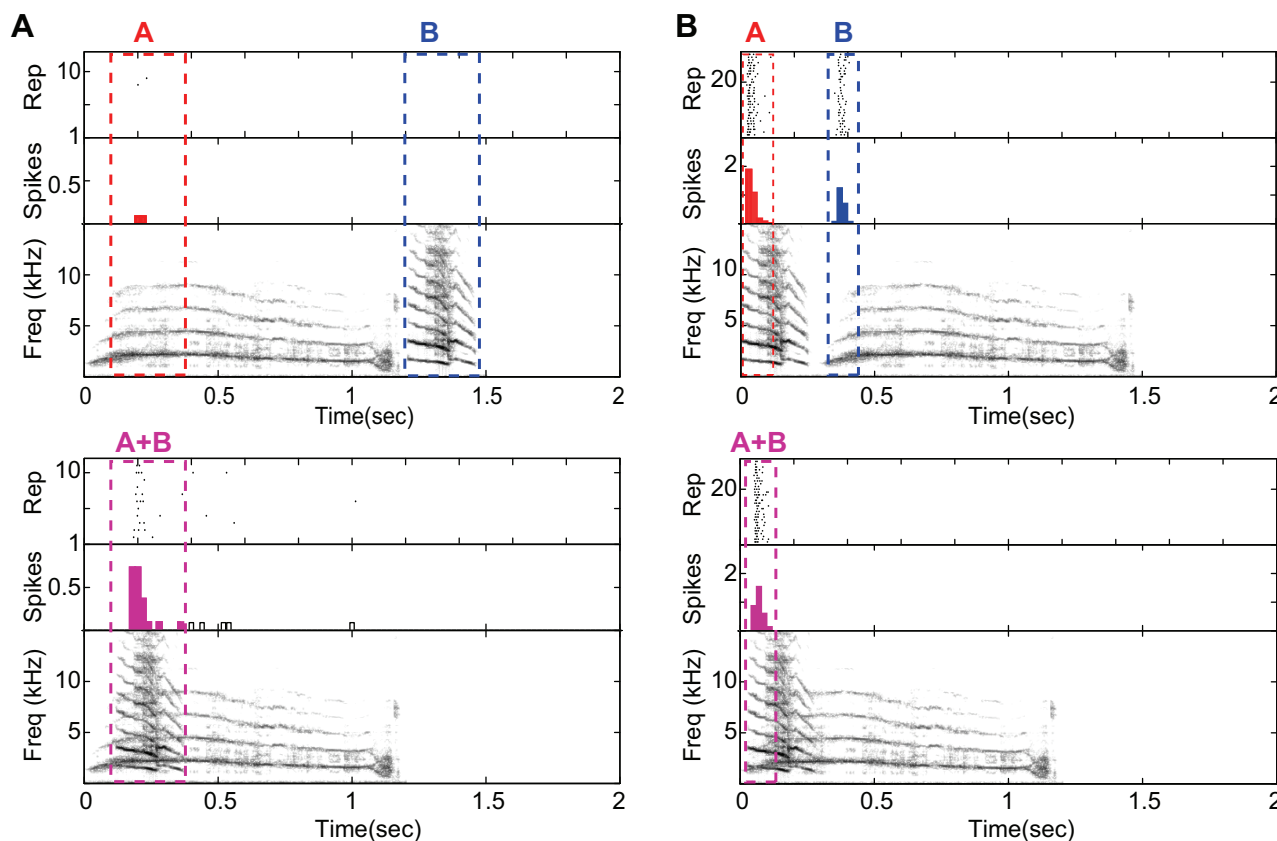


Fig. 1. Examples of “OR”-like and “AND”-like responses in individual auditory neurons. Spike raster plots, peristimulus histograms, and spectrograms for 2 stimuli (“A” and “B”) presented sequentially (*top*) and together (*bottom*) illustrating an AND-like response [summation index (*Sml*) = ∞ ; *A*] and an OR-like response ($Sml = -0.05$) in a different neuron (*B*). The ordinate in the histograms indicates the average number of spikes per repetition assorted in 20-ms bins.

