OPEN

A Mathematical Theory of Cortex-Receptor Artificial Extension

You-Lu Xing

Many physiology experiments demonstrate that an organism's cortex and receptor system can be artificially extended, giving the organism new types of perceptual capabilities. To examine artificial extension of the cortex-receptor system, I propose a computational model that allows new types of sensory pathways to be added directly to the computational model itself in an online manner. A synapse expandable artificial neuron model that can grow new synapses, forming a bridge between the novel perceptual information and the existing neural network is introduced to absorb the novel sensory pathway. The experimental results show that the computational model can effectively integrate sudden emerged sensory channels and the neural circuits in the computational model can be reused for novel modalities without influencing the original modality.

Many physiological experiments have demonstrated the expandability of an organism's cortex-receptor system. Human L-pigment gene knock-in mice, which express a human long-wavelength-sensitive cone photopigment, acquired a new capacity for chromatic discrimination[1]. Rats that receive information from the infrared environment in their somatosensory cortex can perceive infrared light[2]. Monkeys have learned to discriminate artificial tactile stimuli in an active tactile exploration task, where a brain-machine-brain interface was used to deliver the stimuli to their primary somatosensory cortex[3]. The rat and monkey experiments imply that the sensing capability of an organism can be expanded artificially throughout the organism's lifetime, differing from the natural way, where organisms acquire new experiences through genetic variation and those sensory transducers are fixed at birth[4,5]. To provide a mathematical theory for the cortex-receptor artificial extension, in this study, I build a computational model that can adapt to a novel sensory pathway in an online manner.

According to the experiments[1–3], I divide the extension into two situations: (i) emergence of novel sensory receptors in an existing perceptual channel (Fig. 1a) and (ii) emergence of novel perceptual channels in the whole system (Fig. 1b).

Inspired by the phenomenon that novel experience can induce a formation of new spines in the brain[6,7], a synapse expandable artificial neuron model that can grow new synapses to absorb extended sensory signal during learning is designed. The model is different from the classic artificial neuron model whose synapse structure is fixed[8], which implies a non-extendibility for the novel sensory signals. Meanwhile, to enable the extension of the sensory pathway in the above two situations, a hierarchical and modularized computational structure inspired by the brain structure[9] is designed. The synapse expandable artificial neuron can work on any level of the computational structure, which provides an adequate theoretical explanation to the cortex-receptor artificial extension.

Results

Computational framework. The computational framework for situation (i) (Fig. 2a) and situation (ii) (Fig. 2b) is based on a perception coordination network[10], which is a hierarchical and modularized neural network (see Methods section perception coordination network for algorithmic details). The framework includes the primary sensory areas, the unimodal association areas, and the multimodal association areas. The primary sensory areas contain feature neurons, that respond to elementary features, for example, color, shape, or syllable features. The unimodal association areas contain concept neurons, which combine the elementary features to represent unimodal object concepts; for example, in the visual channel, they combine information from color and shape neurons to represent object images. The multimodal association areas contain association neurons, which connect concept neurons in different perceptual channels, for example, connecting concept neurons in visual and auditory channels to associate the visual and auditory concepts of objects. Figure 2a shows an example of a novel type of receptor added to the retina of an organism, which receives stimuli synchronously with the inherent receptors. Impulses generated by both the novel and inherent receptors are transmitted to the primary sensory areas where they are integrated. Figure 2b shows a novel perceptual channel X being added to the network, which

Department of Computer Science and Technology, Nanjing University, Nanjing, 210023, China. email: youluxing@sina.com
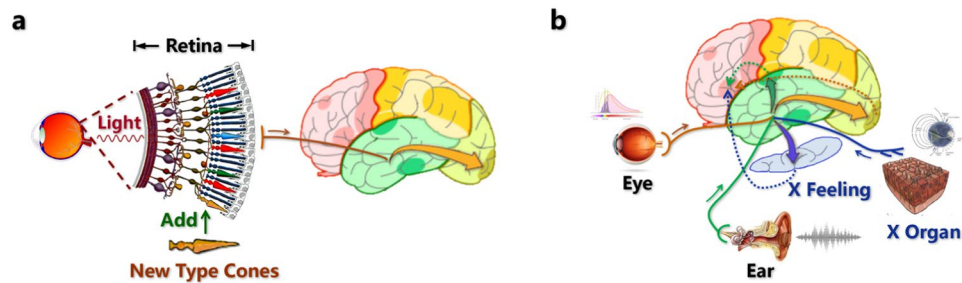
Figure 1. Two situations of extensions of the cortex-receptor system. (**a**) Emergence of novel sensory receptors in an existing perceptual channel. The figure illustrates a new type of cone cells being added to the retina, which is inspired by the experiment in which human L-pigment gene knock-in mice showed enhanced long wavelength sensitivity[1]. In this situation, novel receptors receive environmental stimuli and transmit signals synchronously with the pre-existing receptors, which means that the perceptual channel spectrum is expanded. (**b**) Emergence of novel perceptual channels in the whole system. The figure illustrates a new type of sensory channel X being added to the organism and the channel communicates with other parts of the cortex that are responsible for different sensory types. As a result, the organism gains the ability to sense X. This situation corresponds to the experiment in which rats were able to perceive infrared light through electrodes implanted in their cortex[2] and monkeys were able to feel artificial tactile stimuli via a brain-machine-brain interface[3].

then receives stimuli synchronously with the congenital perceptual channels. Impulses generated by both the novel and congenital channels are transmitted to the primary sensory areas and unimodal association areas, respectively, and are finally combined in the multimodal association area. The neural interactions launched by the association neurons, working as an adhesive, integrates new structure into the original structure to form a single system.

Synapse expandable artificial neuron model. To integrate novel sensory receptors or perceptual channels into the pre-existing neural network, I design a synapse expandable artificial neuron model (Fig. 3a). New synapses (shown in red) can be created to allow pathways for the novel receptor and novel perceptual channel. The expanded synapses on the dendrite (input) side can transmit signals received from new types of sensory receptors. Then, the new type sense has a signal pathway to the existing system; the expanded synapses on the axon (output) side can transmit signals from the existing system to the new type of sense channel. Then, the system has a signal pathway to the new type sense channel. When new synapses are created, the activation function of the synapse expandable artificial neuron model evolves as follows,

$$z_0 = f(x_0, w_0, \theta) \rightarrow [z_0, z_1] = f([x_0, x_1], [w_0, w_1], \theta) \tag{1}$$

where x_0 and z_0 are the input and output in the original synapses, respectively, w_0 is the weights of the original input synapses, x_1 and z_1 are the input and output in the new synapses, respectively, w_1 is the weights of the new input synapses, θ is the threshold of the artificial neuron. The implanted electrodes in the IR rat experiment[2] can be regarded as another form of synapses. Then, a special case of the activation function can be used to explain the IR rat system, where x_1 equals the input infrared light, w_1 transforms the input infrared signal x_1 to the microstimulation which is sent to the neuron, and z_1 is empty here because the IR rat system reuses the original output pathway of the neuron with high probability.

