This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Artificial Evolution Network: A Computational Perspective on the Expansibility of the Nervous System

You-Lu Xing, Hui Sun, Gui-Huan Feng, Fu-Rao Shen, *Member, IEEE*, and Jian Zhao, *Senior Member, IEEE*

***Abstract*— Neurobiologists recently found the brain can use sudden emerged channels to process information. Based on this finding, we put forward a question whether we can build a computation model that is able to integrate a sudden emerged new type of perceptual channel into itself in an online way. If such a computation model can be established, it will introduce a channel-free property to the computation model and meanwhile deepen our understanding about the extendibility of the brain. In this article, a biologically inspired neural network named artificial evolution (AE) network is proposed to handle the problem. When a new perceptual channel emerges, the neurons in the network can grow new connections to connect the emerged channel according to the Hebb rule. In this article, we design a sensory channel expansion experiment to test the AE network. The experimental results demonstrate that the AE network can handle the sudden emerged perceptual channels effectively.**

***Index Terms*— Artificial evolution (AE), online learning, perceptual channel extensibility, unsupervised learning.**

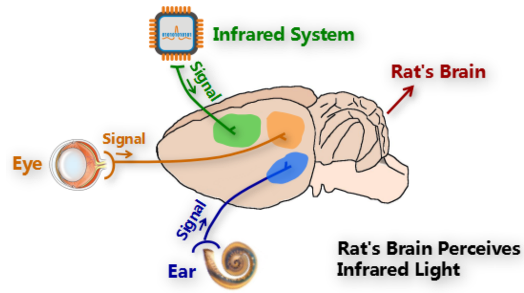


Fig. 1. Rat brain perceives IR light through an implanted IR receiving and processing machine. Experimental results show the rats can learn the IR light as a reward signal, which means that the rats can ground the pattern of the IR light as a concept of "reward." We can regard such a system as a kind of organism–machine hybrid intelligence.

I. Introduction

EVOLUTION plays one of the most important roles for the development of the biological world. Many experiments show that no matter how evolution is achieved, it can indeed help creatures gain new perceptual abilities. For example, in 2007, using genetic engineering, Jacobs *et al.* [1] inserted human L-pigment genes into female mice one X-chromosome and found the heterozygous female mice, whose retinas contain both native mouse pigments (S and M pigments) and human L pigment, show enhanced long-wavelength sensitivity, and acquire a new capacity for chromatic discrimination, meaning that the knock-in mice are able to discriminate among green, yellow, orange, and red panels that, to ordinary mice, look exactly the same. In 2013, Thomson *et al.* [2] implanted stimulating microelectrodes in rat's somatosensory cortex, as shown in Fig. 1. The electrodes connect to an infrared light device that detects and processes infrared light signals and then transmits the processed signals to the microelectrodes. They found that the rat can learn to perceive the invisible infrared light that totally cannot be perceived by natural rats.

Inspired by these experiments, we introduce the concept of artificial evolution (AE) in this article.

Artificial Evolution: It realizes evolution at the living level by receptor-brain artificial expanding, and the living body itself gains new perception abilities. Especially, when the AE is realized using electronic equipments, it implies a kind of organism–machine hybrid intelligence, e.g., the rat model in [2], as shown in Fig. 1.

Evolution in nature usually refers to change in the heritable characteristics of biological populations over successive generations. Changes are a result of genetic recombination, mutations, and other sources of genetic variation. Evolution occurs when natural selection acts on these variations, resulting in certain characteristics becoming common within the next populations. As mentioned earlier, on a macro level, evolution can bring creatures new perceptual receptors and channels, which can lead to sensory augmentation. The AE defined here

Manuscript received June 15, 2018; revised March 11, 2019 and February 17, 2020; accepted June 7, 2020. This work was supported in part by the National Science Foundation of China under Grant 61703002, Grant 61876076, and Grant 61702004, in part by the Jiangsu NSF under Grant BK20171344, and in part by the Dengfeng Scholars Program of Nanjing University under Grant B1512002. *(You-Lu Xing and Hui Sun contributed equally to this work.) (Corresponding authors: Gui-Huan Feng; Jian Zhao.)*

You-Lu Xing and Fu-Rao Shen are with the National Key Laboratory for Novel Software Technology, Department of Computer Science and Technology, Nanjing University, Nanjing 210023, China (e-mail: youluxing@sina.com; frshen@nju.edu.cn).

Hui Sun is with the Department of Computer Science and Technology, Anhui University, Hefei 230601, China (e-mail: sunhui@ahu.edu.cn).

Gui-Huan Feng is with the National Key Laboratory for Novel Software Technology, Software Institute, Nanjing University, Nanjing 210093, China (e-mail: fenggh@nju.edu.cn).

Jian Zhao is with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: jianzhao@nju.edu.cn).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2020.3002556

also brings creatures sensory augmentation. However, different from evolution in nature, it can help creatures themselves get and use new perceptual receptors and channels, not only the offspring generations. This article is also very relevant to neuroplasticity, which is the ability of the brain to change throughout an individual's life, e.g., neural networks of the brain undergo reorganization, and synapses may grow or vanish and strengthen or weaken over time. Neuroplasticity is one important factor to make creatures absorb signals from new perceptual receptors and channels. However, neuroplasticity itself does not involve the concept of sensory augmentation. Thus, the definition of AE here is related to both evolution and neuroplasticity.

Here, we focus on building a computational model to stimulate the third type evolution, i.e., how to integrate a sudden emerged new type of perceptual channel to an existing neural network in an online way. If such a computational model can be established, the following achievements will be obtained.

1) A channel-free property will be introduced to the neural network, which enables the neural network to adapt to suddenly emerged channels.
2) Deepen our understanding about the extendibility of the central nervous system.

Meanwhile, many potential useful applications for the robot platform will be introduced. For example, if we install new sensors to a robot, with such a neural network model, we do not need to retrain the robot offline, and the robot can automatically use the newly installed sensors in the online fashion.

II. Related Work

Biologically inspired computational model is one of the most important driving forces for the development of artificial intelligence. A large number of biologically inspired models have been proposed, which made many pioneering contributions.

Inspired by the tonotopy organization of the auditory cortex, Kohonen [3] designed the self-organizing map (SOM) that arranges features orderly in an unsupervised way. Martinetz and Schulten [4] introduced the Hebb learning rule [5] to the SOM, which made that the connections between neurons can be automatically learned. Later, Fritzke [6] proposed a growing neural gas (GNG) that improved the SOM to learn the number of the neurons in an unsupervised way. Faigl and Hollinger [7] proposed a growing SOM that can adapt the number of neurons during learning. Gorzałczany and Rudziński [8] presented a generalization of SOM with 1-D neighborhoods. It allows dynamic adjustment of the neuron chain and the number of neurons.

Neurophysiological studies show that the nervous system has a high plasticity while maintaining a strong stability [9]. It is one of the most important advantages of the nervous system. For example, the nervous system can quickly learn new things (e.g., the appearance, the name, and the taste of an apple) without forgetting previously learned ones (e.g., the appearance, the name, and the taste of a banana). Carpenter and Grossberg [10] found many algorithms suffer the stability–plasticity dilemma, meaning that they cannot guarantee a good plasticity and stability at the same time. To solve this problem, Grossberg [11] proposed a series of incremental learning algorithms that create a new memory unit when no match occurs between the current input sample and the current learned category set. Meng *et al.* [12] developed the ART to multichannel model that can be used for multimodal features fusion. Shen and Hasegawa [13] introduced this idea to the SOM and proposed a self-organizing incremental neural network (SOINN), which is able to learn representative points and topological structure of the training data in an online incremental way. After that, He *et al.* [14] used the SOINN to handle the incremental word grounding problem. Nakamura and Hasegawa [15] introduced kernel density estimation to the SOINN to estimate the probability density function of online data.

Brain science studies show that the brain has a modular structure [16]. To stimulate such a structure, the modular neural network (MNN) [17] is developed. The MNN contains a series of independent neural networks that perform some subtask and cooperate together to complete the whole task. Schyns [18] proposed an MNN for concept learning, where an SOM is used to build a concept prototype and a brain-state-in-a-box (BSB) [19] is used to associate concept names. Yamauchi *et al.* [20] designed a sensory integration system, where several neural network modules, which consist of forward and backward parts, are fused by some integrating units. Jantvik *et al.* [21] developed a MultiModal Self-Organizing Network (MMSON) for sensory integration with SOM modules. Positional coordinates of unimodal SOM, which receive sensory data, are fused by a high-level SOM. Based on the MMSON, a bimodal incremental self-organizing network (BiSON) [22] is designed, which can incrementally integrate stimuli in visual and auditory modalities.

Besides modular structure, the brain also has a hierarchical structure, which means that the neurons in the high layer integrate the functions of the neurons in the low layer [23]. Many computational models are proposed to realize this structure. Inspired by the simple cells and complex cells in the Hubel–Wiesel model, Fukushima [24] designed a hierarchical neural network named neocognitron that is formed of alternating layers with S-cells and C-cells corresponding to simple cells and complex cells, respectively. Rajapakse and Acharya [25] used the neocognitron to handle the multisensor data fusion, and an ART-1-type fast learning algorithm [26] was introduced to update the weights of the network. Combining neocognitron's structure with backpropagation algorithm, LeCun *et al.* [27] and Krizhevsky *et al.* [28] built a deep convolutional neural networks (CNNs), which achieved a great success in the handwriting recognition task and ImageNet competition. Recently, Palomo and López-Rubio [29] proposed a hierarchical GNG, which can learn a tree of graphs. To combine the advantages of modular structure and hierarchical structure, Xing *et al.* [30], [31] designed a perception coordination network (PCN) that is used to handle the multimodal concept acquisition and binding problem. In the network, each modular charges unimodal concept acquisition

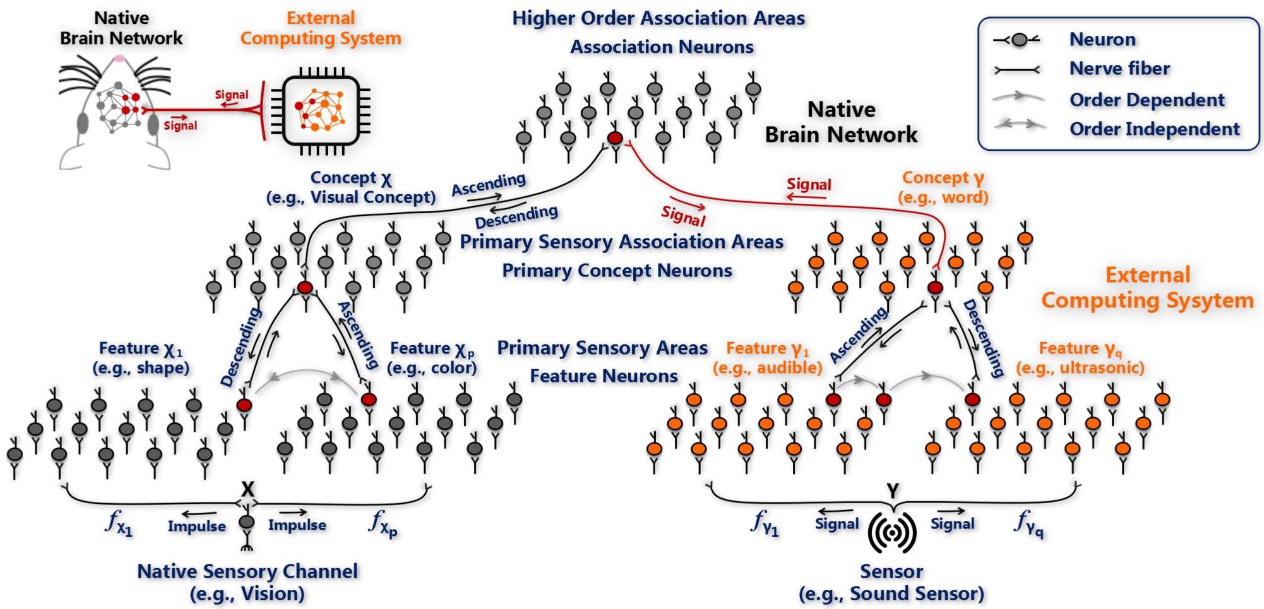


Fig. 2. Computational model of the AE. The gray part is the native neural network and the yellow part is the suddenly emerged channel. The framework is a brain-like hierarchical structure which is divided into the PSA, the primary SAA, and the HAA. The PSA includes feature neurons that respond to some elementary features, e.g., shape features, color features, audible syllable features, and ultrasonic syllable features. The SAA includes primary concept neurons that connect the feature neurons to represent some unimodal concepts, e.g., to form visual concepts by connecting shape and color feature neurons, to form auditory concepts (words) by connecting audible and ultrasonic syllable feature neurons. The HAA includes association neurons that connect primary concept neurons in different SAAs to form a multimodal concept representation. The association neurons and neuroplasticity play an adhesive function between the native channel and the suddenly emerged channel. The connections of the association neurons are extendable (the red color connection in the figure) during online learning, which makes the native neural network can integrate sudden emerged channels easily.

and the concept in each modular will be bound through some vertical connections in the hierarchical structure.