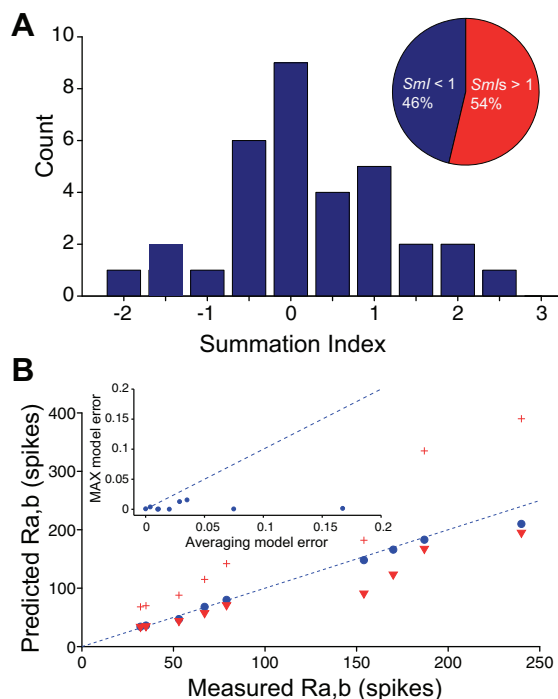


Fig. 2. *Sml* distribution from the population of neurons and quantification of the putative MAX responses. *A*: *Sml* distribution contains 54 values [22 in caudal mesopallium (CM) and 32 in caudo-medial nidopallium (NCM)] from 39 cells in 17 birds. The histogram is truncated at *Sml* = 3 for comparison with a similar distribution in Lampl et al. (2004). *Inset*: chart showing all the 54 *Sml*s, partitioned into those <1 (blue) and those >1 (red). *B*: measured responses R_{A+B} plotted against R_{A+B} predicted by the linear sum (+), arithmetic average (▼) and the MAX models (●) for the nine *Sml* values closest to zero in *A*. The dashed line represents the diagonal (also in the *inset*). *Inset*: normalized squared error SE between the predicted response and the measured response for the MAX model plotted against SE for the arithmetic average model (the error for the linear sum model was 10 times greater than that for the average model and is not shown).

ron. Verifying that the MAX response holds for independent inputs in our opinion greatly strengthens the evidence for this operation. Studies reporting the MAX operation in the visual system did not include this control (Sato 1989; Gawne and Martin 2002; Lampl et al. 2004). As explained above, we also controlled for potential *Sml* biases induced by output saturation and by temporal misalignment of putative evoked EPSPs. Thus we think that our results provide strong and novel evidence for the MAX operation in auditory system.

The hierarchical object-recognition algorithms (Hubel and Wiesel 1962; Fukushima 1980; Riesenhuber and Poggio 1999; Serre et al. 2007) assign to individual neurons either an AND-like operation, or an OR-like operation such as the MAX function, but not both. In contrast, we observed both operations in the same neuron. We obtained widely different *Sml*s in response to different stimuli in seven neurons, and in six of them, we observed both an *Sml* close to zero ($-0.15 < Sml < 0.15$) and an *Sml* >1. In other words, we observed the MAX response but no exclusive “MAX” neurons. This property is predicted by a biophysical model of a canonical neural circuit for nonlinear cortical operations that can approximate multiple operations, including the MAX, through the divisive normalization and variable polynomial nonlinearities (Kouh and Poggio 2008). Variable input-output nonlinearities can arise from voltage noise at the spike threshold (Miller and Troyer 2002),

whereas divisive normalization is considered a canonical neural computation (Carandini and Heeger 2011). Note that we obtained this diversity of feature recombination functions even without varying the amplitude of individual stimuli. Variations in stimulus amplitude can add an additional source of variability to the summation rules.