Novel receptor added to existing perceptual channel. Assume that a perceptual channel initially has n sensory receptors, that receive an n-dimensional vector $x_{\mathrm{I}} = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$. After m new sensory receptors, which receives an m-dimensional vector $x_{\mathrm{II}} = (x_{n+1}, x_{n+2}, \ldots, x_{n+m}) \in \mathbf{R}^m$, are added to the perceptual channel, the dimension of the internal mapping space of the channel grows from $\mathbf{R}^n$ to $\mathbf{R}^{n+m}$, and the learned concepts in space $\mathbf{R}^n$ are mapped to a higher-dimensional feature space $\mathbf{R}^{n+m}$, as illustrated in Fig. 3b. For example, assuming that novel receptors that receive very short wavelength light are added to a trichromatic color vision organism, the organism's color perception will evolve from the original 3-dimensional system (R, G, B) to a 4-dimensional system (R, G, B, X). Feature neurons are used to realize the mapping from space $\mathbf{R}^n$ to space $\mathbf{R}^{n+m}$, and new synapses are created between the new sensory neurons and the feature neurons for the x_{II} pathway, as illustrated in Fig. 3c. The main idea of the algorithm is as follows. When an input sample $(x_{\mathrm{I}}, x_{\mathrm{II}})$ arrives, feature extraction is conducted using x_{I} and x_{II}. I assume that the feature vectors y_{I} and y_{II} are obtained through x_{I} and x_{II}, respectively. Then, y_{I} tries to activate a feature neuron in its corresponding primary sensory areas. When some feature neuron is activated but no synapse exists between the feature neuron and the novel sensory neurons, new synapses are created between them and the weights of the synapses are initialized to y_{II}, which means that $w = (w_{\mathrm{I}}, w_{\mathrm{II}})$, where w_{I} represents the original weights of the feature neuron, w_{II} represents the original weights of the feature neuron which is initialized to y_{II}. According to Eq. (1), the activation function of the feature neuron evolves as follows,
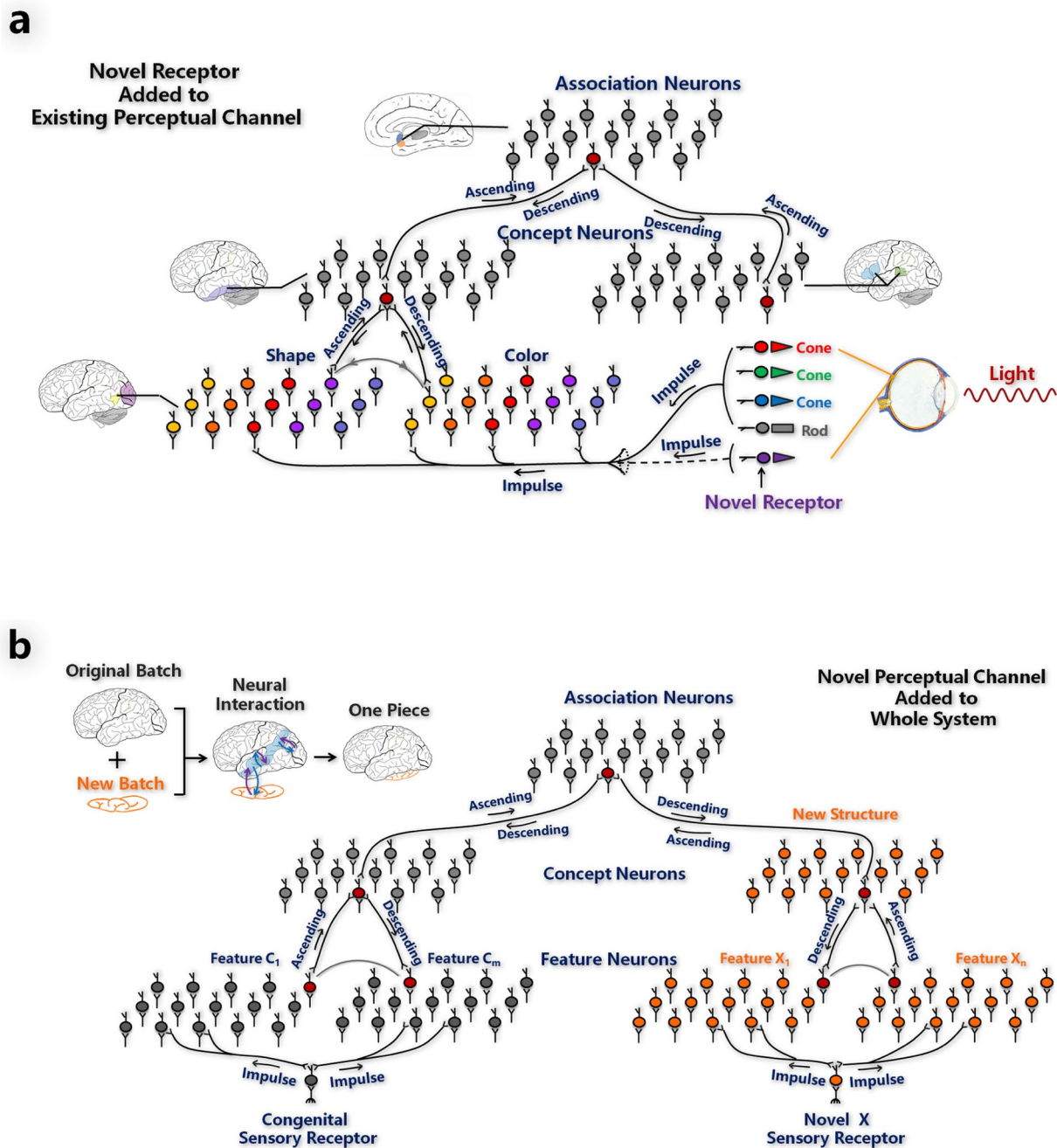
Figure 2. Computational framework of extensions to the cortex-receptor system. (**a**) A novel sensory receptor is added to an existing perceptual channel. (**b**) A novel perceptual channel is added to the entire system.

$$z = f(x_I, w_I, \theta) \rightarrow z = f([x_I, x_{II}], [w_I, y_{II}], \theta) \tag{2}$$

where x_I and w_I are the input and weights in the original synapses, respectively. x_{II} and y_{II} are the input and initialized weights in the new synapses, respectively, θ is the threshold of the artificial neuron. z is the output of the neuron. Finally, a deeper feature reflecting the external world is formed in the network, and the feature neuron can perform calculations in the high dimension feature space (see Methods section dimensionality increasing process of the feature neurons for algorithmic details).

Novel perceptual channel added to whole system. Assume that the system has an inherent perceptual channel **L**, that receives an n-dimensional vector $x^I = (x_1^I, x_2^I, \ldots, x_n^I) \in \mathbf{R}^n$. After a new perceptual channel **X**, which receives an m-dimensional vector $x^{II} = (x_1^{II}, x_2^{II}, \ldots, x_m^{II}) \in \mathbf{R}^m$, is added to the system, a new internal mapping space emerges in the system. The system should associate concepts in the new space with the concepts
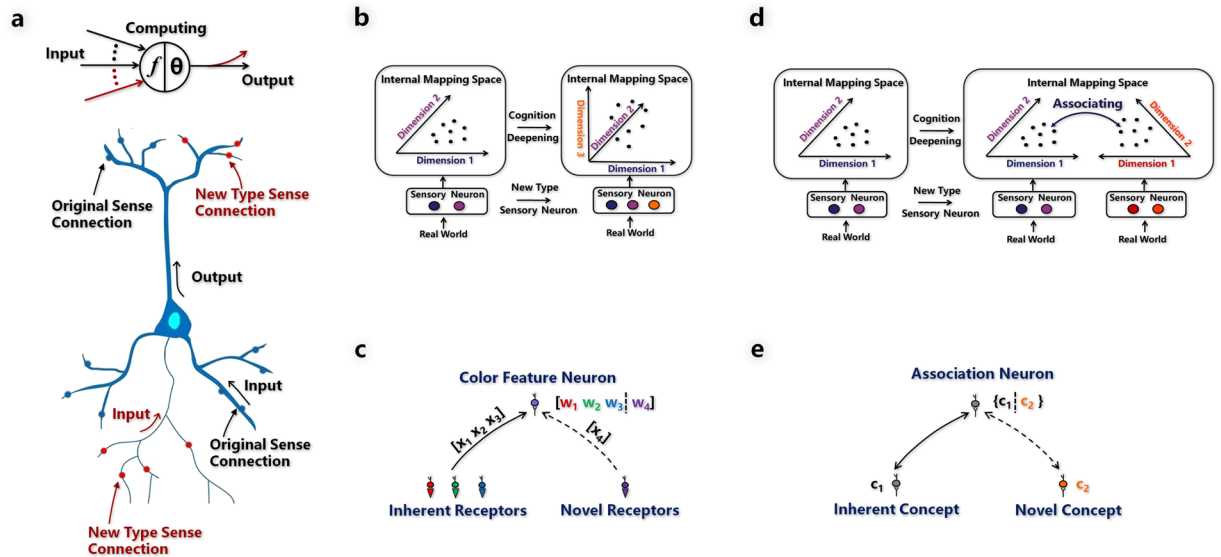
Figure 3. Computational solutions. (**a**) Synapse-expandable artificial neuron model. New synapses (shown in red) can be created to allow a pathway for novel perceptual receptors or channels. (**b**) Mathematical modelling of situation (i). After a new type of sensory receptor is added to an existing perceptual channel, the organism has a new way to perceive the environment. Consequently, the internal mapping space of the organism has a corresponding new dimension. Thus, the points that represent the learned concepts in the original low dimension feature space are mapped to a higher-dimensional feature space. (**c**) The dimensionality increasing process of the feature neuron. Feature neuron grows new synapses to absorb novel features. (**d**) Mathematical modelling of situation (ii). After a new perceptual channel is added to the entire system, the internal mapping space of the organism has a corresponding new space, and new concepts will be generated in this space. Finally, the concepts in the original space and the new space become associated with each other. (**e**) The response modal expansion process of the association neuron. Association neuron grows new synapses to absorb novel modality concept.

