



Sensing-to-Learn and Learning-to-Sense: Principles for Designing Neuromorphic Sensors

Shantanu Chakrabarty, Baranidharan Raman,
and Chetan Singh Thakur

Contents

1	Introduction	3
2	Case Study I: Neuromorphic Audition	5
2.1	Biological Cochlea Model: Cascade of Asymmetric Resonators with Fast-Acting Compression (CAR-FAC)	6
2.2	Jump-Resonance-Based Auditory Filter Banks	11
2.3	Biological Feature Extraction (Spectro-temporal Receptive Field (STRF)-Based Approach)	14
2.4	Statistical Learning-Based Feature Extraction	15
3	Case Study II: Neuromorphic Olfaction	20
3.1	Basic Building Blocks of the Biological Olfactory System	20
4	Biological Signal Processing Principles	26
4.1	Noise-Shaping in Integrate and Fire Spiking Neural Networks	27
4.2	Adaptive Analog-to-Digital Converter Based on Spiking Neural Networks	29
4.3	Signal De-Correlation and Network Stability	31
4.4	Neuromorphic Olfaction: Dimensionality Reduction	33
4.5	Neuromorphic Olfaction: Gain Control	34
4.6	Neuromorphic Olfaction: Contrast Enhancement	35
4.7	Neuromorphic Olfaction: Mixture Segmentation and Background Suppression	36
4.8	Neuromorphic Olfaction: De-correlation of Signals	38
5	Discussion and Future Direction	39
	References	41

S. Chakrabarty · B. Raman
Washington University in St. Louis, St. Louis, MO, USA
e-mail: shantanu@wustl.edu; barani@wustl.edu

C. S. Thakur (✉)
Indian Institute of Science, Bengaluru, India
e-mail: csthakur@iisc.ac.in

Abstract

Neurobiological systems have evolved over a billion years and serve as a good template for some text engineers to mimic when designing intelligent sensors and systems. For instance, neurobiological systems exploit noise and system non-linearity as a computational aid to push the limits of performance and energy efficiency. In contrast, in man-made technologies, these artifacts are generally considered to be a nuisance. This chapter's focus is on the neuromorphic concept of “sensing-to-learn” and “learning-to-sense,” which are grounded in key neuromorphic adaptation principles based on noise exploitation and non-linear sensory processing techniques. “Noise shaping” and “jump-resonance” are two techniques that can extract salient sensing cues by exploring the synergy between noise and system non-linearity. We illustrate these concepts in the context of auditory and olfaction pathways, and we argue how these principles can be used to design the next generation of neuromorphic sensory interfaces.

Keywords

Neuromorphic audition · Neuromorphic olfaction · Biological signal processing

List of Acronyms

DVS	Dynamic vision sensor
ATIS	Asynchronous time imaging sensor
BM	Basilar membrane
IHC	Inner hair cells
CF	Characteristic frequency
CAR	Cascaded asymmetric resonator
CAR-FAC	Cascade of asymmetric resonators with fast-acting compression
STRF	Spectro-temporal receptive field
OHC	Outer hair cells
DIHC	Digital inner hair cell
DOH	Digital outer hair cell
AGC	Automatic gain control
EER	Equal error rate
PD	Probability of detection
SVM	Support vector machine
SNR	Signal-to-noise ratio
AC	Association connection

1 Introduction

For ages, biology has served as an inspiration to scientists and engineers. However, it is only in the last decade that advancements in micro- and nanofabrication technology have reached the point where “truly” neuromorphic sensors and systems can be designed that match the “raw” sensing and computational capabilities observed in biology. For instance, more than 100 silicon transistors can be packed inside a mammalian cell, which is $10\text{ }\mu\text{m}$ in diameter. It is now possible to fabricate piezoelectric nano-fibers whose dimensions are comparable and smaller than a cricket’s or a spider’s mechanoreceptors. Furthermore, the raw computational power of today’s processor is comparable to that of the human brain. Despite these remarkable technological advances, the performance achieved by specialized biological sensing systems makes even the most advanced man-made systems of today look crude and primitive. To understand how biological systems can achieve such a performance, it is essential to acknowledge that these systems have evolved in an evolutionary environment, where energy resources were scarce. The threat of predators was omnipresent. These constraints have led to remarkable designs of nature seeking to operate at fundamental limits of energy dissipation and performance. For example, the filiform hairs in crickets operate at fundamental limits of noise, the power dissipation of a mammalian cell is less than pico-watt, and the texture of shark’s skin is optimized to reduce bacterial adhesion and growth [1, 2, 3, 4]. To mimic these remarkable abilities using man-made technologies, it is important to understand what new sensing principles are implemented by biology. Specifically, neuromorphic sensing principles aim to understand the interplay between the biological sensors and transducers and the sensory signal processing implemented by the neurobiological interfaces.