We have thus found that the same neuron can implement both the AND-like and OR-like operations with different stimuli. The simplest explanation for this result is that different stimuli produce different operations because they activate different inputs, whose strength (i.e., the ability to cause spiking) is different. If this explanation is correct, then the whole range of operations, covering the entire distribution of *Sml* values in Fig. 2*A*, should be observed even with a single pair of stimuli, in the same neuron, provided input strength can be manipulated over a range as broad as with different stimuli. Indeed, it has been shown analytically that neural networks can switch between sublinear and supralinear summation regimes following changes of input strength (Ahmadian et al. 2012).

Because sensory adaptation can change input strength, for example, through synaptic depression, we wondered whether it can control a transition between OR-like and AND-like operations in a given neuron. By using adaptation to manipulate input strength, we confirm this prediction and show a full range of input combination operations executed for the same pair of stimuli in the same neuron. In seven neurons in seven birds, using a different stimulus pair for each neuron, we determined the *Sml* as described above, but with each stimulus (*A*, *B*, and *A + B*) repeated to cause adaptation. In five of these neurons, individual stimuli evoked stimulus-specific adaptation. In the remaining two neurons they did not, and therefore those two neurons could not be used further in this experiment. As the

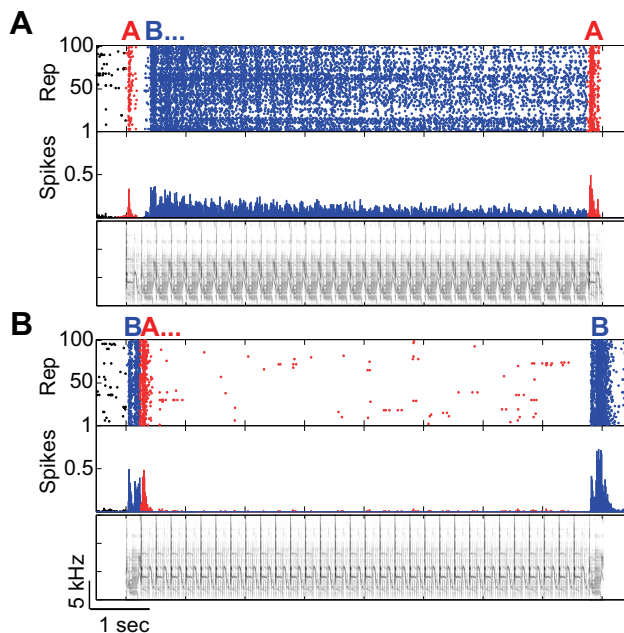


Fig. 3. Stimulus-specific adaptation. *A*: stimulus “A” evokes robust responses (red) both before and immediately after 30 repetitions of stimulus “B” (blue). *B*: same as *A* but with stimuli “A” and “B” interchanged. Note a fast and complete adaptation of responses to stimulus “A” but not to stimulus “B” after only 1 presentation (resembling rapid adaptation in visual cortex: Müller et al. 1999), the distinct kinetics of adaptation to the 2 stimuli, and the absence of cross-adaptation.