in the inherent space, as illustrated in Fig. 3d. Association neurons are used to associate the concepts in the two spaces by creating synapses between the concept neurons in channel **X** and the association neurons that already connect to some concept neurons in the inherent perceptual channel, as illustrated in Fig. 3e. This process is based on the Hebbian theory that neurons that fire together wire together. The main idea of the algorithm are as follows. When an input sample pair (x^I, x^{II}) arrives, x^I tries to activate some association neuron through the inherent perceptual channel. If the activated association neuron is not connected with any concept neurons in channel **X**, a new synapse is created between the association neuron and the concept neuron, which is activated by x^{II} in channel **X**. According to Eq. (1), the activation function of the association neuron evolves as follows,

$$z^I = f(x^I, w^I, \theta) \rightarrow [z^I, z^{II}] = f([\, x^I, x^{II}], [w^I, w^{II}], \theta) \tag{3}$$

where z^{II} is the new output synapse to channel **X** which provides a signal pathway from channel **L** to channel **X**. Meanwhile, z^I can provide a reverse signal pathway from channel **X** to channel **L** now. x^I and x^{II} are the input from the concept neurons in **L** and **X**, respectively. w^I and w^{II} are the weights of the synapses from the concept neurons in **L** and **X** to the association neuron, respectively. These synapses are two independent groups. Subsequently, the system can respond to the new perceptual channel, which means that the system can respond to a new world by feeling **X** (see Materials and Methods section response modal expansion of the association neurons for algorithmic details).

Experiments. An experiment (Supplementary Fig. 1 shows the objects used) is designed to validate our computational model, which is called the CRAET (cortex-receptor artificial extension theory) network. Figure 4a shows the experimental setting, I first give the computational model a visual channel. To simulate the visual system evolving from achromatopsia to dichromatopsia and then to trichromatopsia, I initially give the visual channel a brightness receptor (corresponding to the Rods) that receives grayscale object images, and let the network learn object shapes. After a period of learning with all objects, I provide the visual channel with new green and blue color receptors (corresponding to the M-Cones and L-Cones) so that it additionally receives the green and blue light, respectively. Now, the network can receive a grey + green + blue image of the object. Then, I let the network perceive all objects again to learn color information of each object. After that, I add a red color receptor (corresponding to the S-Cones), allowing it to receive the red light. The network can now receive a Grey + RGB image of the object. Similarly, I let the network perceive all objects another round to update its color feature neurons. Next, I introduce an auditory channel that enables audible sound input to the network. At this time, visual and auditory input can be received simultaneously. I let the network learn the Chinese name of each object by receiving pairs consisting of images accompanied by the Chinese names of each object. After all objects are
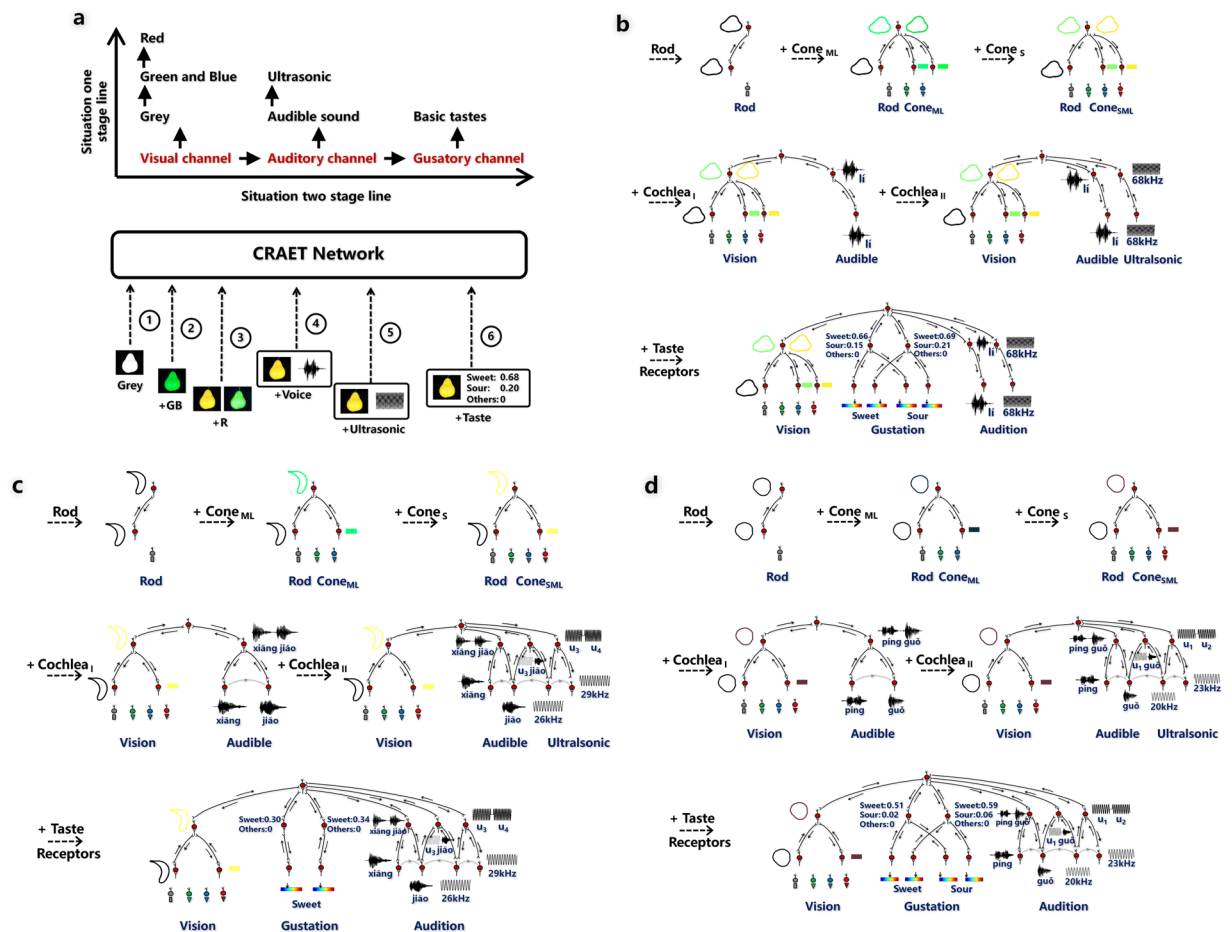
Figure 4. Experiment. (**a**) Experimental settings. (**b**–**d**) Neurons in the learned network that associate with concept pear, banana and apple. The icons next to the neuron are the objects to which the neuron maximally responds. Thresholds for neurons to be activated are set as follows: in the visual channel, thresholds of the shape and color neurons are 1/4 times the L2 norm of their weights; in the auditory channel, thresholds of the audible and ultrasonic syllable neurons are 200 and 9, respectively; in the gustatory channel, thresholds of the basic flavor neurons are 0.015.

learned, I add an ultrasonic receptor to the auditory channel and use ultrasonic sounds to name all objects again. Finally, a gustatory channel is added to the network. Pairs of image and taste samples are provided to the network simultaneously at this stage to let the network learn the taste of each object.