It is hypothesized that neurobiological interfaces achieve this performance by their inherent ability to adapt and learn not only at a cognitive level but also at the level of the sensor. Learning involves adaptation and altering the parameters over a slow time-scale during the lifetime of an organism or even altering the topology of the sensor at a much slower rate during the evolution process. This adaptation makes the system efficient in sensing the signals of interest without dissipating too much energy and without requiring precision computing or sensing devices. In this regard, most of the sensory processing in biology is inherent “analog,” and efficiency arises out of the exploitation of computing and sensing primitives inherent in the substrate’s physics, like biochemical diffusion or feedback regulation. Also, unlike man-made sensors, which consider device and sensor noise as nuisances, biology has evolved to use non-linear sensing techniques to exploit noise to its advantage and operate at or below fundamental limits [5]. Thus, the neuromorphic concept of “sensing-to-learn” and “learning-to-sense” represents the need for a symbiotic understanding between biology and technology for designing the next generation

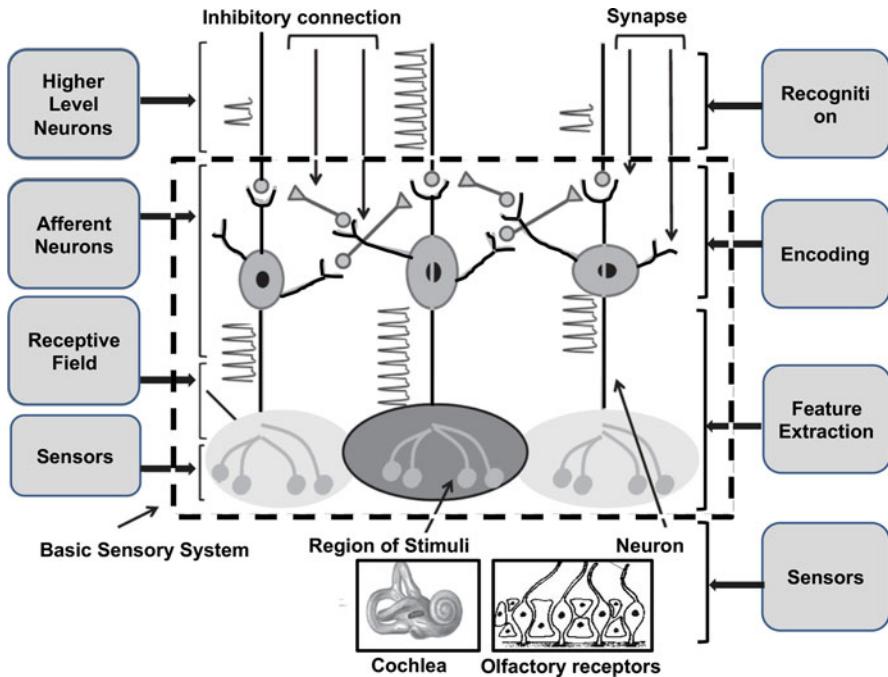


Fig. 1 Visualization of organization principle of neurobiological sensory systems

of instrumentation. The neuromorphic principles described in this chapter can be used for designing sensing interfaces for applications ranging from olfactory and acoustic sensors to radio-frequency systems (antenna arrays). The principles are generic enough because the neurosensory interfaces follow a common architecture as shown in Fig. 1. An array of transduction elements, e.g., olfactory receptors or cochlear hair cells, interfaces with a layer of receptor neurons, referred to as the afferent neurons. Each of the afferent neurons could receive electrical stimuli from multiple sensors (as shown in Fig. 1), an organization commonly referred to as the sensory receptive field. The neurons are connected through specialized junctions known as synapses.

One another most popular sensing modality is neuromorphic vision sensor [6, 7, 8] which has been designed to emulate human retina, commonly known as event camera. This sensor has displayed incredible solutions to the setbacks of the conventional camera in terms of sparsity, dynamic range, temporal resolution, etc. These sensors have led to a revolutionary way of registering only changes in the scene with microsecond temporal resolution. With present-day technology, these sensors have started to show up in the commercial field. An increase in the application of event cameras has driven many companies to get involved in the research and development of event cameras such as Dynamic Vision Sensor (DVS) and Asynchronous Temporal Imaging Sensor (ATIS).

This chapter will focus on neuromorphic signal acquisition techniques that “learn” to extract salient sensing cues by exploring the synergy between noise and specific types of non-linear processing. One principle described in the chapter is the principle of “jump-resonance,” which could be used to improve the performance of acoustic recognition systems. Another principle described in the chapter is noise shaping, which is a signal processing technique whereby the energy contained in noise and interference is shifted out of the spectral or spatial regions in which the desired signal resides [9].