Inspired by the human L-pigment genes, knocked-in mice show an enhanced long-wavelength sensitivity and acquire a new capacity for chromatic discrimination, and recently, a perception evolution network (PEN) [32]–[34] is introduced to handle a problem of emergence of novel sensory inputs during online learning. The PEN directly added the weight vector learned from novel sensory inputs to the tail of the weight vector of a neuron when novel sensory inputs are introduced to the network. One disadvantage of the PEN is that it treats old and novel features as a unit input after novel sensory inputs is added. As a result, PEN needs data from all old sensory inputs to integrate emerged sensory inputs, which is very inefficient when the number of sensory inputs increases. Meanwhile, PEN is a nonhierarchical and nonmodular single-layer structure that works in an SOM [3] way, which is not flexible enough to integrate emerged inputs conveniently, e.g., it cannot combine features and concepts effectively and cannot handle novel sensory input with different dimensions, such as voice inputs.

After the abovementioned exhaustive investigations on studies of the neural network society in recent several decades, we find a few of them discuss the artificial sensory augment phenomenon [2], which is defined as a kind of AE here. Obviously, giving the AE, computational modeling is a very attractive problem because, as mentioned before, it can deepen our understanding about the extendibility of the central nervous system and introduce a channel-free property to the artificial neural networks.

III. Computational Model

In this section, a biologically inspired neural network model named AE network is proposed to integrate a sudden emerged new type perceptual channel in an online way. Fig. 2 shows the computational model of the AE. In the figure, the gray part is the native neural network and the yellow part is the suddenly emerged channel. The framework is a hierarchical structure that is divided into the primary sensory area (PSA), the primary sensory association area (SAA), and the higher order association area (HAA). The PSA includes feature neurons that respond to some elementary features, e.g., shape features, color features, audible syllable features, and ultrasonic syllable features. The SAA includes primary concept neurons that connect the feature neurons to represent some unimodal concepts, e.g., to form visual concepts by connecting shape and color feature neurons, to form auditory concepts[1] by connecting audible and ultrasonic syllable feature neurons. There are two activation ways of the primary concept neurons, which are the Order InDependent Activation (OIDA) model and order-dependent activation (ODA) model. For example, visual concept neurons have an OIDA model because different activation orders of color and shape features do not affect the activation of their corresponding visual concept. Auditory concept neurons (words) have an ODA model because different activation orders of the same group of syllables may refer to different concepts (words). The HAA includes

[1]For example, auditory concepts are denoted by words. For convenience, in this article, auditory concepts refer to words.

association neurons that connect primary concept neurons in different SAAs to form a multimodal concept representation. Importantly, the association neurons play an adhesive function between a native channel in the existing neural network and the suddenly emerged channel. The connections of the association neurons are extendable during online learning, which makes the native neural network can integrate sudden emerged channels easily.

In the following, we first give the problem formulation. Then, we give the learning algorithm of the AE network.

A. Problem Formulation

Assume that a neural network originally has a perceptual channel $\mathbf{X}$, which receives n-dimensional data $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$. The neural network receives the input data from $\mathbf{X}$ in an online way. After a period of time, a new type perceptual channel $\mathbf{Y}$ is added to the neural network, which receives m-dimensional data $\boldsymbol{y} = (y_1, y_2, \ldots, y_m) \in \mathbf{R}^m$. Then, the neural network has one more way to perceive the external world and the received data by the neural network can be a pair of $(\boldsymbol{x}, \boldsymbol{y})$, where $\boldsymbol{x}$ comes from channel $\mathbf{X}$ and $\boldsymbol{y}$ comes from channel $\mathbf{Y}$. Now, the neural network should learn through the new channel $\mathbf{Y}$ and meanwhile integrate the concepts learned through channel $\mathbf{Y}$ with the concepts learned through channel $\mathbf{X}$, which means integrating channel $\mathbf{Y}$ into the existing neural network in an online way. In this setting, we can call channel $\mathbf{X}$ as a reference channel, which is indispensable when the system integrates the newly emerged channel $\mathbf{Y}$. Similarly, if another channel $\mathbf{Z}$ emerges after channel $\mathbf{Y}$, a reference channel is also needed for integrating channel $\mathbf{Z}$. At this time, we can choose either channel $\mathbf{X}$ or channel $\mathbf{Y}$ as the reference channel, then the input of the neural network is a pair of $(\boldsymbol{x}, z)$ or $(\boldsymbol{y}, z)$, where z comes from channel $\mathbf{Z}$. In the following, we take channel $\mathbf{X}$ and channel $\mathbf{Y}$ as an example to describe our method.

B. Learning Algorithm

The AE network is an online learning model, which means that samples are fed into the network sequentially. When a pair of inputs $(\boldsymbol{x}, \boldsymbol{y})$ arrive, the PSA extracts the features from $\boldsymbol{x}$ and $\boldsymbol{y}$ first. Then, competitive learning among feature neurons is conducted, and firing feature neurons transmit activation signals to their corresponding SAA. When the SAA receives the ascending signal from PSA, unimodal concept learning is executed in channel $\mathbf{X}$ and channel $\mathbf{Y}$. The firing concept neurons then transmit activation signals to the HAA. In the HAA, the activation signals from the SAA activate some association neurons first. Then, the activate association neurons integrate the firing concept neuron in channel $\mathbf{Y}$ with the firing concept neuron in channel $\mathbf{X}$ by creating new connections between them (see the red color connection in the figure). It can be found that the sudden emerged channel $\mathbf{Y}$ is integrated into the existing neural network through this area. By now, the current input pair is complete, and AE will then deal with the next input pair. In the following, we describe the method in detail.

1) Primary Sensory Area: The PSA includes the feature neurons that respond to particular features, e.g., shape, color, audible syllable, or ultrasonic syllable features. Feature neurons that respond to the same type feature are located in the same area D and we use set N^{F_D} to store them. $N_i^{F_D} \triangleq \{\boldsymbol{w}_i^{F_D}, \sigma_i^{F_D}\}$ denotes feature neuron i in feature area D, where $\boldsymbol{w}_i$ and σ_i represent the weights and the activation times of feature neuron i, respectively. The activating domains (ADs) of $N_i^{F_D}$ are defined as

$$\boldsymbol{ADs}(N_i^{F_D}) \triangleq \{\boldsymbol{x} \mid \boldsymbol{Dis}(\boldsymbol{x}, \boldsymbol{w}_i) \leq \theta\} \tag{1}$$

where θ is a parameter that controls the range of the ADs. Different types of feature neurons may have different strategies of deciding θ, usually according to the characteristic of the used feature. $\boldsymbol{Dis}(\cdot, \cdot)$ is the distance function. Euclidean distance and dynamic time warping are used for vectors with the same and different dimensions.

In each feature area, horizontal connections between feature neurons are developed to organize the features neurons into a feature map. The connection between feature neuron $N_i^{F_D}$ and feature neuron $N_j^{F_D}$ is defined as

$$c_{(i,j)}^h \triangleq \left\{N_i^{F_D}, N_j^{F_D}, \tau_{(i,j)}\right\} \tag{2}$$

where $\tau_{(i,j)}$ is a time parameter that records the age of the connection.

When a pair sample $(\boldsymbol{x}, \boldsymbol{y})$ comes, feature extraction is conducted first. We use the following equations to abstract the feature extraction processing:

$$\begin{aligned} \boldsymbol{x}_i' &= f_{\chi_i}(\boldsymbol{x}), \quad 1 \leq i \leq p \\ \boldsymbol{y}_i' &= f_{\gamma_i}(\boldsymbol{y}), \quad 1 \leq i \leq q \end{aligned} \tag{3}$$

where $D = \{\chi_1, \chi_2, \ldots, \chi_p, \gamma_1, \gamma_2, \ldots, \gamma_q\}$, and p and q are the number of different types of features used in channel $\mathbf{X}$ and $\mathbf{Y}$. For example, in the experiment, we set two feature extraction functions in a visual channel $\mathbf{X}$, which are the normalized Fourier descriptors (NFDs) of the object's boundary and the color histogram (CH) of the object's area in the image. Then, here, $p = 2$, f_{χ_1} is the NFD function, and f_{χ_2} is the CH function. Meanwhile, we set one feature extraction function in an auditory channel $\mathbf{Y}$, which is the mel-frequency cepstral coefficients (MFCCs) of the syllable contained in an input voice. Then, here, $q = 1$ and f_{γ_1} is the MFCC function. Details of the features will be described in Section IV.

After the feature extraction step, competitive learning in the PSA is executed. First, a winner feature neuron in each feature area χ_i and γ_i is calculated as follows:

$$N_{k_i}^{F_{\chi_i}} = \underset{N_j^{F_{\chi_i}} \in N^{F_{\chi_i}}}{\operatorname{argmin}} \boldsymbol{Dis}\big(\boldsymbol{x}_i', \boldsymbol{w}_j^{F_{\chi_i}}\big), \quad 1 \leq i \leq p \tag{4}$$

$$N_{k_i}^{F_{\gamma_i}} = \underset{N_j^{F_{\gamma_i}} \in N^{F_{\gamma_i}}}{\operatorname{argmin}} \boldsymbol{Dis}\big(\boldsymbol{y}_i', \boldsymbol{w}_j^{F_{\gamma_i}}\big), \quad 1 \leq i \leq q \tag{5}$$

where $\boldsymbol{Dis}(\cdot, \cdot)$ is the distance function. Euclidean distance and dynamic time warping are used for vectors with the same and different dimension.