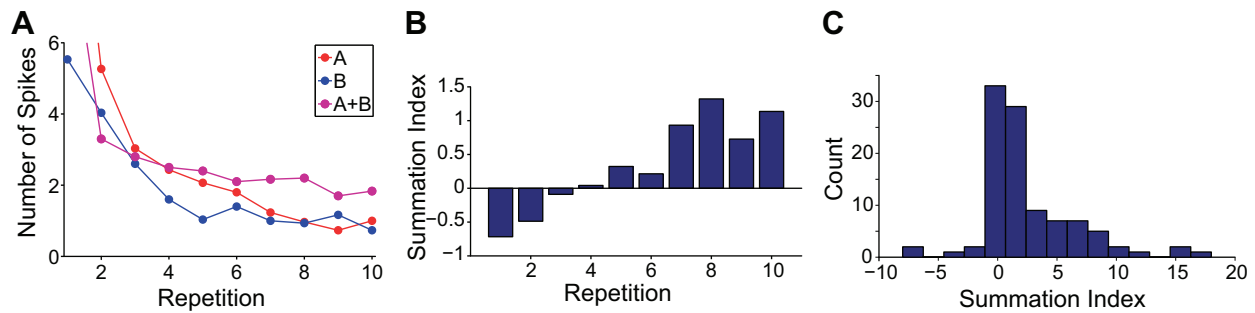


Fig. 4. Feature recombination functions vary broadly with adaptation. *A*: average number of action potentials per repetition obtained with 2 stimuli presented either alone or together with the optimal temporal overlap, determined for this neuron and these stimuli in advance. Each stimulus configuration was repeated to cause adaptation. For each repetition in the sequence, an *Sml* is displayed in *B*. *C*: a pooled distribution of *Smls* for 5 (out of 7 tested) neurons that displayed stimulus-specific adaptation. The distribution contains 106 *Sml* values (5 infinite *Smls* are not depicted), 54% of which are >1.

responses decreased with adaptation (Fig. 4A), the associated *Smls* increased greatly, varying from less than -0.5, to zero, to greater than 1 (Fig. 4B). Similarly broad *Sml* distributions were obtained in the other four neurons. The pooled distribution of *Smls* for the five neurons (Fig. 4C) obtained with the same stimulus for each neuron is very similar to the one obtained with different stimuli in 39 neurons (Fig. 2A). Thus single neurons do not maintain a specific computation over their input and output dynamic ranges. Because adaptation is a ubiquitous property of sensory systems, this result raises a general and important question of whether neuronal computations remain stable during adaption and, if so, how. Even in the absence of adaptation, different operations will be executed on stimuli that activate inputs of different strength. In general, anything that can change input activities may also change how these activities are combined.

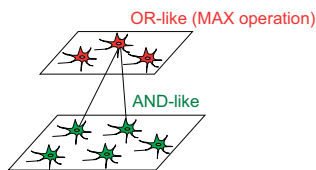
DISCUSSION

This study builds on the insights from natural and computer vision to address two questions (Fig. 5). First, does the MAX

function exist in auditory system? Our results show that the MAX operation is implemented in central auditory neurons. The results, therefore, expand the list of sensory systems known to implement this important operation. Because the MAX operation is a key building block of several powerful visual model architectures, such as HMAX and convolutional nets, we suggest that those models may be extended to auditory pattern recognition.

The second question we address is more general: do neurons implement a singular feature recombination function, as assumed in hierarchical object-recognition models (Hubel and Wiesel 1962; Fukushima 1980; Riesenhuber and Poggio 1999; Serre et al. 2007), or a broad range of functions, as predicted by generic, modality-independent, models of canonical cortical microcircuits (Kouh and Poggio 2008)? Our results show that the key algorithmic division between the two types of neurons, those optimized for selectivity and those optimized for invariance, is not valid in the auditory system. Instead, we find that the same neuron can implement both a MAX-like function, and the logical AND-like function. This result is consistent with

A Fixed neuron-to-computation mapping in visual models (HMAX, convolutional nets)



Our results in the context of previous work

	Visual Models	Visual Experiments	Auditory Experiment (this study)
MAX function	yes	yes	yes
Exclusive MAX neurons	yes	unknown	no