Figure 4b shows the neurons in the learned network that associate with concept pear. The top line shows the change of the network structure of the concept pear as the network acquires new visual receptors. Initially, the network receives only a greyscale image of the object and learns the shape of the object. After the green and blue color receptors are added, the network can receive certain object colors, and color feature neurons are created in the network. The visual concept neuron is then able to respond to color information. In this case, two colors are associated with the pear shape, one appears a little brighter and the other is somewhat darker. However, they are not easy to distinguish in this color space. After the red color receptor is added, the color feature neurons are mapped to a higher-dimensional space: the RGB color space. One color feature neuron becomes responsive to yellow, and another becomes responsive to green. Thus, the bright green pear and dark green pear become more easily distinguishable in the RGB color space: one is yellow, and the other is green. This gives trichromatic organisms an advantage compared to dichromatic organisms. The middle line shows the change in the network structure of the concept pear after adding an auditory channel. On the left, the auditory channel which provides audible frequencies is added. On the right, the auditory receptor which provides ultrasonic frequencies is added. The association neuron connects the images and names (audible and artificial ultrasonic words) of the object correctly. The bottom line shows the change in the network structure of the concept pear after a gustatory channel is added. The artificial taste data of the pear contain sweet and sour flavors. The network learns these two flavors (sweet 0.67, sour 0.22) and (sweet 0.63, sour 0.14). The response modal of the associate neuron is expanded with the concept of taste. Figure 4c,d show two similar results. The experiment demonstrates the CRAET network can effectively integrate newly introduced sensory receptors and channels in an online manner (Supplementary Fig. 2 shows more learning results).

Inspired by the phenomenon found in the IR rat experiment in which a new receptive field was embedded in S1 neurons without hijacking their original receptive field[2], I designed a modality embedding experiment.
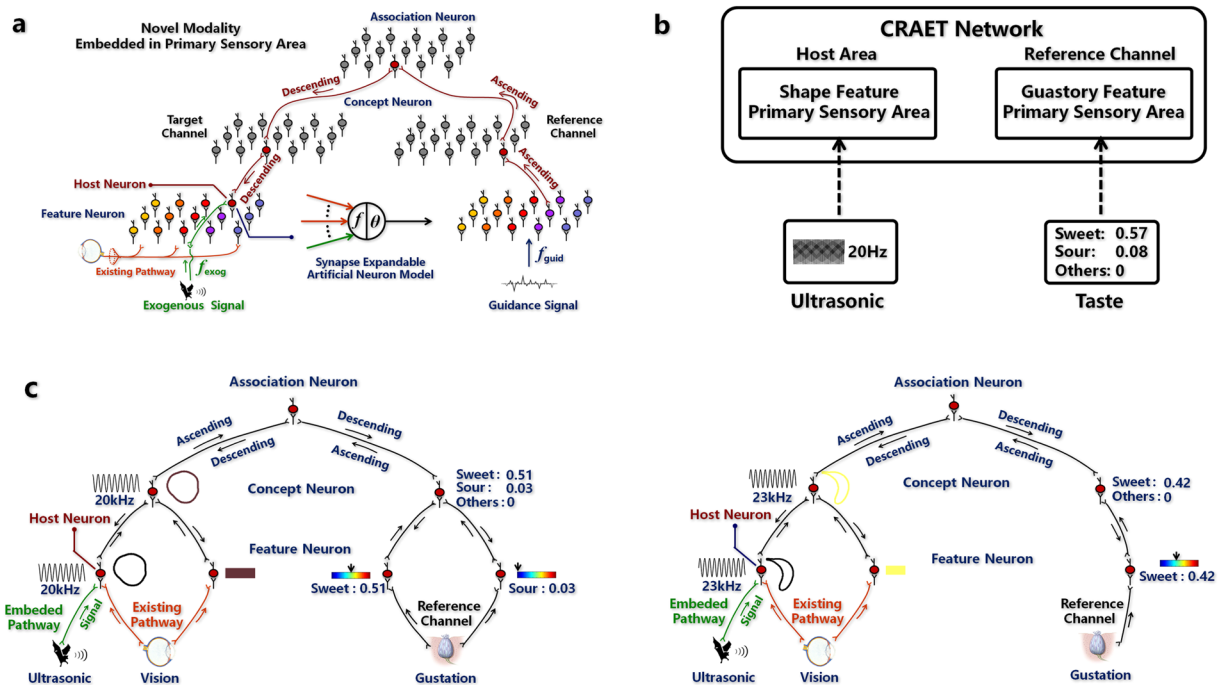
Figure 5. Modality embedding experiment. (**a**) Computational framework of the modality embedding. (**b**) Experimental settings. I set the gustatory channel as the reference channel and the visual channel as the target channel. The ultrasonic modality is embedded into the shape feature primary sensory area. (**c**) Examples of the embedding result. The shape feature neurons are embedded with an ultrasonic modality without being deprived of their original receptive field, which means the ultrasonic modality reuses the vision-gustation circuit effectively. Thresholds for neurons to be activated are set as follows: thresholds of the ultrasonic syllable neurons are 200 and 9, respectively; thresholds of the basic flavor neurons are 0.015.

Figure 5a shows the computational modelling. The network receives a pair of samples that includes an exogenous signal x_e to be embedded and a guidance signal x_g, which leads x_e to the host neuron. The guidance signal first transmits in ascending fashion in the reference channel to activate some association neuron. Then, the activated association neuron transmits signals in descending fashion to the host area, where the new modality is embedded, to activate some host neuron used to absorb the novel modality. If the host neuron does not have synapses connected to the novel modal signal pathway, new synapses will be created, and their weights will be initialized as $w = f_{exog}(x_e)$ to store the pattern of the exogenous signal x_e, where $f_{exog}(\cdot)$ is the feature extraction function of the novel modality. If the host neuron already has synapses connected to the novel modal signal pathway, the weights of the synapses are updated using the competitive learning rule, $w = w + \delta(f_{exog}(x_e) - w)$, where δ is the learning rate (see Methods section modality embedding for algorithmic details).

Figure 5b shows the experimental setting, where an ultrasonic modality is embedded into the neurons in the shape feature primary sensory area using a taste sample as a guidance signal. Figure 5c shows two examples of embedding results. In the results, the shape feature neurons are embedded with an ultrasonic modality without being deprived of their original receptive field. The ultrasonic modality reuses the vision-gustation circuit effectively (Supplementary Fig. 3 shows more embedding results).

Discussion

In this study, a mathematical theory for the cortex-receptor artificial extension is studied. I design a synapse expandable artificial neuron mode to absorb novel information flow. A hierarchical and modularized computational structure is proposed to enable novel information flow be integrated in different concept levels. Meanwhile, I design different computational models for different types of neurons found in many physiological experiments including the feature neurons[11], concept neurons[12] and association neurons[13] (see Methods section perception coordination network for different neuron models).