If $\boldsymbol{x}_i'$ or $\boldsymbol{y}_i'$ belongs to the ADs of neuron $N_{k_i}^{F_{\chi_i}}$ or $N_{k_i}^{F_{\gamma_i}}$, $N_{k_i}^{F_{\chi_i}}$ and $N_{k_i}^{F_{\gamma_i}}$ are activated. Meanwhile, the neurons are updated as follows[2]:

$$\sigma_{k_i}^{F_{\chi_i}} = \sigma_{k_i}^{F_{\chi_i}} + 1; \quad \boldsymbol{w}_{k_i}^{F_{\chi_i}} = \boldsymbol{w}_{k_i}^{F_{\chi_i}} + (\boldsymbol{x}_i' - \boldsymbol{w}_{k_i}^{F_{\chi_i}})/\sigma_{k_i}^{F_{\chi_i}} \tag{6}$$

$$\sigma_{k_i}^{F_{\gamma_i}} = \sigma_{k_i}^{F_{\gamma_i}} + 1; \quad \boldsymbol{w}_{k_i}^{F_{\gamma_i}} = \boldsymbol{w}_{k_i}^{F_{\gamma_i}} + (\boldsymbol{x}_i' - \boldsymbol{w}_{k_i}^{F_{\gamma_i}})/\sigma_{k_i}^{F_{\gamma_i}}. \tag{7}$$

If $\boldsymbol{x}_i'$ or $\boldsymbol{y}_i'$ does not belong to the ADs of neuron $N_{k_i}^{F_{\chi_i}}$ or $N_{k_i}^{F_{\gamma_i}}$, the AE network recognizes such a feature as a new feature and remember the new feature in its memory, which means that a new neuron is created to store the current features

$$N_{\text{new}}^{F_{\chi_i}} = \{\boldsymbol{x}_i', 1\}, \quad N_{\text{new}}^{F_{\gamma_i}} = \{\boldsymbol{y}_i', 1\}. \tag{8}$$

Then, $N_{\text{new}}^{F_{\chi_i}}$ and $N_{\text{new}}^{F_{\gamma_i}}$ are activated. For convenience, we rewrite $N_{\text{new}}^{F_{\chi_i}}$ and $N_{\text{new}}^{F_{\gamma_i}}$ as $N_{k_i}^{F_{\chi_i}}$ and $N_{k_i}^{F_{\gamma_i}}$, respectively.

Meanwhile, self-organizing [3] among feature neurons is conducted according to the competitive Hebbian learning rule [4]. Take the feature area χ_i for example, and if the winner neuron $N_{f_i}^{F_{\chi_i}}$ is activated when an input sample $\boldsymbol{x}$ comes, the following condition will be checked for all other neurons in feature area χ_i:

$$\boldsymbol{Dis}(\boldsymbol{x}_i', \boldsymbol{w}_j) \le \delta \cdot \theta, \text{ where } N_j^{F_{\chi_i}} \in N^{F_{\chi_i}} \setminus \{N_{f_i}^{F_{\chi_i}}\} \tag{9}$$

where δ is a constant to control the activation behavior of $N_j^{F_{\chi_i}}$ during self-organizing processing, and usually, it is set larger than or equal to 1. We set $\delta = 3$ in this article.

If feature neuron $N_j^{F_{\chi_i}}$ satisfies (9), a horizontal connection is established between $N_{f_i}^{F_{\chi_i}}$ and $N_j^{F_{\chi_i}}$ if no connection exists between them, that is

$$c_{(f_i,j)}^h = \{N_{f_i}^{F_{\chi_i}}, N_j^{F_{\chi_i}}, 1\}. \tag{10}$$

If there is already a connection between them, the time parameter $\tau_{(f_i,j)}$ will be set to 0 to renew the connection

$$\tau_{(f_i,j)} = 0. \tag{11}$$

During learning, (6) and (7) change neuron's weights continuously. As a result, feature neurons that are connected to each other at an early stage may not have similar weights at an advanced stage. The connections between such neurons should be weakened or removed. Therefore, if a winner neuron $N_{f_i}^{F_{\chi_i}}$ is activated and updated, the connections emanating from it will be weakened by increasing the age parameter as follows:

$$\tau_{(f_i,k)} = \tau_{(f_i,k)} + 1, \quad N_k^{F_{\chi_i}} \in S_{f_i}^{F_{\chi_i}} \tag{12}$$

where $S_{f_i}^{F_{\chi_i}}$ represents the neighbor neuron set of $N_{f_i}^{F_{\chi_i}}$, which means that neurons in $S_{f_i}^{F_{\chi_i}}$ are connected to $N_{f_i}^{F_{\chi_i}}$ directly. Connections $c_{i,j}^h$ whose time parameter is larger than a predefined threshold t_τ will be removed, and t_τ is set as 50 here.

Finally, the activated signals of the firing feature neurons in the PSA, $N_{k_1}^{F_{\chi 1}}, N_{k_2}^{F_{\chi 2}}, \ldots, N_{k_p}^{F_{\chi p}}$ and $N_{k_1}^{F_{\gamma 1}}, N_{k_2}^{F_{\gamma 2}}, \ldots, N_{k_q}^{F_{\gamma q}}$, are transmitted to their corresponding SAA.

[2]If the dimension of the weights is not fixed in some area χ_i or γ_i, e.g., the weights of different syllable feature neurons, we do not update the weights of the neurons during online learning using this way.

2) Primary Sensory Association Area: The SAA includes the primary concept neurons that connect the feature neurons in the same perceptual channel to represent unimodal concepts, e.g., connecting shape and color feature neurons to form a visual concept or connecting syllable feature neurons to form a word (auditory concept). Primary concept neurons in the same SAA R are stored in set N^{C_R}. $N_i^{C_R}$ denotes the concept neuron i in area R. The connection between a concept neuron $N_i^{C_R}$ and a feature neuron $N_j^{F_D}$ is defined as

$$c_{(i,j)} \triangleq \{N_i^{C_R}, N_j^{F_D}, \rho_{(i,j)}\} \tag{13}$$

where $\rho_{(i,j)}$ is an activity parameter of $c_{(i,j)}$, which records the number of times that the connection is activated.

The ADs of the concept neurons are defined as

$$\boldsymbol{ADs}(N_i^{C_R}) \triangleq \begin{cases} \overrightarrow{(N_1^{F_D}, N_2^{F_D}, \ldots, N_n^{F_D})}, & \text{ODA} \\ (N_1^{F_D}, N_2^{F_D}, \ldots, N_n^{F_D}), & \text{OIDA} \end{cases} \tag{14}$$

where feature neurons $N_1^{F_D}$, $N_2^{F_D}$, …, $N_n^{F_D}$ connect the concept neuron $N_i^{C_R}$. Note that the arrow over the vector means that $N_i^{C_R}$ can only be activated by the firing of $N_1^{F_D}$, $N_2^{F_D}$,…, $N_n^{F_D}$ through the arrow's direction. In the following, we assume that the primary concept neurons in channel $\boldsymbol{X}$ have an OIDA model and the primary concept neurons in channel $\boldsymbol{Y}$ have an ODA model.

When the SAA receives the activation signals from $N_{k_1}^{F_{\chi 1}}$, $N_{k_2}^{F_{\chi 2}}$, …, $N_{k_p}^{F_{\chi p}}$ and $N_{k_1}^{F_{\gamma 1}}$, $N_{k_2}^{F_{\gamma 2}}$, …, $N_{k_q}^{F_{\gamma q}}$, the signals are checked whether equal to any concept neuron's ADs, which means to find the solution to the following equations:

$$\boldsymbol{ADs}(N_i^{C_\chi}) = (N_{k_1}^{F_{\chi 1}}, N_{k_2}^{F_{\chi 2}}, \ldots, N_{k_p}^{F_{\chi p}}), \quad N_i^{C_\chi} \in N^{C_\chi} \tag{15}$$

$$\boldsymbol{ADs}(N_i^{C_\gamma}) = \overrightarrow{(N_{k_1}^{F_{\gamma 1}}, N_{k_2}^{F_{\gamma 2}}, \ldots, N_{k_q}^{F_{\gamma q}})}, \quad N_i^{C_\gamma} \in N^{C_\gamma} \tag{16}$$

where $R = \{\chi, \gamma\}$, χ means the SAA in channel $\mathbf{X}$ and γ means the SAA in the emerged channel $\mathbf{Y}$.

If some concept neurons $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ are found by (15) and (16), then $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ are activated, and the activity parameters of connections between $N_a^{C_\chi}$ and $N_{k_1}^{F_{\chi 1}}$, $N_{k_2}^{F_{\chi 2}}$, …, $N_{k_p}^{F_{\chi p}}$, and $N_b^{C_\gamma}$ and $N_{k_1}^{F_{\gamma 1}}$, $N_{k_2}^{F_{\gamma 2}}$, …, $N_{k_q}^{F_{\gamma q}}$, are increased by one, that is

$$\rho_{(a,k_i)} = \rho_{(a,k_i)} + 1, \quad 1 \le i \le p \tag{17}$$

$$\rho_{(b,k_i)} = \rho_{(a,k_i)} + 1, \quad 1 \le i \le q. \tag{18}$$

If no concept neuron is found according to (15) or (16), it means that the concept represented by the current input has not been encountered by AE network before. The SAA will remember this new concept and a new concept neuron $N_{\text{new}}^{C_\chi}$ or $N_{\text{new}}^{C_\gamma}$ will be created as follows:

$$\boldsymbol{ADs}(N_{\text{new}}^{C_\chi}) = (N_{k_1}^{F_{\chi 1}}, N_{k_2}^{F_{\chi 2}}, \ldots, N_{k_p}^{F_{\chi p}}) \tag{19}$$

$$\boldsymbol{ADs}(N_{\text{new}}^{C_\gamma}) = \overrightarrow{(N_{k_1}^{F_{\gamma 1}}, N_{k_2}^{F_{\gamma 2}}, \ldots, N_{k_q}^{F_{\gamma q}})}. \tag{20}$$

Then, $N_{\text{new}}^{C_\chi}$ and $N_{\text{new}}^{C_\gamma}$ are activated. For convenience, we rewrite $N_{\text{new}}^{C_\chi}$ and $N_{\text{new}}^{C_\gamma}$ as $N_a^{C_\chi}$ and $N_b^{C_\gamma}$, respectively.

Meanwhile, connections between the new concept neuron and each feature neurons are created as follows:

$$\rho_{(a,k_i)} = \left\{N_a^{C_\chi}, N_{k_i}^{F_{\chi_i}}, 1\right\}, \quad 1 \leq i \leq p \tag{21}$$

$$\rho_{(b,k_i)} = \left\{N_b^{C_\gamma}, N_{k_i}^{F_{\gamma_i}}, 1\right\}, \quad 1 \leq i \leq q. \tag{22}$$

Finally, the activated signals of the firing concept neurons, $N_a^{C_\chi}$ and $N_b^{C_\gamma}$, are transmitted to the HAA.

3) Higher Order Association Area: The HAA includes the association neurons that connect primary concept neurons in different SAAs. The association neurons play a communication roll between the native channel **X** and the sudden emerged channel **Y**. New connections between the association neurons and the concept neurons in channel **Y** will be created during online learning, which makes the native neural network is able to integrate channel **Y**. Association neurons are stored in set N^A, and N_i^A is used to denote association neuron i. The connection between a concept neuron $N_m^{C_R}$ and another concept neuron $N_n^{C_R}$ through N_i^A is defined as follows:

$$c_{(m,i,n)} \triangleq \left\{N_m^{C_R}, N_i^A, N_n^{C_R}, \rho_{(m,i,n)}\right\} \tag{23}$$

where $\rho_{(m,i,n)}$ is an activity parameter of $c_{(m,i,n)}$, which records the number of times that the connection is activated, $N_m^{C_R}$ and $N_n^{C_R}$ come from different PSAs, e.g., $N_m^{C_R}$ is a concept neuron in channel **X**, and $N_n^{C_R}$ is a concept neuron in channel **Y**. The ADs of N_i^A is the primary concept neurons that N_i^A connects, which means

$$\boldsymbol{ADs}\left(N_i^A\right) \triangleq \left\{N_1^{C_R}, N_2^{C_R}, \ldots, N_n^{C_R}\right\}. \tag{24}$$

Note that the ADs of N_i^A are a set of primary concept neurons, and it means that any concept neurons in the set can activate it.