B Flexible neuron-to-computation mapping

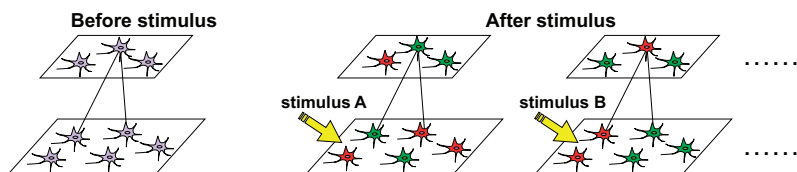


Fig. 5. Schematic representation of our findings and a possible implication for hierarchical object-recognition models. *A, left*: schematic drawing of 2 connected layers representing a fragment of a hierarchical, multilayer, object-recognition model such as HMAX or convolutional nets. Each layer contains units that are assigned a specific operation, such as AND-like Gaussian tuning or a MAX-like function. *A, right*: while we find the MAX function in auditory system, we also find evidence against exclusive “MAX” neurons. Constraining and exploiting this flexibility to produce invariant responses are challenges for future models. *B*: schematic illustration of one potential solution incorporating operational flexibility. *Left*: depiction of the same 2 layers of neurons as in *A* but with a flexible neuron-to-computation mapping. Before a stimulus activates inputs, it is impossible to assign a specific operation to each neuron. When inputs are activated, depending on their activation level, different stimuli (*middle* and *right*) will result in different computation-to-neuron mappings. With sparse coding, there will be little overlap between neuronal populations performing the same set of computations activated by different stimuli. It remains to be tested whether this principle can be implemented and work well for object recognition.

evidence in the literature against the simple/complex cell dichotomy (Mechler and Ringach 2002). It is also consistent with a modeling study of competitive dynamics in cortical responses to multiple visual stimuli that predicted a transition between normalization mode and winner-take-all behavior (Moldakarimov et al. 2005).

Our results demonstrate that changes in the feature recombination functions implemented by any single neuron can be related directly to changes in input strength. Using adaptation to manipulate the strength of inputs activated by the same pair of stimuli in the same neuron, we recapitulate the computational flexibility observed across both neurons and stimuli in our initial experiments. In principle, the flexibility observed here in the auditory system should extend to all sensory modalities, including vision. Indeed, biophysical models of the canonical neural circuits using the *softmax* function (Kouh and Poggio 2008) allow a broad range of feature recombination functions depending on the relative strength of excitatory and inhibitory inputs. Furthermore, a rule phenomenologically equivalent to the logical “AND” describes how weak sub-threshold inputs interact to evoke a spike. Thus, all feature recombination functions must eventually converge to the logical AND-like operation when inputs become sufficiently weak. Nothing in this picture, or in the biophysical models of the canonical neural circuits (Kouh and Poggio 2008), depends on a sensory modality. Therefore, it should apply to all modalities. Nevertheless, additional studies in the visual system are needed to test whether the MAX operation is implemented by exclusive “MAX” neurons.

We emphasize that our study represents a departure from the traditional approach in central auditory neuroscience focused on the description of neuron receptive field properties. Here, instead, we started from a powerful pattern, recognition model, supported by numerous data in visual neuroscience, and tested its key predictions, the existence of the MAX operation and of the exclusive “MAX” neurons implementing this operation-in the auditory system on the premise that different sensory systems may share the same fundamental computations. One advantage of this approach is that no knowledge of a neuron receptive field is required to determine the summation index characterizing the specific operations. Indeed, the notion of receptive field does not figure in the equation for *SmI* determination. More generally, one can determine whether a neuron performs an AND-like or an OR-like operation with a given pair of stimuli without any explicit knowledge (model) of the neuron receptive field, as long as the stimuli activate the neuron. Thus progress can be made now in understanding computations in high-order sensory neurons whose receptive fields remain difficult to characterize using natural stimuli, as is the case in the visual areas V4 and IT where *SmIs* were first determined, as well as in the present study. Once appropriate statistical tools become available, however, knowledge of which operation (AND-like or OR-like) a neuron performs over specific receptive-field components may be key for their functional interpretation (Kaardal et al. 2013).

Our results indicate that object-recognition models should allow multiple operations per neuron. We suggest that the required flexibility can be achieved by delaying the choice of how a given neuron combines its inputs until those inputs are activated (Fig. 5). In other words, a neuron would combine inputs using one among several possible feature recombination

functions (which biophysically may correspond to different parameter regimes of the same function, such as the *softmax*), depending on the input activation. Stimuli that only partially match a neuron template (its “preferred” stimulus) would individually cause weak depolarization and would combine supralinearly according to the AND-like rule. This is how simpler features would assemble into more complex ones. The “preferred” stimulus of a neuron (i.e., the full complement of simpler features) would cause a stronger input activation, delivering more electric charge and switching to the OR-like rule, such as the MAX function for invariance in the presence of distracters. In this scenario, those neurons that have achieved full selectivity would automatically acquire invariance.