Currently, the computational mechanism is built only at the neuron level. The computational mechanism of the cortex and receptor extension at the large-scale neural circuit level should be studied. This raises an interesting question: can we design new modality feelings with electronic circuits and then integrate them into organisms? Regarding the question of whether organisms experience novel sensory inputs as a new distal sensory modality[2], I feel the computational and behavioral dimensions of this question are extremely interesting avenues for future research, with particular potential to be studied in human subjects. Organisms' cortex and receptor systems tend to become more complex over time; As more information arrives at the brains of the organisms, and in response, there is increasing number of information processing modes of the brain which seem to become increasingly complicated. The organisms are able to understand the real world more deeply. The brain-machine interface
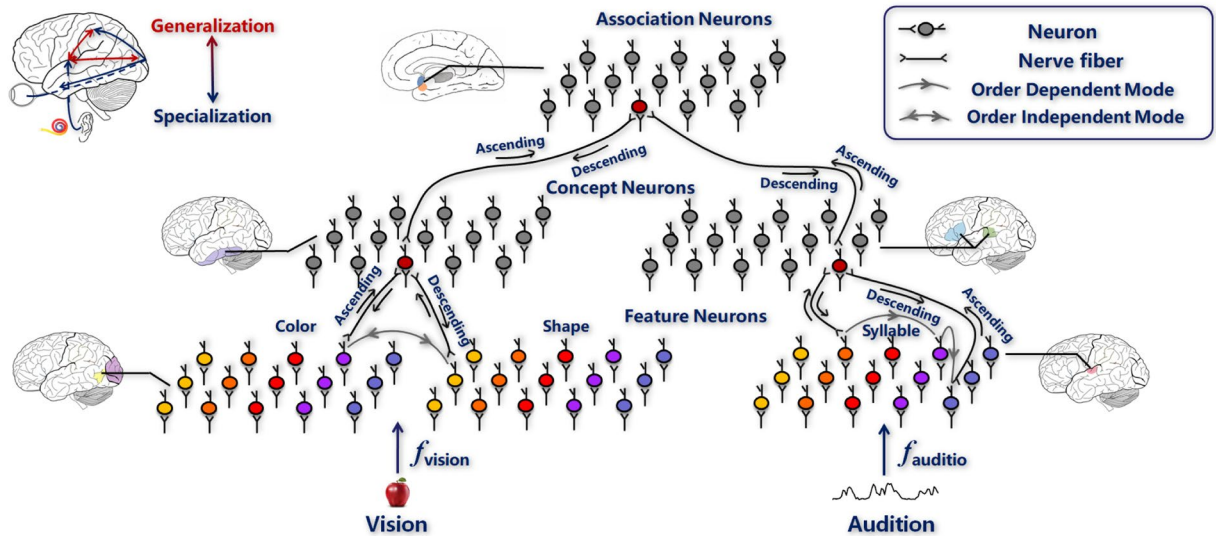
Figure 6. Computational model of the perception coordination network. The hierarchical and modularized structure of the network is inspired by the brain's structure. Note that the figure depicts only vision and audition, other sensations could also be included in the structure.

experiments[2,3] showed the potential to expand a species' normal perceptual range. Building up a corresponding computational theory will deepen our understanding about the extensibility of the central nervous system and the organism-machine hybrid intelligence. It seems clear from experimental results that we have not come close to exhausting the potential for incorporating novel sources of information into cortical processing modes[14], this also raises a very interesting question: can we know or prove that we can already perceive all the dimensions of the world?

Methods

Perception coordination network.   The perception coordination network is an online learning framework that aims to incrementally learn and bind concepts. As shown in Fig. 6, the network includes the primary sensory areas, the unimodal association areas and the multimodal association areas.

The primary sensory areas include feature neurons that respond to particular features, for example, shape features, color features, or syllable features. Feature neurons that respond to the same feature type are located in the same area α, and I use the set N^{F_α} to represent them. As mentioned above, α can be the shape area b, the color area c, or the syllable area s. $N_i^{F_\alpha}$ denotes feature neuron i in area α and $N_i^{F_\alpha} = \{w_i, \sigma_i\}$, where w_i represents the weights and σ_i is the cumulative number of times the neuron has been activated. The activation function of $N_i^{F_\alpha}$ is $f(x) = \begin{cases} 1, & \|x - w_i\| \leq \theta \\ 0, & \|x - w_i\| > \theta \end{cases}$, where θ is a parameter that controls the response range of the feature neuron.

The unimodal association areas include concept neurons, which connect the feature neurons to represent a unimodal concept, for example, to form visual concepts by connecting shape and color feature neurons, or to form words (auditory concepts) by connecting syllable feature neurons. Concept neurons with the same modality are located in the same area, β, and I use set N^{C_β} to represent them. β can be a visual association area v, an auditory association area a, or some other sensory association area. $N_i^{C_\beta}$ denotes the concept neuron i in area β. The activation mode of the concept neuron is one of two types: order-independent activation mode and order-dependent activation mode. For example, a visual concept neuron has an order-independent activation mode because different sequences in the activation of shape and color feature neurons to which a visual concept neuron connects do not affect the activation of the visual concept neuron. In contrast, auditory concept neurons (for example, words) have an order-dependent activation mode because a word is composed of a sequence of syllables. Correspondingly, two types of circuit structure between a concept neuron and a feature neuron are defined as follows

<div align="center">
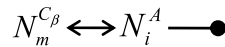
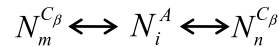Type I: order-independent circuit   Type II: order-dependent circuit

</div>
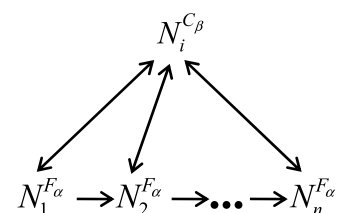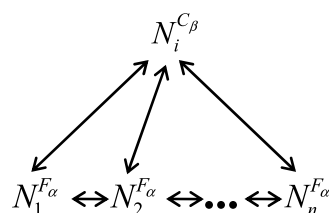
where the arrows represent connections between neurons. Then, the activation function of $N_i^{C_\beta}$ is

$$f(N_1^{F_\alpha}, N_2^{F_\alpha}, …, N_n^{F_\alpha}) = \begin{cases} 1, & N_i^{F_\alpha} = 1, 1 \leq i \leq n, \\ 0, & \text{else} \end{cases}$$ where feature neurons $N_1^{F_\alpha}, N_2^{F_\alpha}, …, N_n^{F_\alpha}$ connect to $N_i^{C_\beta}$, and

$N_i^{F_\alpha} = 1$ means $N_i^{F_\alpha}$ is activated. For the order-dependent activation mode, $N_1^{F_\alpha}, N_2^{F_\alpha}, …, N_n^{F_\alpha}$ must be activated in sequence. The connection between concept neuron $N_i^{C_\beta}$ and feature neuron $N_j^{F_\alpha}$ is denoted as $c_{(i,j)} = \{N_i^{C_\beta}, N_j^{F_\alpha}, \rho_{(i,j)}\}$, where $\rho_{(i,j)}$ is the cumulative number of times the connection is activated.

The multimodal association areas include association neurons, which connect the concept neurons in different unimodal association areas, for example, they connect an image of an object represented by a visual concept neuron with its name, which is represented by an auditory concept neuron. The association neurons are stored in set N^A, and N_i^A is used to denote the association neuron for i. There are four types of circuit structures between an association neuron and a concept neuron,
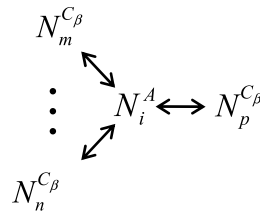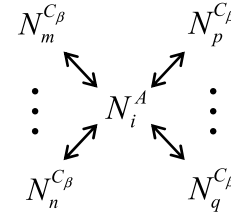
Type O: 1-to-? circuit

$$N_m^{C_\beta} \longleftrightarrow N_i^A \longrightarrow\bullet$$

Type I: 1-to-1 circuit

$$N_m^{C_\beta} \longleftrightarrow N_i^A \longleftrightarrow N_n^{C_\beta}$$

Type II: 1-to-n circuit

$$\begin{matrix} N_m^{C_\beta} \\ \vdots \\ N_n^{C_\beta} \end{matrix} \searrow N_i^A \longleftrightarrow N_p^{C_\beta}$$

Type III: n-to-m circuit

$$\begin{matrix} N_m^{C_\beta} && N_p^{C_\beta} \\ \vdots & N_i^A & \vdots \\ N_n^{C_\beta} && N_q^{C_\beta} \end{matrix}$$

Each concept neuron in the circuit can activate the association neuron. Thus, the association neuron has a multimodality activation mode, and the activation function of N_i^A is $f(N_1^{C_\beta}, N_2^{C_\beta}, …, N_n^{C_\beta}) = \begin{cases} 1, & N_i^{C_\beta} = 1, i \in \{1, 2, …, n\}, \\ 0, & \text{else} \end{cases}$, where the concept neurons $N_1^{C_\beta}, N_2^{C_\beta}, …, N_n^{C_\beta}$ connect to N_i^A, and $N_i^{C_\beta} = 1$ means $N_i^{C_\beta}$ is activated. The connection between concept neuron $N_m^{C_\beta}$ and another concept neuron $N_n^{C_\beta}$ through association neuron N_i^A is denoted as $c_{(m,i,n)} = \{N_m^{C_\beta}, N_i^A, N_n^{C_\beta}, \rho_{(m,i,n)}\}$, where $\rho_{(m,i,n)}$ is the cumulative number of times the connection is activated.