The learning algorithm in this area is divided into two situations: 1) auditory channel for communication is included and 2) auditory channel for communication is not included. In the following, we first describe the learning algorithm in situation 1, where channel **Y** is assumed as an auditory channel for communication, and then, we go to learning algorithm in situation 2.

Situation 1: When the HAA receives the activation signals from $N_a^{C_\chi}$ and $N_b^{C_\gamma}$, the signals are checked whether fall into any association neuron's ADs, i.e., to find the solution of the following equations:

$$\boldsymbol{ADs}\left(N_i^A\right) \ni N_a^{C_\chi}, \quad N_i^A \in N^A \tag{25}$$

$$\boldsymbol{ADs}\left(N_i^A\right) \ni N_b^{C_\gamma}, \quad N_i^A \in N^A. \tag{26}$$

Assume that the solution sets of (25) and (26) are S^χ and S^γ.

Then, the activated association neuron(s) in set S^χ and S^γ will try to activate the concept neuron(s) that they connect in channels **Y** and **X**, respectively, i.e., to find the solution of the following equations:

$$N_i^{C_\gamma} \in \boldsymbol{ADs}\left(N_j^A\right), \quad N_j^A \in S^\chi \tag{27}$$

$$N_i^{C_\chi} \in \boldsymbol{ADs}\left(N_j^A\right), \quad N_j^A \in S^\gamma. \tag{28}$$

Assume that the solution sets of (27) and (28) are U^γ and U^χ. Concept neuron(s) in set U^γ is the conceptual representation of $N_a^{C_\chi}$ in channel **Y**, and concept neuron(s) in set U^χ is the conceptual representation of $N_b^{C_\gamma}$ in channel **X**.

For (25) and (26), there are four conditions of the solutions.

Fisrt, $S^\chi \neq \varnothing$ and $S^\gamma = \varnothing$, which means that the concept neuron $N_b^{C_\gamma}$ in the emerged channel **Y** has not yet been connected to the neural network through any association neurons. Meanwhile, the concept neuron $N_a^{C_\chi}$ in the native channel **X** has been connected to some association neuron(s) in set S^χ.

Now, if $U^\gamma = \varnothing$, i.e., association neuron(s) in set S^χ has not yet been connected to any concept neurons in channel **Y**, this implies that concept neuron $N_a^{C_\chi}$ has not been grounded with any concept neurons in channel **Y**. In this case, the AE network creates a new connection between concept neuron $N_b^{C_\gamma}$ and each association neuron in set S^χ to associate concept $N_b^{C_\gamma}$ with concept neuron $N_a^{C_\chi}$ as follows:

$$c_{(a,i,b)} = \left\{N_a^{C_\chi}, N_i^A, N_b^{C_\gamma}, 1\right\}, \quad N_i^A \in S^\chi. \tag{29}$$

Meanwhile, the ADs of each association neuron in set S^χ are expanded with $N_b^{C_\gamma}$ as follows:

$$\boldsymbol{ADs}\left(N_i^A\right) = \boldsymbol{ADs}\left(N_i^A\right) \cup N_b^{C_\gamma}, \quad N_i^A \in S^\chi. \tag{30}$$

Now, the concept neuron $N_b^{C_\gamma}$ in channel **Y** is integrated into the existing network.

If $U^\gamma \neq \varnothing$, this means that some concept neurons in channel **Y** are activated by the association neuron(s) in S^χ, i.e., concept neuron $N_a^{C_\chi}$ has been grounded with some concept neurons in channel **Y**. However, $N_b^{C_\gamma}$ is not in set U^γ. Then, a contradiction happens and AE network will ask the communicator a question: " The concept $N_a^{C_\chi}$ that I recognized in channel **X** is represented by the concept neurons (words) in set U^γ before. The current input concept (word) $N_b^{C_\gamma}$ in channel **Y** is different from those concept neurons (words) in set U^γ. $N_b^{C_\gamma}$ can also represent concept $N_a^{C_\chi}$?" An answer from the communicator is needed to help make a judgment[3].

If the answer is positive, i.e., $N_b^{C_\gamma}$ can represent concept $N_a^{C_\chi}$, then the network is updated with formula (29) and formula (30). If the answer is negative, the network will not associate $N_b^{C_\gamma}$ with $N_a^{C_\chi}$.

Second, $S^\chi \neq \varnothing$ and $S^\gamma \neq \varnothing$, which means that the concept neurons $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ both have been connected to some association neuron(s). Then, the AE network will check the coherence of the current input.

If $S^\chi \cap S^\gamma \neq \varnothing$, i.e., $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ activate some association neuron(s) in common, it implies that the AE network recognizes the current input pair as an encountered one. For convenience, we use set I to represent the activated association neuron(s) in common. Then, the connections between concept neuron $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ through each activate association neuron in set I are strengthened by increasing their activity parameters

$$\rho_{(a,i,b)} = \rho_{(a,i,b)} + 1, \quad N_i^A \in I. \tag{31}$$

If $S^\chi \cap S^\gamma = \varnothing$, i.e., $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ do not activate any association neurons in common, it implies that the current

[3]In the experiment, the answer is input to the method by the communicator, e.g., input "N" means negative and input "Y" means positive.

combination $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ is inconsistent with some previous combinations. The AE network will ask the communicator a question: "The current input pair $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ is inconsistent with some previous pairs, is it an expected combination?" An answer from the communicator is needed to help make a judgment.

If the answer is positive, it means that the current input concepts $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ recognized by the AE network are indeed an expected combination. Then, concept neurons $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ are added to the ADs of each association neuron in set S^χ and S^γ, respectively

$$\boldsymbol{ADs}\left(N_i^A\right) = \boldsymbol{ADs}\left(N_i^A\right) \cup N_b^{C_\gamma}, \quad N_i^A \in S^\chi \tag{32}$$

$$\boldsymbol{ADs}\left(N_i^A\right) = \boldsymbol{ADs}\left(N_i^A\right) \cup N_a^{C_\chi}, \quad N_i^A \in S^\gamma. \tag{33}$$

Meanwhile, connections between the new combinations are initialized as

$$c_{(a,i,b)} = \left\{N_a^{C_\chi}, N_i^A, N_b^{C_\gamma}, 1\right\}, \quad N_i^A \in S^\chi \cup S^\gamma. \tag{34}$$

If the answer is negative, it means that the combination is not an expected one. The network will not associate $N_b^{C_\gamma}$ with $N_a^{C_\chi}$.

Third, $S^\chi = \varnothing$ and $S^\gamma \neq \varnothing$, which means that $N_a^{C_\chi}$ in channel $\mathbf{X}$ has not been connected to any association neurons, i.e., channel $\mathbf{X}$ recognizes the current input $\boldsymbol{x}$ as a new concept, but $N_b^{C_\gamma}$ in channel $\mathbf{Y}$ has been connected to some association neuron(s). The activated concept neuron $N_a^{C_\chi}$ is not in set U^χ. A contradiction happens and AE will ask the communicator a question: "The concept $N_b^{C_\gamma}$ that I recognized in channel $\mathbf{Y}$ is represented by the concept neurons in set U^χ before. The current input concept (word) $N_b^{C_\gamma}$ in channel $\mathbf{Y}$ can also represent concept $N_a^{C_\chi}$?" An answer from the communicator is needed.

If the answer is positive, a new connection between concept neuron $N_a^{C_\chi}$ and each association neuron in set S^γ is created to associate concept neuron $N_a^{C_\chi}$ with concept neuron $N_b^{C_\gamma}$

$$c_{(a,i,b)} = \left\{N_a^{C_\chi}, N_i^A, N_b^{C_\gamma}, 1\right\}, \quad N_i^A \in S^\gamma. \tag{35}$$

Meanwhile, the ADs of each association neuron in set S^γ are expanded with $N_a^{C_\chi}$

$$\boldsymbol{ADs}(N_i^A) = \boldsymbol{ADs}\left(N_i^A\right) \cup N_a^{C_\chi}, \quad N_i^A \in S^\gamma. \tag{36}$$

If the answer is negative, it means that the combination is not an expected one. The network will not associate $N_a^{C_\chi}$ with $N_b^{C_\gamma}$.

Fourth, $S^\chi = \varnothing$ and $S^\gamma = \varnothing$, which means that neither $N_a^{C_\chi}$ nor $N_b^{C_\gamma}$ has been connected to any association neurons. This implies that $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ are new to the AE network. Then, a new association neuron N_{new}^A is initialized to associate concept neuron $N_a^{C_\chi}$ and concept neuron $N_b^{C_\gamma}$

$$\boldsymbol{ADs}\left(N_{\text{new}}^A\right) = \left\{N_{k_1}^{C_\chi}, N_{k_2}^{C_\gamma}\right\}. \tag{37}$$

A connection between $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ through N_{new}^A is created as follows:

$$c_{(a,\text{new},b)} = \left\{N_a^{C_\chi}, N_{\text{new}}^A, N_b^{C_\gamma}, 1\right\}. \tag{38}$$

Now, the learning for the current input pair $(\boldsymbol{x}, \boldsymbol{y})$ finishes. The AE network will go to the next input pair. It can be found that among the four conditions, conditions 1 and 4 are in charge of integrating channel $\mathbf{Y}$ into the existing network; conditions 2 and 3 are in charge of samples from channel $\mathbf{Y}$ that already has been integrated into the network. During online learning, new connections are created to integrate the emerged channel $\mathbf{Y}$ with the native channel $\mathbf{X}$ according to the Hebb rule, i.e., neurons that fire together, wire together [35], which makes the associated neurons expand their response modality with the modality of channel $\mathbf{Y}$. This is also conformed to the experimental finding in [2], which is neurons that respond to IR light inputs maintaining their original sensory modality, the response modality of these neurons is expanded.

By now, the learning algorithm of the HAA in situation 1, i.e., auditory channel for communication is included, is finished. In the following, we go to the learning process in situation 2 auditory channel for communication is not included.

Situation 2: The learning process of the HAA in this situation is very similar to the learning process in situation 1. However, communication between the AE network and the communicator is not needed because auditory channel for communication is not included in this situation. For example, channel $\mathbf{X}$ is a visual channel and the emerged channel $\mathbf{Y}$ is a gustatory channel.

When the HAA receives the activation signals from $N_a^{C_\chi}$ and $N_b^{C_\gamma}$, we also solve (25) and (26) to find the association neurons whose ADs include $N_a^{C_\chi}$ and $N_b^{C_\gamma}$. Assume that the solution sets of (25) and (26) are S^χ and S^γ. Similarly, there are four conditions of S^χ and S^γ.

First, $S^\chi \neq \varnothing$ and $S^\gamma = \varnothing$, which means that the concept neuron $N_b^{C_\gamma}$ in the emerged channel $\mathbf{Y}$ has not yet been connected to any association neurons and the concept neuron $N_a^{C_\chi}$ in the native channel $\mathbf{X}$ has been connected to some association neuron(s) in set S^χ. Then, a new connection between concept neuron $N_b^{C_\gamma}$ and each association neuron in set S^χ will be created to associate concept $N_b^{C_\gamma}$ with concept neuron $N_a^{C_\chi}$ using formulas (29) and (30).