Finally, we point out that our third key result, the demonstration of adaptation changing the summation rules in individual neurons, would be trivial if it were obtained in an isolated neuron in vitro. In an isolated neuron, adaptation (synaptic depression) must eventually make all adapting inputs subthreshold, after which point they will combine following the AND-like rule, regardless of the rule that governed their combination before adaptation. However, the state-of-the-art object-recognition models implicitly assume that when neurons are connected in functional networks, this trivial behavior of isolated neurons is replaced by a unique response rule.

Instability in the operations produced by adaptation poses a challenge to current models of object recognition. On one hand, the main goal of a model such as HMAX is to provide an invariant signature, a feature vector that is invariant as much as possible to transformations of the input and to the presence of distracters. On the other hand, when the operations in the network follow changes in input strength, the model will fail to provide this invariant feature vector. How could this challenge be addressed? Hierarchical models such as HMAX could still work with linear nodes, but they would suffer from clutter. It therefore seems essential to pool activities nonlinearly, e.g., using the MAX operation that is supported by strong experimental evidence in the published studies. We think, therefore, that the key question is not how to make the model work without using the MAX operation but rather how to make it work using the MAX operation (or its biophysically more plausible computational equivalents such as the *softmax*) in such a way that this operation remains stable across different conditions and associated parameter values.

This may be achieved if the network were able to stabilize input activities in individual neurons, for example, using attention, effectively canceling large changes in input strength and undoing the effect of adaptation (and also raising interesting questions about the role of adaptation itself). Some evidence points to this possibility: attention can counteract effects of adaptation on the contrast-response functions in the visual area V4 (Hudson et al. 2009). If attention can stabilize the specific computations, then the HMAX model in its present form would work well within the locus of attention. It will be important to test this hypothesis and to examine how a balance between the specific computations in each neuron is controlled and exploited at different time scales in awake behaving animals.

ACKNOWLEDGMENTS

We thank the members of the Gentner laboratory for comments on the manuscript.

GRANTS

The research was funded by National Institute of Deafness and Other Communications Disorders Grant R01-DC-008358.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: A.K. and T.Q.G. conception and design of research; A.K. performed experiments; A.K. analyzed data; A.K. and T.Q.G. interpreted results of experiments; A.K. prepared figures; A.K. drafted manuscript; A.K. and T.Q.G. edited and revised manuscript; A.K. and T.Q.G. approved final version of manuscript.