Without loss of generality, in the following I use a pair of visual (an object image) and auditory (the name of an object) inputs to describe the method in detail.

When an image and voice (name) pair arrives, feature extraction is conducted first. Currently, the normalized Fourier descriptors of the object's boundary d and the color histogram of the object h are used for visual features. The Mel-Frequency Cepstral Coefficients (MFCCs) of the syllables contained in the voice m are used for the auditory features, where short-time energy and short-time zero crossing are used to extract the syllables from a voice wave. Then, a competitive learning process is executed among the feature neurons.

In the visual channel, a winner neuron $N_b^{F_b}$ in the shape feature area and a winner neuron $N_c^{F_c}$ in the color feature area are found using $N_b^{F_b} = \underset{N_i^{F_b} \in N^{F_b}}{\arg\min} \|d - w_i\|_2$ and $N_c^{F_c} = \underset{N_i^{F_c} \in N^{F_c}}{\arg\min} \|h - w_i\|_2$, where N^{F_b} and N^{F_c} are the sets of neurons in the shape area and color area, respectively. In the shape area, if $\left\|d - w_b\right\|_2 \leq \frac{1}{4}\left\|w_b\right\|_2$, $N_b^{F_b}$ is activated and updated using the winner-take-all principle: $\sigma_b = \sigma_b + 1$, $w_b = w_b + (d - w_b)/\sigma_b$; If $\left\|d - w_b\right\|_2 > \frac{1}{4}\left\|w_b\right\|_2$, the network recognizes d as a new feature and creates a new feature neuron to record it: $N_{new}^{F_b} = \{h, 1\}$. Then, $N_{new}^{F_b}$ is activated. A similar process is executed in the color area. Finally, the activated feature neurons in the shape area (assumed to be $N_{f_b}^{F_b}$) and color area (assumed to be $N_{f_c}^{F_c}$) transmit their activation signals to the visual unimodal association area.

In the auditory channel, dynamic time warping (DTW) is used to find the winner neuron for each syllable contained in the voice input: $N_{f_i}^{F_s} = \underset{N_j^{F_s} \in N^{F_s}}{\arg\min} \mathbf{dtw}(m_i, w_j)$, where $1 \leq i \leq k$, w_j represents the MFCCs of syllable feature neuron j, m_i is the MFCCs of the i-th syllable in the voice input, and there are k syllables in total in the voice input. If $\mathbf{dtw}(m_i, w_{f_i}) < 200$, $N_{f_i}^{F_s}$ is activated; otherwise, the network recognizes m_i as a new feature and creates a new feature neuron to record it: $N_{new}^{F_s} = \{m_i, 1\}$. Then, $N_{new}^{F_s}$ is activated. Finally, the activated feature neurons in the syllable area (assumed to be $N_{f_i}^{F_s}, 1 \leq i \leq k$) transmit their activation signals to the auditory unimodal association area.

Meanwhile, self-organization is conducted among feature neurons using the competitive Hebbian rule (Supplementary Figs. 4–6 show examples of self-organization among feature neurons).

When the visual unimodal association area and auditory unimodal association area receive the activation signals $(N_{f_b}^{F_b} = 1, N_{f_b}^{F_b} = 1)$ and $(N_{f_1}^{F_s} = 1, N_{f_2}^{F_s} = 1, ..., N_{f_k}^{F_s} = 1)$, a unimodal-concept incremental learning process is executed.

In the visual channel, if the activation signal $(N_{f_b}^{F_b} = 1, N_{f_b}^{F_b} = 1)$ satisfies the activation function of a visual concept neuron $N_{f_v}^{C_v}$, which means that the activation function of $N_{f_v}^{C_v}$ is $f(N_{f_a}^{F_a}, N_{f_c}^{F_c}) = \begin{cases} 1, & N_{f_a}^{F_a} = 1, N_{f_c}^{F_c} = 1, \\ 0, & \text{else} \end{cases}$, then $N_{f_v}^{C_v}$ is activated and the cumulative number of activations of the connection between $N_{f_v}^{C_v}$ and $N_{f_b}^{F_b}$, $N_{f_v}^{C_v}$ and $N_{f_c}^{F_c}$ is increased by one. If the activation signal $(N_{f_b}^{F_b} = 1, N_{f_b}^{F_b} = 1)$ does not satisfy the activation function of any visual concept neuron, the network assumes the current input is a new concept and creates a new visual concept neuron $N_{new}^{C_v}$ to record it. Therefore, an order-independent circuit with $N_{new}^{C_v}$, $N_{f_b}^{F_b}$, and $N_{f_b}^{F_b}$ is introduced into the network. Then, $N_{new}^{C_v}$ is activate. A similar process is executed in the auditory channel. Finally, the activated concept neurons in the visual channel (assumed to be $N_{f_v}^{C_v}$) and auditory channel (assumed to be $N_{f_a}^{C_a}$) transmit their activation signals to the multimodal association area.

When the multimodal association area receives signals $N_{f_v}^{C_v} = 1$ and $N_{f_a}^{C_a} = 1$, the network first checks whether $N_{f_v}^{C_v} = 1$ and $N_{f_a}^{C_a} = 1$ can activate some association neurons. I assume that the association neurons in set N_v^A can be activated by $N_{f_v}^{C_v} = 1$ and that association neurons in set N_a^A can be activated by $N_{f_a}^{C_a} = 1$. Next, the network activates the auditory concept neurons that connect to the association neurons in N_v^A and the visual concept neurons that connect to the association neurons in N_a^A. I use sets $N_v^{C_a}$ and $N_a^{C_v}$ to represent these auditory concept neurons and visual concept neurons, respectively. Obviously, there are four possible combinations between set N_v^A and set N_a^A.

When $N_v^A = \varnothing$ and $N_a^A \neq \varnothing$, the view of the current input of the object is new to the network, but the voice has been encountered previously and has been used to name some other views. However, the current object should look like the views symbolized by the visual concept neurons in set $N_a^{C_v}$ according to the current voice input. Now, the network asks the user a question: "I find the current input name $N_{f_a}^{C_a}$ has been used to call other views in $N_a^{C_v}$; can it also represent the current input view?" If the answer from the user is positive, a connection is created between $N_{f_v}^{C_v}$ and $N_{f_a}^{C_a}$ through each association neuron in set N_a^A; otherwise, the view $N_{f_v}^{C_v}$ is stored as a Type O circuit using a new association neuron.

When $N_v^A \neq \varnothing$ and $N_a^A = \varnothing$, the current voice input is new to the network, but the view of the current object has been encountered previously. However, the object should be named as symbolized by the auditory concept neurons in set $N_v^{C_a}$. Therefore, the network asks the user a question: "The object was called $N_v^{C_a}$ previously. Is it also called $N_{f_a}^{C_a}$?" If the answer is positive, a connection is created between $N_{f_v}^{C_v}$ and $N_{f_a}^{C_a}$ through each association neuron in set N_v^A; otherwise, the current input name $N_{f_a}^{C_a}$ is rejected by the network.

When $N_v^A \neq \varnothing$ and $N_a^A \neq \varnothing$, both the view and the voice input have been encountered previously. Their coherence should be checked. When $N_v^A \cap N_a^A \neq \varnothing$, then $N_{f_v}^{C_v}$ and $N_{f_a}^{C_a}$ activate some association neurons in common. The current image/voice pair is consistent with some previous pairs. The cumulative number of activations of the connections between $N_{f_v}^{C_v}$ and $N_{f_a}^{C_a}$ through the activated association neurons in common are increased by one to strengthen the association. When $N_v^A \cap N_a^A = \varnothing$, it means the current combination of $N_{f_v}^{C_v}$ and $N_{f_a}^{C_a}$ is inconsistent with previous pairs. Therefore, the network asks the user a question: "The current input pair is inconsistent with previous pairs, is this pair an expected combination?" If the answer is positive, the connections between $N_{f_v}^{C_v}$ and $N_{f_a}^{C_a}$ through the activated association neurons in common are created; otherwise, the current input pair is rejected by the network.