Second, $S^\chi \neq \varnothing$ and $S^\gamma \neq \varnothing$, which means that the concept neurons $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ both have been connected to some association neuron(s). If $S^\chi \cap S^\gamma \neq \varnothing$, i.e., $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ activate some association neuron(s) in common, it implies that AE recognizes the current input pair as an encountered one. Then, the AE network strengthens the activated connections using formula (31). If $S^\chi \cap S^\gamma = \varnothing$, concept neurons $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ are added to the ADs of each association neuron in set S^γ and S^χ with formulas (32) and (33), and the connections between the two concept neurons are initialized using formula (34).

Third, $S^\chi = \varnothing$ and $S^\gamma \neq \varnothing$, which means that $N_a^{C_\chi}$ has not been connected to any association neurons, but $N_b^{C_\gamma}$ has been connected to some association neuron(s). Then, a connection between $N_a^{C_\chi}$ and each association neuron in set S^γ will be created using formula (35). Meanwhile, the ADs of each association neuron in S^γ will be expanded with $N_a^{C_\chi}$ using formula (36).

Fourth, $S^{\chi} = \varnothing$ and $S^{\gamma} = \varnothing$, which means that neither $N_a^{C_\chi}$ nor $N_b^{C_\gamma}$ has been connected to any association neurons. In this case, a new association neuron is initialized to associate $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ using formula (37), and a new connection between $N_a^{C_\chi}$ and $N_b^{C_\gamma}$ through N_{new}^A is created using formula (38).

After that, the learning for the current input pair $(\boldsymbol{x}, \boldsymbol{y})$ finishes. The AE network will go to the next input pair.

Now, the learning algorithm of the HAA in situation 2, i.e., auditory channel for communication is not included, is finished.

As a summary, we give the complete algorithm of the AE network in Algorithm 1.

Algorithm 1 AE

Premise: A native perceptual channel $\mathbf{X}$ and a sudden emerged perceptual channel $\mathbf{Y}$ which receive input data $\boldsymbol{x}$ and $\boldsymbol{y}$.

1: Receive a pair of sample $(\boldsymbol{x}, \boldsymbol{y})$ from the native channel $\mathbf{X}$ and the sudden emerge channel $\mathbf{Y}$.

2: PSA: Execute feature extraction and incremental competitive learning in channel $\mathbf{X}$ and channel $\mathbf{Y}$, formula (3) to formula (8). Perform self-organizing on feature neurons, formula (10) to formula (12).

3: SAA: Execute unimodal concept incremental learning in channel $\mathbf{X}$ and channel $\mathbf{Y}$, formula (15) to formula (22).

4: HAA: Integrate channel $\mathbf{Y}$ with channel $\mathbf{X}$. If auditory channel for communication is included, using procedure **Situation (1)**. If auditory channel for communication is not included, using procedure **Situation (2)**.

5: Waiting for the next input pair and go to step 1.

IV. Experiment

In this section, we design a sensory channel expansion experiment to test the AE network. As shown in Fig. 3, 20 objects are used in the experiment. We give the sensory channel expansion process as shown in the following formula:

$$\underbrace{\mathbf{V} + \mathbf{A}}_{\text{Stage One}} \rightsquigarrow \underbrace{\mathbf{V} + \mathbf{A} + \mathbf{G}}_{\text{Stage Two}} \rightsquigarrow \underbrace{\mathbf{V} + \mathbf{A} + \mathbf{G} + \mathbf{U}}_{\text{Stage Three}}. \tag{39}$$

Formula (39) means that at the first stage, there are two perceptual channels in the AE network: one is a vision channel $\mathbf{V}$, which is a camera, and the other is an auditory channel $\mathbf{A}$, which is a microphone. At this stage, channel $\mathbf{V}$ is treated as the native channel and channel $\mathbf{A}$ is treated as the emerged channel. Pairs of image and Chinese name are used to train the network. At each round of learning, an object is put in front of the camera. Then, we start the audition program and pronounce the Chinese name of the object to the microphone. The audition program records the voice, and at the same time, the vision program captures an image of the object. When the pronunciation is finished, the audition program is closed. One round of learning is finished, and we go to the next round. There are 22 names in total and we let AEN learn each object from eight different view angles when pronouncing the object's name. Thus, this stage includes 176 rounds of learning with 176 different pairs of images and words.

Fig. 3. Examples of the objects used in the experiments.

After the first stage, the AE network enters the second stage, where a gustatory channel $\mathbf{G}$ emerges. The network integrates channel $\mathbf{G}$ into itself. We give pairs of image and gustatory data to the network. The learning procedure of this stage is similar to that of the first stage. At each round of learning, an object is put in front of the camera and vision program captures the image of the object. Meanwhile, we input the gustatory data of the object to the network. After that, one round of learning is finished, and we go to the next round. Similarly, we let AEN learn each object from eight different view angles accompanying with different gustatory samples. Thus, the second stage also includes 176 rounds of learning.

After the second stage, the AE network enters the third stage, where an ultrasonic channel $\mathbf{U}$ emerges which receives ultrasonic sounds. Pairs of image and ultrasonic words are used to train the network. The learning procedure is also similar to that of the first stage. At each round of learning, an object is put in front of the camera and vision program captures the image of the object. Meanwhile, the ultrasonic name of the object is input to the network. Then, one round of learning is finished and we go to the next round. Similar to the previous stages, 176 rounds of learning with 176 different pairs of images and ultrasonic words are conducted. After the third stage, one time of the sensory channel expansion experiment finishes.

Visual features include the CH of the object and the NFD of the object's boundary. To extract the object's boundary, a Gaussian filter with $[Hsize = 15, \sigma = 9]$ is used to smooth the image first. Then, the image is converted into a binary image with a gray threshold of 0.7. After the coordinates of the object's boundary are obtained from the binary image, which is assumed to be $b_i = (x_i, y_i)$, $0 \le i \le n-1$, the fast Fourier transform is applied to the coordinates to get the Fourier descriptors of the object's boundary, which is assumed to be $\boldsymbol{d} = (d_0, d_1, d_2, \ldots, d_{n-1})$. Finally, the NFD is calculated by $d_i = \|d_i\|/\|d_1\|$, $1 \le i \le n-1$ and vector $(d_2, d_3, \ldots, d_{25})$ is chosen for the final shape feature. For the color feature, we first get the area of the object in the image, which is the area surrounded by the coordinates $b_i = (x_i, y_i)$, $0 \le i \le n-1$. Then, CH is applied to this area, where the container size is set to 0.05.

Auditory features are the MFCCs of the syllables contained in the input voice. First, all syllables in the input voice wave need to be extracted. To do this, the voice wave is first filtered by a high-pass filter, where the system function is $H(z) = 1 - \mu z^{-1}$, and we set μ as 0.9375. Following this, the amplitude of the voice wave is normalized. Next, frame blocking is executed, and the frame size and frame shift are set to 256 and 128, respectively. Assuming that m frames are gained. Then, the short-time energy and short-time zero-crossing rate of each frame are calculated. The frame whose short-time energy is larger than 0.5 or short-time zero-crossing rate is larger than 100 will be marked as a candidate frame. The consecutive candidate frames are spliced together as the candidate syllables. Meanwhile, the adjacent two candidate syllables with gap no more than one frame length will be merged together. Also, candidate syllables whose length is less than two frame lengths will be marked as noise and removed. Finally, the syllables contained in the voice wave are obtained and the MFCCs of each syllable are extracted.

Gustatory samples are artificially designed. They are 6-D vectors that have a format of (sweet, sour, hot, salt, bitter, and umami). The value of each attribute is in the range [0, 1]. For example, the gustatory samples of the object apple are designed as follows, the values of sweet and sour attribute are uniformly distributed in the range [0.5, 0.6] and [0, 0.1], respectively, and other attributes are 0.

For ultrasonic data, we generate 35 kinds of sin waves every 3 kHz with an amplitude of 0.5. The frequency range is from 20 to 122 kHz, and the sampling frequency is 10 MHz. Random noise in the interval of [0 0.001] is added to the sin waves. We use $\boldsymbol{s}_i, 1 \leq i \leq 35$, to denote them and each $\boldsymbol{s}_i$ contains 100k sampling points. We use these waves to generate ultrasonic words, for example, we name apple as "[$\boldsymbol{s}_1$ $\boldsymbol{bg}$ $\boldsymbol{s}_2$]," where $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are the samples of 20 and 23 kHz sin waves, and $\boldsymbol{bg}$ is a background noise wave that links the two sin waves. The ultrasonic words contain one to three sin waves. Features in this channel are the real cepstrum of each ultrasonic syllable. Similarly, short-time energy and short-time zero-crossing rate are used to extract the ultrasonic syllables contained in the input ultrasonic word. The thresholds of the short-time energy and short-time zero crossing are 0.001 and 100, respectively. To obtain the real cepstrum of each syllable, the frame size and frameshift are set to 4096 and 4096, respectively.

As mentioned in Section III-B, parameter θ of the feature neuron is set according to the characteristic of particular features. In practice, for each input channel, we can collect a small data set beforehand and estimate a suitable value for θ using some statistical methods based on the interclass distances and intraclass distances of the features extracted from the data set, e.g., a factor times the minimum value of the intraclass distances. Using such a method, parameter θ of the feature neuron in the experiment is set as follows. In the visual channel, θ for shape and color neuron is set to 1/6 times the norm of the neuron's weights. In the audible channel, θ for the audible syllable neuron is set to 200. In the ultrasonic channel, θ for the ultrasonic syllable neuron is set to 9. In the gustatory channel, θ for the basic taste neuron is set to 0.015.

We conduct the experiment in two environments: 1) the closed environment and 2) the open-ended environment (for the stability–plasticity dilemma [10]). In a closed environment, object is randomly chosen from the 20 objects in each round of learning. In an open-ended environment, we first let the method learn ten objects. In each round of learning, an object is randomly chosen. After that, we give the method the remaining ten new objects. Similarly, in each round of learning, the object is also randomly chosen. The open-ended learning environment is designed to test whether the AE network can learn new objects without forgetting previously learned ones. To test the stability of the method, we conduct the experiment 20 times for both the closed environment and the open-ended learning environment.

We also compare the AE network with some classical and state-of-the-art methods, including IKR1 system [14], GHF-ART [12], and PEN [34]. Among the three comparison methods, the IKR1 system and GHF-ART do not focus on integrating suddenly emerged channels, so they do not have channel extensibility. IKR1 system is a vision–audition bimodal model and cannot own three or more channels, so we use visual, auditory, and ultrasonic data to train two models for each method: one is a vision–audition (audible) model and the other one is a vision–ultrasonic model. The GHF-ART cannot deal with data with different dimensions, so we train a vision–gustation–ultrasonic model (the dimensions of the auditory samples are different). PEN can adapt to the emergence of new dimensional sensory inputs. However, as mentioned in Section II, it cannot deal with samples with different dimensions. Thus, we can say that it has limited channel extensibility. During the training of the PEN, we first give PEN the visual channel, then the gustatory channel, and finally the ultrasonic channel. We also run the three methods in both the closed environment and the open-ended environment.