REFERENCES

- Ahmadian Y, Rubin DB, Miller KD. Analysis of the stabilized supralinear network. *Neural Comput* 25: 1994–2037, 2013.
- Bernander O, Douglas RJ, Martin KA, Koch C. Synaptic background activity influences spatiotemporal integration in single pyramidal cell. *Proc Natl Acad Sci USA* 88: 11569–11573, 1991.
- Boureau YL, Bach F, LeCun Y, Ponce J. Learning mid-level features for recognition. *2010 IEEE Conf Comput Vis Pattern Recognition (CVPR)*, 2010, p. 2559–2566.
- Butler AB, Reiner A, Karten HJ. Evolution of the amniote pallium and the origins of mammalian neocortex. *Ann NY Acad Sci* 1225: 14–27, 2011.
- Carandini M, Heeger DJ. Normalization as a canonical neural computation. *Nat Rev Neurosci* 13: 51–62, 2011.
- Chew SJ, Mello C, Nottebohm F, Jarvis E, Vicario DS. Decrements in auditory responses to a repeated conspecific song are long-lasting and require two periods of protein synthesis in the songbird forebrain. *Proc Natl Acad Sci USA* 92: 3406–3410, 1995.
- DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition? *Neuron* 73: 415–434, 2012.
- Dugas-Ford J, Rowell JJ, Ragsdale CW. Cell-type homologies and the origins of the neocortex. *Proc Natl Acad Sci USA* 109: 16974–16979, 2012.
- Eens M. Understanding the complex song of the European starling: an integrated approach. *Adv Study Behav* 26: 355–434, 1997.
- Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern* 36: 193–202, 1980.
- Gawne TJ, Martin JM. Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. *J Neurophysiol* 88: 1128–1135, 2002.
- Gentner TQ, Margoliash D. Neuronal populations and single cells representing learned auditory objects. *Nature* 424: 669–674, 2003.
- Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160: 106–154, 1962.
- Hudson AE, Schiff ND, Victor JD, Purpura KP. Attentional modulation of adaptation in V4. *Eur J Neurosci* 30: 151–171, 2009.
- Kaardal J, Fitzgerald JD, Berry MJ 2nd, Sharpee TO. Identifying functional bases for multidimensional neural computations. *Neural Comput* 25: 1870–1890, 2013.
- Kouh M, Poggio T. A canonical neural circuit for cortical nonlinear operations. *Neural Comput* 20: 1427–1451, 2008.
- Lampl I, Ferster D, Poggio T, Riesenhuber M. Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *J Neurophysiol* 92: 2704–2713, 2004.
- Lee H, Grosse R, Ranganath R, Ng A. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. *Proc 26th Annual Int Conf Machine Learning*, 2009, p. 609–616.
- Mechler F, Ringach DL. On the classification of simple and complex cells. *Vision Res* 42: 1017–1033, 2002.
- Miller KD, Troyer TW. Neural noise can explain expansive, power-law nonlinearities in neural response functions. *J Neurophysiol* 87: 653–659, 2002.
- Moldakarimov S, Rollenhagen JE, Olson CR, Chow CC. Competitive dynamics in cortical responses to visual stimuli. *J Neurophysiol* 94: 3388–3396, 2005.
- Movshon JA, Lennie P. Pattern-selective adaptation in visual cortical neurons. *Nature* 278: 850–852, 1979.
- Müller JR, Metha AB, Krauskopf B, Lennie P. Rapid adaptation in visual cortex to the structure of images. *Science* 285: 1405–1408, 1999.
- Poggio T, Bizzi E. Generalization in vision and motor control. *Nature* 431: 768–774, 2004.
- Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. *Nat Neurosci* 2: 1019–1025, 1999.
- Rifkin R, Bouvrie J, Schutte K, Chikkerur S, Kouh M, Ezzat T, Poggio T. Phonetic classification using hierarchical, feed-forward, spectro-temporal patch-based architectures. In *AI Memo 2007-007*. Cambridge, MA: MIT Press, 2007.
- Sato T. Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Exp Brain Res* 77: 23–30, 1989.
- Scholl B, Gao X, Wehr M. Nonoverlapping sets of synapses drive on responses and off responses in auditory cortex. *Neuron* 65: 412–421, 2010.
- Serre T, Oliva A, Poggio T. A feedforward architecture accounts for rapid categorization. *Proc Natl Acad Sci USA* 104: 6424–6429, 2007.
- Theunissen FE, Shaevitz SS. Auditory processing of vocal sounds in birds. *Curr Opin Neurobiol* 16: 400–407, 2006.
- Thompson JV, Gentner TQ. Song recognition learning and stimulus-specific weakening of neural responses in the avian auditory forebrain. *J Neurophysiol* 103: 1785–1797, 2010.
- Walther DB, Koch C. Attention in hierarchical models of object recognition. *Prog Brain Res* 165: 57–75, 2007.
- Yang J, Yu K, Gong Y, Huang T. Linear spatial pyramid matching using sparse coding for image classification. *2010 IEEE Conf Comput Vis Pattern Recognition (CVPR)*, 2009, p. 1794–1801.