When $N_v^A = \varnothing$ and $N_a^A = \varnothing$, the current combination of $N_{f_v}^{C_v}$ and $N_{f_a}^{C_a}$ has not been encountered previously. A new association neuron is created to associate $N_{f_v}^{C_v}$ and $N_{f_a}^{C_a}$ with a 1-to-1 circuit.

At this point, the entire learning process for the visual and auditory input pair is complete, and the network continues with the next input pair.

Dimensionality increasing process of the feature neurons.

When novel receptors are added to an existing perceptual channel, feature neurons handle the signals from the novel receptors. As illustrated in Fig. 3c, new synapses will grow to connect the new sensory receptors, which means that the spectrum range of the feature neuron is broadened by the new receptors. I call it a dimensionality increasing process.

I denote $x_I = (x_1, x_2, ..., x_n) \in \mathbf{R}^n$ as the data received from the originally existing n sensory neurons, $x_{II} = (x_{n+1}, x_{n+2}, ..., x_{n+m}) \in \mathbf{R}^m$ as the data received from the m new sensory neurons, $S_I = \mathbf{R}^n$ as the original sensory space, and $S_{II} = \mathbf{R}^{n+m}$ as the new sensory receptor space. During the dimensionality increasing process, two types of feature neurons can exist: type one consists of neurons that already completed the dimensionality increasing process, and type two is the neurons that have not completed the dimensionality increasing process. I use the sets $N_I^{F_\alpha}$ and $N_{II}^{F_\alpha}$ to represent the type one and type two feature neurons, respectively, and use $N^{F_\alpha} = N_I^{F_\alpha} \cup N_{II}^{F_\alpha}$ to store the feature neurons in area α.

When an input sample $x = (x_I, x_{II})$ arrives, feature extraction is conducted using x_I and x_{II} first, $y_I = g(x_I)$, $y_{II} = g(x_{II})$, where $g(\cdot)$ is the feature extraction function, and $y = (y_I, y_{II})$ is the total number of features in the current input sample x. Next, among the neurons in set N^{F_α}, a winner neuron is found through com-

petition in the original sensory space S_I, such that $N_a^{F_\alpha} = \underset{N_j^{F_\alpha} \in N^{F_\alpha}}{\arg\min} D(y_I, \hat{w}_j)$, where $\hat{w}_j = (w_1, w_2, \ldots, w_n)$ represents the weight vector of $N_j^{F_\alpha}$ in space S_I, and function $D(\cdot, \cdot)$ measures the similarity between the two variables. Among the neurons in set $N_{II}^{F_\alpha}$, a winner neuron is found in the evolved sensory space S_{II}, such that $N_b^{F_\alpha} = \underset{N_j^{F_\alpha} \in N_{II}^{F_\alpha}}{\arg\min} D(y, w_j)$, where $\hat{w}_j = (w_1, w_2, \ldots, w_{n+m})$ represents the weight vector of $N_j^{F_\alpha}$ in space S_{II}. Then, I check whether y_I and y can activate the feature neurons $N_a^{F_\alpha}$ and $N_b^{F_\alpha}$, respectively, which requires calculating equations $f_{N_a^{F_\alpha}}(y_I)$ and $f_{N_b^{F_\alpha}}(y)$, where $f_{N_a^{F_\alpha}}(\cdot)$ and $f_{N_b^{F_\alpha}}(\cdot)$ are the activation functions of $N_a^{F_\alpha}$ and $N_b^{F_\alpha}$, respectively.

If $f_{N_a^{F_\alpha}}(y_I) = 1$ and $f_{N_b^{F_\alpha}}(y) = 0$, then y_I activates $N_a^{F_\alpha}$ but y does not activate $N_b^{F_\alpha}$. This implies that a familiar feature y_I is accompanied by a novel feature in the unity form y that is not recognized by the network. As a result, the dimensionality increasing process should be applied to $N_a^{F_\alpha}$:

$$\sigma_a = \sigma_a + 1, \quad w_a = w_a + \frac{1}{\sigma_a}(y_I - w_a)$$

$$w_a = (w_a, y_{II}) \tag{4}$$

where σ_a represents the cumulative number of times that $N_a^{F_\alpha}$ has been activated and w_a is the weight vector of $N_a^{F_\alpha}$. It is worth noting that Eq. (4) increases the number of synapses for a neuron. The weights of the new added synapses are set to y_{II} which causes $N_a^{F_\alpha}$ to connect to the novel feature y_{II}. Consequently, the response range of $N_a^{F_\alpha}$ is broadened. Equation (4) denotes the dimensionality increasing process for a feature neuron.

When $f_{N_a^{F_\alpha}}(y_I) = 1$ and $f_{N_b^{F_\alpha}}(y) = 1$, then y_I activates $N_a^{F_\alpha}$ and y activates $N_b^{F_\alpha}$. Therefore, the input feature y is recognized by the network. In this situation, $N_b^{F_\alpha}$ inhibits $N_a^{F_\alpha}$, and $N_b^{F_\alpha}$ is updated as follows:

$$\sigma_b = \sigma_b + 1, \quad w_b = w_b + \frac{1}{\sigma_b}(y - w_b)$$

When $f_{N_a^{F_\alpha}}(y_I) = 0$ and $f_{N_b^{F_\alpha}}(y) = 1$, then y_I does not activate $N_a^{F_\alpha}$ but y activates $N_b^{F_\alpha}$. The input feature y is recognized by the network, and $N_b^{F_\alpha}$ is updated using the above equations.

When $f_{N_a^{F_\alpha}}(y_I) = 0$ and $f_{N_b^{F_\alpha}}(y) = 0$, it means y_I does not activate $N_a^{F_\alpha}$ and y does not activate $N_b^{F_\alpha}$. A new feature neuron $N_{new}^{F_\alpha} = \{y, 1\}$ is created to record the input feature y. Then, $N_{new}^{F_\alpha}$ is activated.

Finally, the activation signal generated by the activated feature neuron is transmitted to the unimodal association areas.

Response modal expansion of the association neurons. When a new perceptual channel **X** is added to the network, the association neurons handle the signals transmitted from **X**. As illustrated in Fig. 3e, new synapses will grow to connect the concept neurons in channel **X**, which implies that the response modal of the association neuron is extended.

I denote $x^I = (x_1^I, x_2^I, \ldots, x_n^I) \in \mathbf{R}^n$ as the data received from the pre-existing perceptual channel **L** and $x^{II} = (x_1^{II}, x_2^{II}, \ldots, x_m^{II}) \in \mathbf{R}^m$ as the data received from the new channel **X**.

When an input sample pair (x^I, x^{II}) arrives, x^I and x^{II} activate the concept neurons (the same as in Algorithm 1, steps 2 and 3; therefore, I omit them here). Assume that the concept neurons $N_a^{C_L}$ and $N_b^{C_X}$ are activated in channels **L** and **X**, respectively. Then, $N_a^{C_L}$ and $N_b^{C_X}$ transmit their activated signals to the multimodal association areas to activate the association neurons. Assume the association neurons in set N_L^A are activated by $N_a^{C_L}$ and the association neurons in set N_X^A are activated by $N_b^{C_X}$. Obviously, four combinations exist between set N_L^A and set N_X^A.

When $N_X^A = \varnothing$ and $N_L^A \neq \varnothing$, then x^I activates some association neurons but x^{II} does not. That is, x^I has been encountered previously, but x^{II} is new to the network. The network obtains a new modality feature for the current object through channel **X**. The association neurons in N_L^A should associate $N_b^{C_X}$ with $N_a^{C_L}$. Thus, a new connection is created between $N_b^{C_X}$ and $N_a^{C_L}$ through each association neuron in set N_L^A:

$$c_{(m,i,n)} = \{N_a^{C_L}, N_i^A, N_b^{C_X}, 1\}, \quad \text{where } N_i^A \in N_L^A \tag{5}$$

It is worth noting that Eq. (5) connects the concept neurons in channel **X** to the association neuron which was not previously able to respond to channel **X**. Thus, this process realizes the response-modal expansion process of the association neuron. Subsequently, the network can respond to channel **X**.