A. Learning Results

For the 20 times experiments in the closed environment and open-ended environment, after stage 1, the AE network obtains 71–76 shape feature neurons, 47–49 color feature neurons, 71–76 visual concept neurons, 231–236 Chinese syllable neurons, 138–144 auditory concept neurons (Chinese words), and 64–66 association neurons. After stage 2, the AE network obtains 76–81 shape feature neurons, 47–49 color feature neurons, 76–81 visual concept neurons, 33–35 sweet feature neurons, 27 and 28 sour feature neurons, 4 and 5 salt feature neurons, 4 and 5 bitter feature neurons, 5 and 6 umami feature neurons, 12–14 hot feature neurons, 119123 taste concept neurons, and 67–71 association neurons, and the number of other types of neurons does not change. After stage 3, the AE network obtains 78–84 shape feature neurons, 78–84 visual concept neurons, 170–177 ultrasonic syllable neurons, 291–296 auditory concept neurons (Chinese words and ultrasonic words), and 696972 association neurons, and the number of other types of neurons does not change.

Fig. 4 shows some examples of the learning results at different stages. It can be found that the AE network can integrate the learned concepts in the emerged perceptual channel with the learned concepts in the native perceptual correctly; in other words, the newly emerged perceptual channel is effectively integrated into the existing network in an online way.

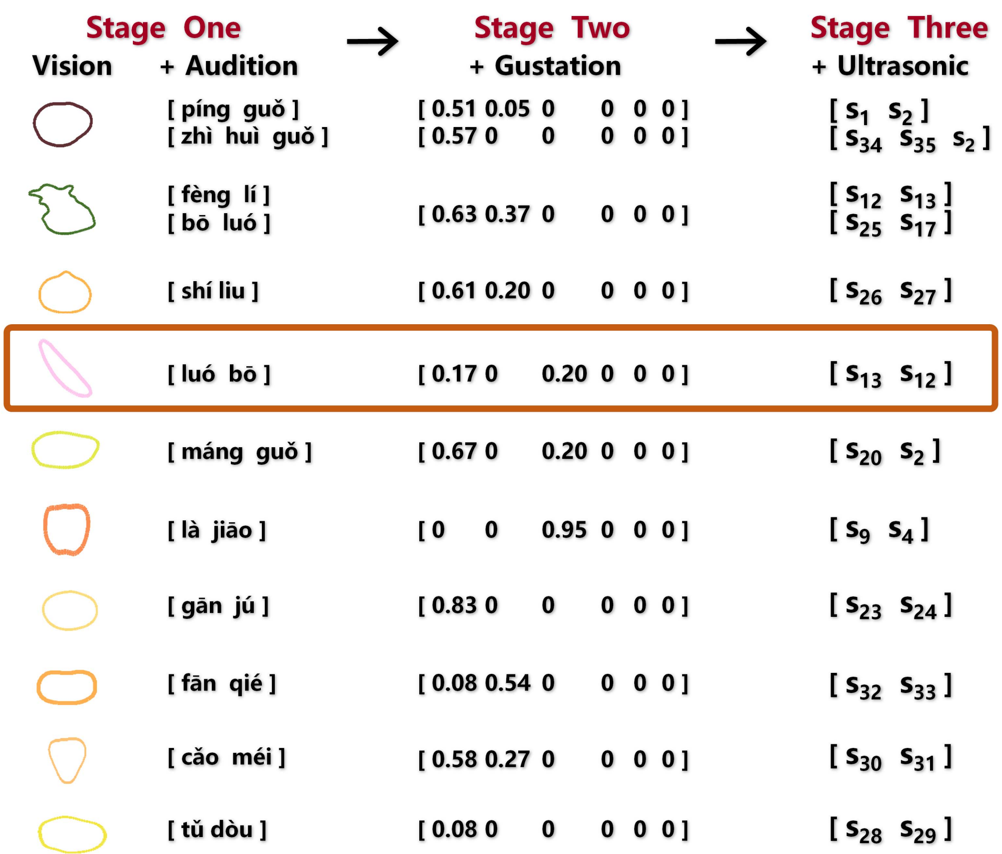


Fig. 4. Examples of the learning results at different stages. As shown in the box, the learned concepts in each line are the maximum response concepts of some neurons, and these neurons are connected together in the neural network. A detailed structure can be found in Figs. 5–7. At stage 1, the AE network has a visual channel and an auditory channel that receive the image and the Chinese name of an object. At this stage, we treat the auditory channel as the emerged channel. It can be found that the network learns the view and the name of each object and associates them correctly. At stage 2, a gustatory channel emerges, which receives the taste data of an object. The AE network associates the taste of each object with its view and name correctly. After this stage, it can be said that the AE network comes to a new world with a concept of taste. At stage 3, an ultrasonic channel emerges, which receives the object's name in the ultrasonic field. Again, the AE network associates the ultrasonic name of each object with its view, Chinese name, and taste correctly. The AE network then comes to a new world with a concept of ultrasonic sounds. In brief, the AE network can integrate novel emerged perceptual channels effectively in an online way.

For a more detailed observation, Fig. 5 shows the change of the network structure of the concept apple after new perceptual channels emerged in the network. It can be found that when a new perceptual channel emerged in the network, the association neurons grow new connections to connect the concept neurons in the new channel, and these connections build a bridge for the communication between the new perceptual channel and the native network. The new channel is effectively integrated into the native network through such connections. The response modality of the association neurons is expanded, which is conformed to the experimental finding in [2] that neurons responding to IR light inputs maintain their original sensory modality. Note that the probability of the connection between visual concept neuron and each auditory concept neuron through the association neuron is 0.5. They are gained by the normalization of the activity parameter ρ of these connections in formula (23), and the value of the probability means that the two tastes appear at the same probability in this case. Figs. 6 and 7 show two similar results.

Examples of self-organizing of the feature neurons in the PSA are shown in Fig. 8. In Fig. 8(a), the coordinates of the points are obtained by applying the principal component analysis on the weights of the shape neurons. The features are organized by the Euclidean distance and similar shapes under the Euclidean distance metric are connected together. In Fig. 8(b), the coordinate of each point is the RGB value of the color, which is transformed from the CH. The features are organized by the Euclidean distance. It can be found that similar colors under the Euclidean distance metric are connected together. Fig. 8(c) shows the Chinese syllables that connected to the syllable [ji]. The coordinates of the points are obtained by applying the principal component analysis on the weights of the Chinese syllable neurons. The features are organized by the dynamic time warping distance. It can be found that the pronunciations of these syllables are very similar to [ji]. Fig. 8(d) shows a similar phenomenon.

B. Testing Results

1) Recall Experiments: To test the learning results, we use one kind of sensory input to recall the other three kinds of sensory output. Similar to the learning procedure, for each modality, we do 176 rounds of recalls. For example, when we use audible word input to recall vision, gustation, and

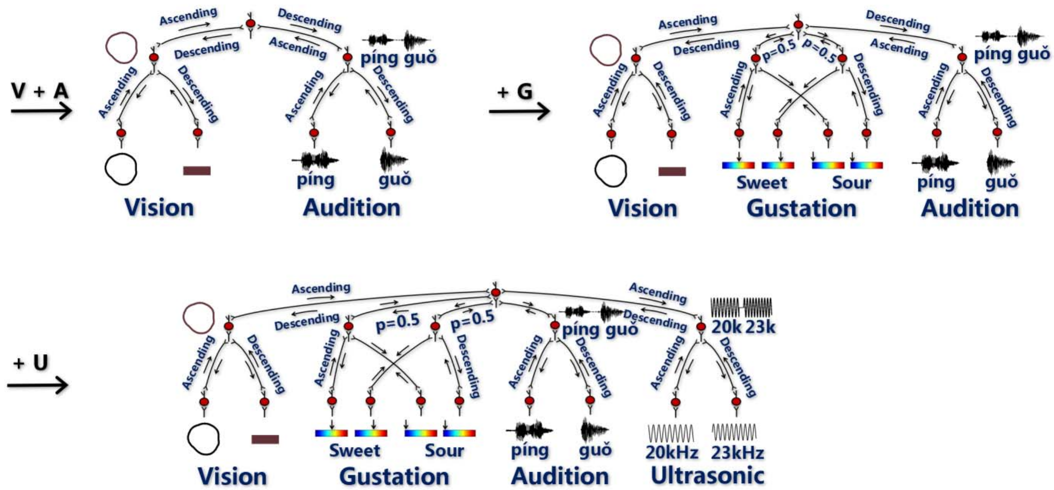


Fig. 5. Change of the network structure of the concept apple after new perceptual channel emerged in the network. Icons next to the neurons represent the objects to which the neurons maximally respond.

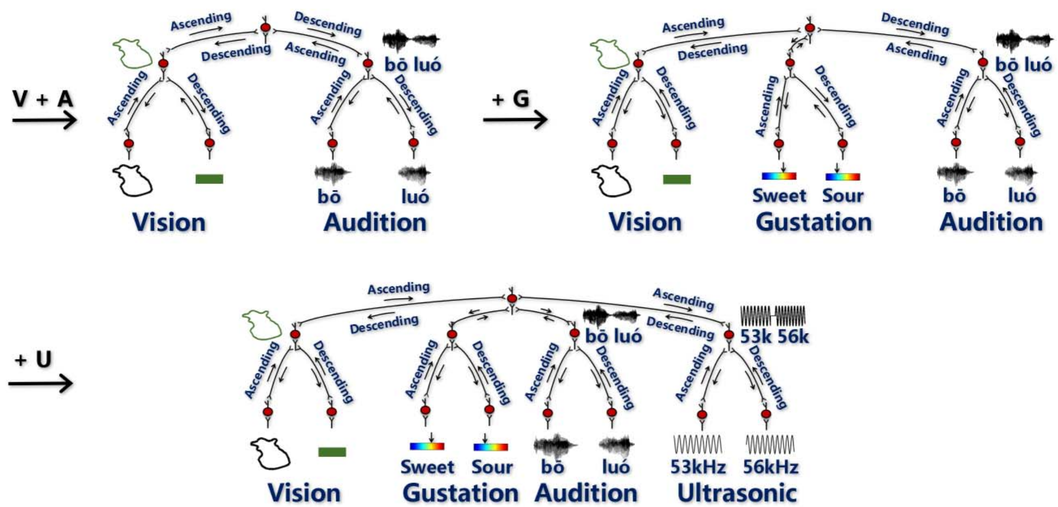


Fig. 6. Change of the network structure of the concept pineapple after a new perceptual channel emerged in the network. Icons next to the neurons represent the objects to which the neurons maximally respond.

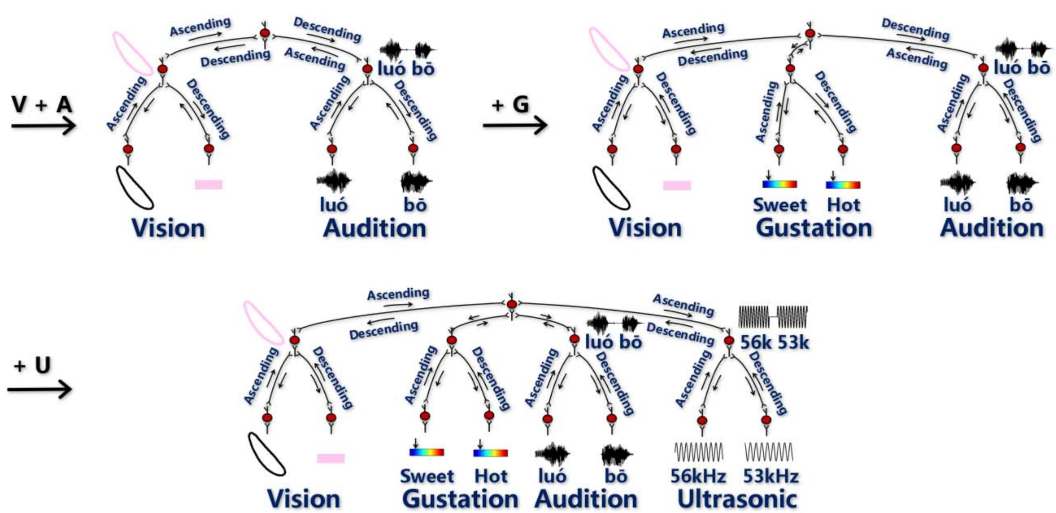


Fig. 7. Change of the network structure of the concept turnip after a new perceptual channel emerged in the network. Icons next to the neurons represent the objects to which the neurons maximally respond.

ultrasonic word output, for each audible word, we pronounce it eight times to do eight recalls.

Because IKR1 learns a vision–audition (audible) model and a vision–ultrasonic model, we only do vision recalls audition

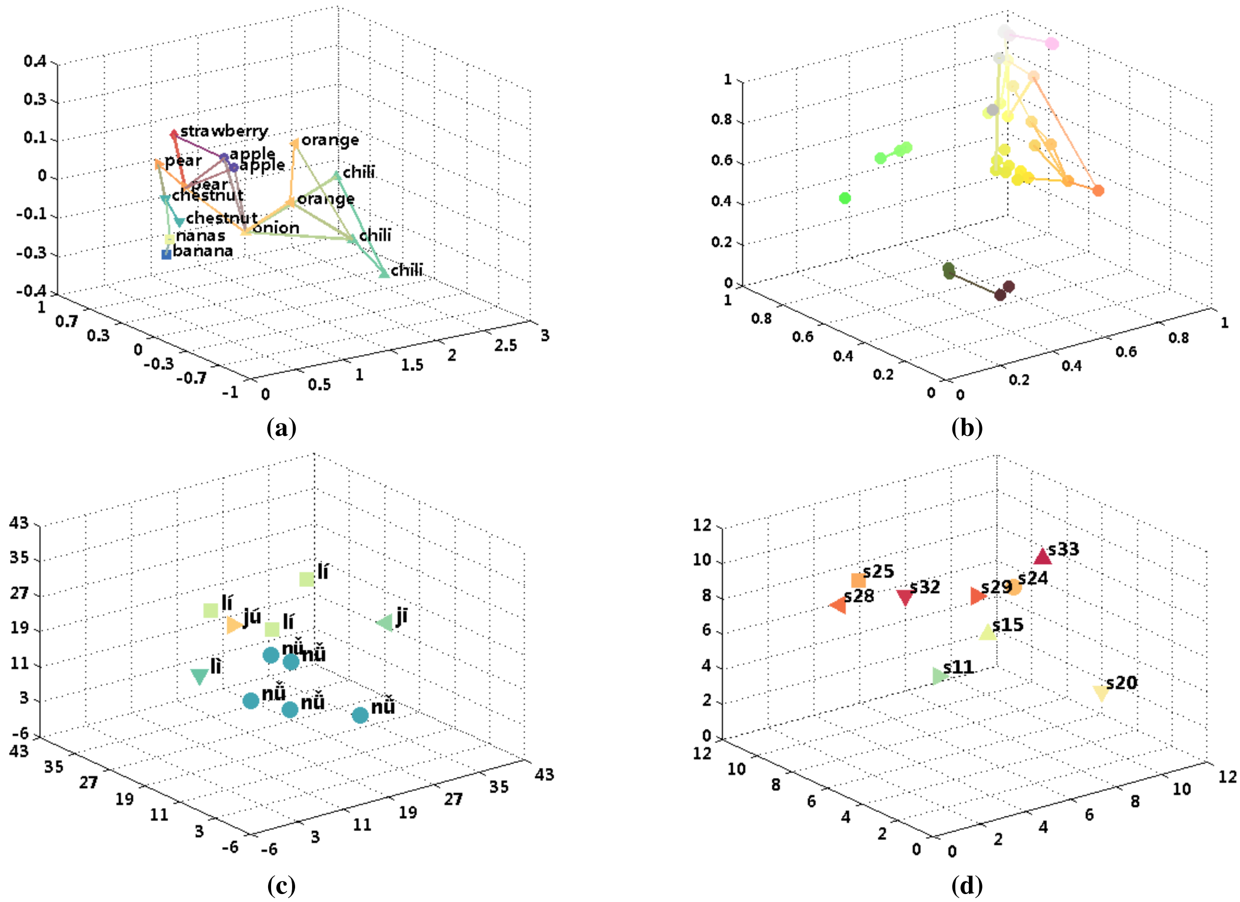


Fig. 8. Examples of the self-organizing among feature neurons. (a) Self-organizing of 15 shape feature neurons. (b) Self-organizing of the color feature neurons. (c) Chinese syllables that connected to the syllable [ji]. (d) Ultrasonic syllables that connected to the ultrasonic syllable s_{33}.

and ultrasound, audition recalls vision and ultrasound, and ultrasound recalls vision and audition in the testing experiments. GHF-ART and PEN learn a vision–gustation–ultrasonic model, so we do vision recalls others, gustation recalls others, and ultrasound recalls others. Testing is conducted after each time of learning experiment, which means that 20 testing experiments are done in both the closed environment and the open-ended environment. To test the learned structure of the network, incremental learning is disabled during testing.

Tables I and II show the statistical results of the testing experiments. Significant decrease of the accuracy in one environment is marked by ↓ in the tables. It can be found that the AE network gets a higher accuracy than the other methods. We also find that the accuracy of IKR1 is unstable, which has a gap of about 5%–10% between the two learning environments. This is due to the IKR1 that usually forgets previously learned objects when learning the latter ten new objects. Thus, the drop of the accuracy is mainly due to the recognition of the previous ten objects. The accuracy of the AE network is stable in both environments.

More importantly, the AE network can integrate new perceptual channels in an online way, which means that a new channel can be plugged in whenever it is needed. However, computational models without such an ability have no such convenience, e.g., for the bimodal method IKR1 system, we need to train two models when we use visual, auditory, and ultrasonic data. As a consequence, it is inconvenient to find the relationship between the audible name and the ultrasonic name of an object. One possible way is to use the first model to find the image of an object by the input audible name and then use the second model to find the ultrasonic name of the image. When we use the AE network, the audible name and the ultrasonic name of the object are linked together automatically through the association neuron, whereas the audible data and the ultrasonic data were not given simultaneously during learning. For the multimodal method GHF-ART, if we want to add a new modality, a new model should be trained from scratch which is also inconvenient.

When the emerged channel comes, PEN needs data from all native channels to integrate the emerged channel. Because PEN treats multimodal data as a unit and adds the weight vector learned from the emerged channel to the tail of the weight vector of a native neuron. This is a very inconvenient and unnatural way. The proposed AE network uses association neurons to grow new connections to connect the emerged channel. Thus, the AE network does not need all native channels when doing integration, which is a more natural way.

2) Channel Expansion in Different Orders: In this section, we test AEN in the following different channel expansion orders. In order I, we first give AEN a visual channel and an ultrasonic channel and feed pairs of visual and ultrasonic sample to AEN. Then, we add a gustatory channel and feed AEN with pairs of gustatory and ultrasonic sample. Finally, we add an auditory channel to AEN and use pairs

TABLE I

STATISTICAL RESULTS OF THE TESTING EXPERIMENTS (MEAN+STD) IN THE CLOSED ENVIRONMENT. N/A MEANS THAT THE TESTING CONDITION IS NOT APPLICABLE. THE BEST RESULTS ARE IN BOLD. V: VISION. A: AUDITION. G: GUSTATION. U: ULTRASOUND

Extensibility	Method	V recalls Other	A recalls Other	G recalls Other	U recalls Other
No	IKR1 [14]	73.55±3.24% ↓	60.58±3.09% ↓	N/A	65.63±2.98% ↓
	GHF-ART [12]	82.33±2.53%	N/A	84.71±2.09%	80.05±2.17%
Limited	PEN [34]	81.27± 2.65%	N/A	83.06±2.57%	78.94±2.85%
Yes	AE	**84.75±2.61%**	**82.11±2.73%**	**89.54±2.76%**	**83.57±2.59%**

TABLE II

STATISTICAL RESULTS OF THE TESTING EXPERIMENTS (MEAN+STD) IN THE OPEN-ENDED ENVIRONMENT. N/A MEANS THAT THE TESTING CONDITION IS NOT APPLICABLE. THE BEST RESULTS ARE IN BOLD. V: VISION. A: AUDITION. G: GUSTATION. U: ULTRASOUND

Extensibility	Method	V recalls Other	A recalls Other	G recalls Other	U recalls Other
No	IKR1 [14]	78.21±3.13%	71.35±3.32%	N/A	74.42±2.75%
	GHF-ART [12]	83.15±2.67%	N/A	86.45±2.23%	81.65±2.21%
Limited	PEN [34]	82.04± 2.74%	N/A	83.78±2.67%	79.36±2.91%
Yes	AE	**85.07±2.58%**	**82.39±2.89%**	**89.76±2.94%**	**83.79±2.47%**

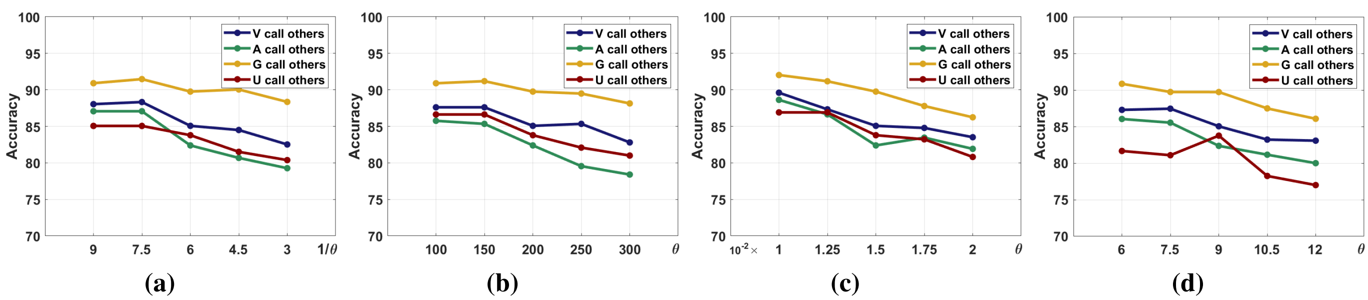


Fig. 9. Parameter influence. (a) Vary θ of the visual channel from 1/9 to 1/3. (b) Vary θ of the audible channel from 100 to 300. (c) Vary θ of the gustatory channel from 0.01 to 0.02. (d) Vary θ of the ultrasonic channel from 6 to 12. While varying one channel's θ value, the θ values of other channels are fixed at 1/6 for the visual channel, 200 for the audible channel, 9 for the ultrasonic channel, and 0.015 for the gustatory channel.

TABLE III

ACCURACY (MEAN+STD) IN DIFFERENT CHANNEL EXPANSION ORDERS

Order	V recalls Other	A recalls Other	G recalls Other	U recalls Other
I	84.59±2.33%	81.75±2.79%	89.93±1.75%	83.29±2.61%
II	84.68±3.24%	82.02±3.01%	89.34±2.46%	83.31±2.71%
III	85.18±2.42%	82.27±3.29%	90.23±2.35%	84.18±3.40%

of visual and audible sample. In order II, we first feed AEN with pairs of visual and ultrasonic sample, then visual and audible sample, and finally gustatory and audible sample. In order III, we first feed AEN with pairs of gustatory and audible sample, then visual and gustatory sample, and finally ultrasonic and gustatory sample. Twenty times of experiments are conducted in each order. Table III shows that the results are stable.