When $N_X^A \neq \varnothing$ and $N_L^A \neq \varnothing$, both x^I and x^{II} activate some association neurons, and x^I and x^{II} have been encountered previously. When $N_X^A \cap N_L^A \neq \varnothing$, x^I and x^{II} activate some association neurons in common. The current input pair x^I and x^{II} is an encountered pair. Consequently, the cumulative number of activations of the connections between $N_a^{C_L}$ and $N_b^{C_X}$ through the activated association neurons they have in common are increased by one to strengthen the association:

$$\rho_{(a,i,b)} = \rho_{(a,i,b)} + 1, \quad \text{where } N_i^A \in N_L^A \cap N_X^A$$

when $N_v^A \cap N_a^A = \varnothing$, the current input pair $\boldsymbol{x}^{\mathbf{I}}$ and $\boldsymbol{x}^{\mathbf{II}}$ is not an encountered pair. New connections are created between $N_a^{C_L}$ and $N_b^{C_X}$ to bind them,

$$c_{(m,i,n)} = \{N_a^{C_L}, N_i^A, N_b^{C_X}, 1\}, \quad where \ \ N_i^A \in N_L^A \bigcup N_X^A$$

when $N_X^A \neq \varnothing$ and $N_L^A = \varnothing$, then $\boldsymbol{x}^{\mathbf{II}}$ activates some association neurons but $\boldsymbol{x}^{\mathbf{I}}$ does not. The association neurons in N_X^A will associate $N_a^{C_L}$ with $N_b^{C_X}$. New connections are created between $N_a^{C_L}$ and $N_b^{C_X}$ through the association neurons in set N_X^A:

$$c_{(m,i,n)} = \{N_a^{C_L}, N_i^A, N_b^{C_X}, 1\}, \quad where \ \ N_i^A \in N_X^A$$

when $N_X^A = \varnothing$ and $N_L^A = \varnothing$, neither $\boldsymbol{x}^{\mathbf{I}}$ nor $\boldsymbol{x}^{\mathbf{II}}$ activates any association neurons, and the concepts $N_b^{C_X}$ and $N_a^{C_L}$ are new to the network. A new association neuron is created to connect $N_b^{C_X}$ and $N_a^{C_L}$ with a 1-to-1 circuit.

At this point, the learning for the current input pair $(\boldsymbol{x}^{\mathbf{I}}, \boldsymbol{x}^{\mathbf{II}})$ is complete, and the network continues with the next pair.

Modality embedding.

To embed a new sensory modality in some particular neurons in an expected channel **T**, a guidance signal is needed to guide the new sensory input to a target neuron in channel **T**. Thus, the input format is $(\boldsymbol{x}_t, \boldsymbol{x}_g)$, where $\boldsymbol{x}_t$ is the sample to be embedded and $\boldsymbol{x}_g$ is the guidance signal that will be fed into the guidance channel **G**.

When the network receives $(\boldsymbol{x}_t, \boldsymbol{x}_g)$, it uses the guidance signal $\boldsymbol{x}_g$ to find the target neuron in channel **T** that will absorb $\boldsymbol{x}_t$. Through channel **G**, some association neurons are activated by $\boldsymbol{x}_g$. Then, these association neurons activate the concept neurons to which they are connected in channel **T**, and subsequently, the concept neurons activate their feature neurons. The activated neurons in both channel **G** and channel **T** form a neural circuit that will be reused for the novel modality.

Assume that concept neuron $N_a^{C_T}$ and feature neurons $N_1^{F_T}, N_2^{F_T}, \ldots, N_m^{F_T}$ are activated in channel **T**. In the following, I choose feature neuron $N_f^{F_T}$ as the target neuron to be embedded with $\boldsymbol{x}_t$. The embedding can be divided into two situations: (1) $N_f^{F_T}$ has not previously been embedded with the novel modality and (2) $N_f^{F_T}$ has already been embedded with the novel modality.

For situation one, neuron $N_f^{F_T}$ creates new synapses to connect $\boldsymbol{x}_t$. This means that neuron $N_f^{F_T}$ now has two sets of synapses: one is the original signal pathway, and the other is used to transmit the exogenous signals. The weights of these expanded synapses are initialized to $\boldsymbol{x}_t$,

$$\boldsymbol{w}_f = \{(\boldsymbol{w}_f), \ (\boldsymbol{x}_t)\} \tag{6}$$

In the equation, I use the parentheses to separate the different sets of synaptic weights. Thus, the novel modality is embedded in the target neuron.

For situation two, the weights of the expanded synapses are updated with sample $\boldsymbol{x}_t$ as follows:

$$\sigma_f^{\mathrm{II}} = \sigma_f^{\mathrm{II}} + 1, \ \ \boldsymbol{w}_f^{\mathrm{II}} = \boldsymbol{w}_f^{\mathrm{II}} + \frac{1}{\sigma_f^{\mathrm{II}}}(\boldsymbol{x}_t - \boldsymbol{w}_f^{\mathrm{II}}) \tag{7}$$

where $\boldsymbol{w}_f^{\mathrm{II}}$ represents the weights of the expanded synapses, and σ_f^{II} represents the cumulative number of times these synapses have been activated.

At this point, the embedding process for the current input sample is finished, and the network continues with the next input sample.

Data availability

Matlab code and data are available at https://github.com/cloudlee711/CREAT.

References

1. Jacobs, G. H., Williams, G. A., Cahill, H. & Nathans, J. Emergence of novel color vision in mice engineered to express a human cone photopigment. *Science* **315**, 1723–1725 (2007).
2. Thomson, E. E., Carra, R. & Nicolelis, M. A. L. Perceiving invisible light through a somatosensory cortical prosthesis. *Nature Communications* **4**, 66–78 (2013).
3. O'Doherty, J. E. *et al*. Active tactile exploration using a brain-machine-brain interface. *Nature* **479**, 228–231 (2011).
4. Sur, M., Garraghty, P. E. & Roe, A. W. Experimentally induced visual projections into auditory thalamus and cortex. *Science* **242**, 1437–1441 (1988).
5. Bach-y-Rita, P., Collins, C. C., Saunders, F. A. A., White, B. & Scadden, L. Vision substitution by tactile image projection. *Nature* **221**, 963–964 (1969).
6. Fu, M., Yu, X.-Z., Lu, J. & Zuo, Y. Repetitive motor learning induces coordinated formation of clustered dendritic spines *in vivo*. *Nature* **483**, 92–95 (2012).
7. Yang, G., Pan, F. & Gan, W. B. Stably maintained dendritic spines are associated with lifelong memories. *Nature* **462**, 920–924 (2009).
8. McCulloch, W. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* **5**, 115–133 (1943).
9. Meunier, D., Lambiotte, R., Fornito, A., Ersche, K. D. & Bullmore, E. T. Hierarchical modularity in human brain functional networks. *Frontiers in Neuroinformatics* **3**, 1–12 (2009).

10. Xing, Y.-L., Shen, F.-R., Zhao, J.-X., Pan, J.-X. & Tan, A.-H. Perception coordination network: a framework for online multi-modal concept acquisition and binding. *The Thirty-Second AAAI Conference on Artificial Intelligence*, 4318–4325 (2018).
11. Mesgarani, N., Cheung, C., Johnson, K. & Chang, E. F. Phonetic feature encoding in human superior temporal gyrus. *Science* **343**, 1006–1010 (2014).
12. Gygi, B. *et al.* Speech-specific tuning of neurons in human superior temporal gyrus. *Cerebral Cortex* **24**(10), 2679–2693 (2014).
13. Quiroga, R. Q. Concept cells: the building blocks of declarative memory functions. *Nature Reviews Neuroscience* **13**(8), 587–597 (2012).
14. Hartmann, K. *et al.* Embedding a panoramic representation of infrared light in the adult rat somatosensory cortex through a sensory neuroprosthesis. *Journal of Neuroscience* **36**(8), 2406–2424 (2016).

Acknowledgements

Author contributions

You-Lu Xing designed the computational model and wrote the paper.

Competing interests

The author declares no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-57591-w.

Correspondence and requests for materials should be addressed to Y.-L.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.