3) Parameter Sensitivity: In this section, we test the influence of parameter θ of the feature neurons. Fig. 9 shows the results. When θ is set in a reasonable range, the influence on the accuracy is not strong. It can also be found in the figure that AEN prefers to small θ in general. However, small θ value means that the neuron is hard to be activated. Thus, in order to make the system have a stronger generalization ability, it is better not to set θ too small.

V. CONCLUSION

In this article, a biologically plausible neural network model AE that aims to integrate sudden emerged perceptual channels is proposed. By creating connections between association neurons and concept neurons in the emerged channel, the native neural network is able to integrate the emerged channel directly. The creation of the new connections follows the Hebb rule, i.e., neurons that fire together, wire together [35], which also conforms with the experiment that new spines are formed in the rat's brain by novel experience [36]. We also introduce many physiological research results to the model, e.g., for different types of neurons found in many physiological experiments, including the feature neurons [37], concept neurons [38], and association neurons [39], we design different computational models.

Experimental results show that the AE network can effectively integrate sudden emerged perceptual channels in an online way. This implies that the AE network can handle many potential practical problems. For example, if we want to install new sensors to a robot, with the AE algorithm, we do not need to retrain the robot off-line from scratch, and the new sensors will be automatically integrated with the robot platform in an online way.

REFERENCES

[1] G. H. Jacobs, G. A. Williams, H. Cahill, and J. Nathans, "Emergence of novel color vision in mice engineered to express a human cone photopigment," *Science*, vol. 315, no. 5819, pp. 1723–1725, Mar. 2007.

[2] E. E. Thomson, R. Carra, and M. A. L. Nicolelis, "Perceiving invisible light through a somatosensory cortical prosthesis," *Nature Commun.*, vol. 4, no. 1, pp. 66–78, Jun. 2013.

[3] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, no. 1, pp. 59–69, 1982.

[4] T. Martinetz and K. Schulten, "Topology representing networks," *Neural Netw.*, vol. 7, no. 3, pp. 507–522, Jan. 1994.

[5] D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory*. New York, NY, USA: Wiley, 1949.

[6] B. Fritzke, "A growing neural gas network learns topologies," in *Proc. Adv. Neural Inf. Process. Syst.*, Denver, CO, USA, Nov. 1994, pp. 625–632.

[7] J. Faigl and G. A. Hollinger, "Autonomous data collection using a self-organizing map," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1703–1715, May 2018.

[8] M. B. Gorzałczany and F. Rudziński, "Generalized self-organizing maps for automatic determination of the number of clusters and their multiprototypes in cluster analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 2833–2845, Jul. 2018.

[9] W. C. Abraham and A. Robins, "Memory retention—The synaptic stability versus plasticity dilemma," *Trends Neurosci.*, vol. 28, no. 2, pp. 73–78, Feb. 2005.

[10] G. A. Carpenter and S. Grossberg, "The art of adaptive pattern recognition by selforganizing neural network," *Comput.*, vol. 21, no. 3, pp. 77–88, 1988.

[11] S. Grossberg, "Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world," *Neural Netw.*, vol. 37, pp. 1–47, Jan. 2013.

[12] L. Meng, A.-H. Tan, and D. Xu, "Semi-supervised heterogeneous fusion for multimedia data co-clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 9, pp. 2293–2306, Sep. 2014.

[13] S. Furao and O. Hasegawa, "An incremental network for on-line unsupervised classification and topology learning," *Neural Netw.*, vol. 19, no. 1, pp. 90–106, Jan. 2006.

[14] X. He, R. Kojima, and O. Hasegawa, "Developmental word grounding through a growing neural network with a humanoid robot," *IEEE Trans. Syst., Man Cybern. B, Cybern.*, vol. 37, no. 2, pp. 451–462, Apr. 2007.

[15] Y. Nakamura and O. Hasegawa, "Nonparametric density estimation based on self-organizing incremental neural network for large noisy data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 1, pp. 8–17, Jan. 2017.

[16] A. Kumar, S. Rotter, and A. Aertsen, "Spiking activity propagation in neuronal networks: Reconciling different perspectives on neural coding," *Nature Rev. Neurosci.*, vol. 11, no. 9, pp. 615–627, Sep. 2010.

[17] G. Auda and M. Kamel, "Modular neural networks: A survey," *Int. J. Neural Syst.*, vol. 9, no. 2, pp. 129–151, Apr. 1999.

[18] P. G. Schyns, "A modular neural network model of concept acquisition," *Cognit. Sci.*, vol. 15, no. 4, pp. 461–508, Oct. 1991.

[19] J. A. Anderson, J. W. Silverstein, S. A. Ritz, and R. S. Jones, "Distinctive features, categorical perception, and probability learning: Some applications of a neural model.," *Psychol. Rev.*, vol. 84, no. 5, pp. 413–451, Sep. 1977.

[20] K. Yamauchi, M. Oota, and N. Ishii, "A self-supervised learning system for pattern recognition by sensory integration," *Neural Netw.*, vol. 12, no. 10, pp. 1347–1358, Dec. 1999.

[21] T. Jantvik, L. Gustafsson, and A. P. Papliński, "A self-organized artificial neural network architecture for sensory integration with applications to letter-phoneme integration," *Neural Comput.*, vol. 23, no. 8, pp. 2101–2139, Aug. 2011.

[22] A. P. Papliński and W. M. Mount, "Bimodal incremental self-organizing network (BiSON) with application to learning Chinese characters," in *Proc. Int. Conf. Neural Inf. Process.*, Daegu, South Korea, Nov. 2013, pp. 121–128.

[23] S. Zeki and S. Shipp, "The functional logic of cortical connections," *Nature*, vol. 335, no. 6188, pp. 311–317, Sep. 1988.

[24] K. Fukushima, "Neocognitron: A hierarchical neural network capable of visual pattern recognition," *Neural Netw.*, vol. 1, no. 2, pp. 119–130, Jan. 1988.

[25] J. Rajapakse and R. Acharya, "Multi sensor data fusion within hierarchical neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, San Diego, CA, USA, Jun. 1990, pp. 17–22.

[26] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self organizing neural pattern recognition machine," *Comput. Vis., Graph. Image Process*, vol. 37, no. 1, pp. 54–115, 1987.

[27] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Carson City, NV, USA, 2012, pp. 1097–1105.

[29] E. J. Palomo and E. López-Rubio, "The growing hierarchical neural gas self-organizing neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 9, pp. 2000–2009, Sep. 2017.

[30] Y.-L. Xing, F.-R. Shen, J.-X. Zhao, J.-X. Pan, and A.-H. Tan, "Perception coordination network: A framework for online multi-modal concept acquisition and binding," in *Proc. AAAI*, New Orleans, LA, USA, Feb. 2018, pp. 4318–4325.

[31] Y.-L. Xing, X.-F. Shi, F.-R. Shen, J.-X. Zhao, J.-X. Pan, and A.-H. Tan, "Perception coordination network: A neuro framework for multimodal concept acquisition and binding," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 4, pp. 1104–1118, Apr. 2019.

[32] Y.-L. Xing, F.-R. Shen, and J.-X. Zhao, "A perception evolution network for unsupervised fast incremental learning," in *Proc. Int. Joint Conf. Neural Netw.*, Dallas, TX, USA, Aug. 2013, pp. 297–304.

[33] Y.-L. Xing, F.-R. Shen, and J.-X. Zhao, "Perception evolution network adapting to the emergence of new sensory receptor," in *Proc. Int. Conf. Artif. Intell.*, Buenos Aires, Argentina, Jul. 2015, pp. 3967–3973.

[34] Y. Xing, F. Shen, and J. Zhao, "Perception evolution network based on cognition deepening model–adapting to the emergence of new sensory receptor," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 3, pp. 607–620, Mar. 2016.

[35] C. J. Shatz, "The developing brain," *Sci. Amer.*, vol. 267, no. 3, pp. 60–67, 1992.

[36] G. Yang, F. Pan, and W.-B. Gan, "Stably maintained dendritic spines are associated with lifelong memories," *Nature*, vol. 462, no. 7275, pp. 920–924, Dec. 2009.

[37] N. Mesgarani, C. Cheung, K. Johnson, and E. F. Chang, "Phonetic feature encoding in human superior temporal gyrus," *Science*, vol. 343, no. 6174, pp. 1006–1010, Feb. 2014.

[38] A. M. Chan *et al.*, "Speech-specific tuning of neurons in human superior temporal gyrus," *Cerebral Cortex*, vol. 24, no. 10, pp. 2679–2693, Oct. 2014.

[39] R. Q. Quiroga, "Concept cells: The building blocks of declarative memory functions," *Nature Rev. Neurosci.*, vol. 13, no. 8, pp. 587–597, Aug. 2012.

You-Lu Xing received the B.S. degree in computer science and technology from the East China University of Science and Technology, Shanghai, China, in 2009, and the M.S. degree in software engineering and the Ph.D. degree in computer software and theory from Nanjing University, Nanjing, China, in 2011 and 2016, respectively.

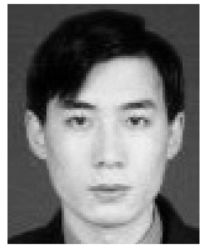
His current research interests include artificial intelligence, robotics, and physiology.

Hui Sun received the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2014.

He is currently an Assistant Professor of computer science with Anhui University, Hefei, China. His research interests include intelligent computer systems and distributed artificial intelligence based on edge computing.

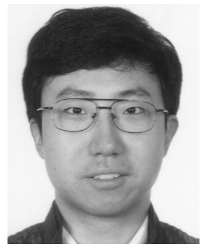
Gui-Huan Feng received the B.S. degree and the Ph.D. degree in computer science and technology from Nanjing University, Nanjing, China, in 2003 and 2009, respectively.

In 2003, she joined the State Key Laboratory for Novel Software Technology, Nanjing University. She is currently an Associate Professor with the Software Institute, Nanjing University. Her research interests include intelligent human–computer interaction and pattern recognition.



Fu-Rao Shen (Member, IEEE) received the B.Sc. and M.Sc. degrees in mathematics from Nanjing University, Nanjing, China, in 1995 and 1998, respectively, and the Ph.D. degree from the Tokyo Institute of Technology, Tokyo, Japan, in 2006.

He is currently a Full Professor of computer science and technology with Nanjing University. His current research interests include neural computing and robotic intelligence.



Jian Zhao (Senior Member, IEEE) received the B.S. degree from Nanjing University, Nanjing, China, in 2001, the M.Sc. degree from the Hamburg University of Technology, Hamburg, Germany, in 2004, and the Dr.Sc. degree in electrical engineering from the Swiss Federal Institute of Technology (ETH) Zürich, Zürich, Switzerland, in 2010.

From 2010 to 2015, he was with the Institute for Infocomm Research, A*STAR, Singapore. He is currently an Associate Professor with the School of Electronic Science and Engineering, Nanjing University. His research interests include deep neural networks, mathematical optimization, and wireless communications.

Dr. Zhao was honored with the Dengfeng Scholars Program of Nanjing University in 2015, the IEEE Globecom 2008 Best Paper Award, and the 2009 Chinese Government Award for Outstanding Self-Financed Students Abroad.