In this chapter, we have taken two sensory pathways, audition and olfaction (Fig. 1), as case studies described in Sects. 2 and 3, respectively, and Sect. 4 discusses several neurobiological signal processing principles such as contrast enhancement, signal de-correlation, dimensionality reduction, etc., which could be used to build man-made technologies. This chapter also describes some robust feature extraction algorithms that combine neurobiological principles with machine learning techniques.

2 Case Study I: Neuromorphic Audition

The human cochlea is a very sophisticated and complex sensory system nature has designed. In this section, we will discuss a neuromorphic cochlea model, which is derived from the underlying physics of the human cochlea by employing various techniques such as feedback, amplification, and dynamic gain control. The human auditory system comprises several stages starting from the outer ear to the auditory cortex. In an early stage of the pathway, sound waves undergo a series of transformations, first travelling through an air-filled space outer ear canal and hitting the eardrum. This causes the eardrum to vibrate, transducing sound energy to mechanical energy. The eardrum is connected to the ossicles (middle ear bones malleus, incus, stapes), which start to vibrate when the eardrum is set in motion. The vibrations travel via the middle ear to the liquid-filled cochlea. Sound waves travel from a less dense medium (air in the outer ear) to a denser medium (liquid in the cochlea), which causes an impedance mismatch. The middle ear helps to match the outer ear’s impedance to the much higher impedance of the fluid in the cochlea. Otherwise, it would cause much of the sound entering the ear to be reflected and resulting in the auditory system’s loss of sensitivity. These smaller bones send the mechanical vibrations to the cochlea, which has three main components Basilar membrane (BM), Inner hair cells (IHC), and Outer hair cells (OHC). These inner hair cells are attached to nerves, and when they bend, it produces an electrical signal, which is carried to the auditory cortex for further perceptual tasks. The OHC control the gain of BM filters dynamically by using local instantaneous non-linearity as well as multi-time-scale non-linearity. The early auditory pathway transduces sound energy to mechanical energy via ossicles and mechanical energy to electrical energy via the inner hair cell of the cochlea. The mathematical function of the cochlea is to transform the acoustic signal into an auditory time-frequency representation, also called a spectrogram.

After the cochlea, the second major transformation occurs at the primary auditory cortex, where more complex processing takes place. A1 neuronal cells have tuned for various spectro-temporal modulations of the signal. From a functional perspective, each Spectro-temporal receptive field (STRF) [10] acts as a selective filter specific to a particular range of spectral resolutions (or scales) and tuned to a limited range of temporal modulations rates. The collection of all STRF would constitute a filter bank spanning the broad range of psychoacoustically observed scale and rate sensitivity in humans and animals.

In the next subsections, we will describe a biologically plausible model of the cochlea, which is hardware friendly and can be used for machine hearing applications. We will also discuss the jump-resonance principle [11] that occurs in auditory filters, which could improve the performance of acoustic recognition systems. Then, we will discuss different audition feature extraction techniques, such as using biological plausible STRF and statistical learning-based feature extraction (SPARK-based auditory feature extraction).

2.1 Biological Cochlea Model: Cascade of Asymmetric Resonators with Fast-Acting Compression (CAR-FAC)

There are various cochlea models that have been presented in the literature [12, 13, 14, 15, 16], but the CAR-FAC model has several properties which makes it unique and the reason to present it here. The CAR-FAC is a biologically inspired model and close enough with known auditory physiology and psychophysics experimental data [17]. It possesses the travelling-wave property of the cochlea by modelling it as a cascade of simple second-order filters. Furthermore, the incorporation of non-linearity for automatic dynamic gain control using the OHC and the AGC blocks, with the ease of digital implementation, makes it very attractive for machine hearing tasks.

The CAR-FAC cochlea model (Fig. 2) is comprised of basilar membranes (BM), which act as resonators and inner and outer hair cell (IHC and OHC) models, and a coupled automatic gain control (AGC). Figure 2a shows a functional physiological element of one location in the cochlea, including the hydrodynamic system involving the basilar membrane afferent and efferent loop through the brain stem of the auditory central nervous system that controls the activity of the outer hair cells, which provides dynamic gain. The Cascaded asymmetric resonator (CAR) part, in the Cascade of Asymmetric Resonators with Fast-Acting Compression (CAR-FAC) model, simulates the basilar membrane's (BM) response to sound. The FAC part models the outer hair cell (OHC), the inner hair cell (IHC), and the medial olivocochlear efferent system functions. The CAR (BM model) part is a linear system when its parameters are held constant, but in the context of fast-acting feedback, the FAC part, which controls the damping in the stages, becomes a non-linear system. This close integration of the linear and non-linear sections of the CAR-FAC makes it a more biologically plausible model of the cochlea.

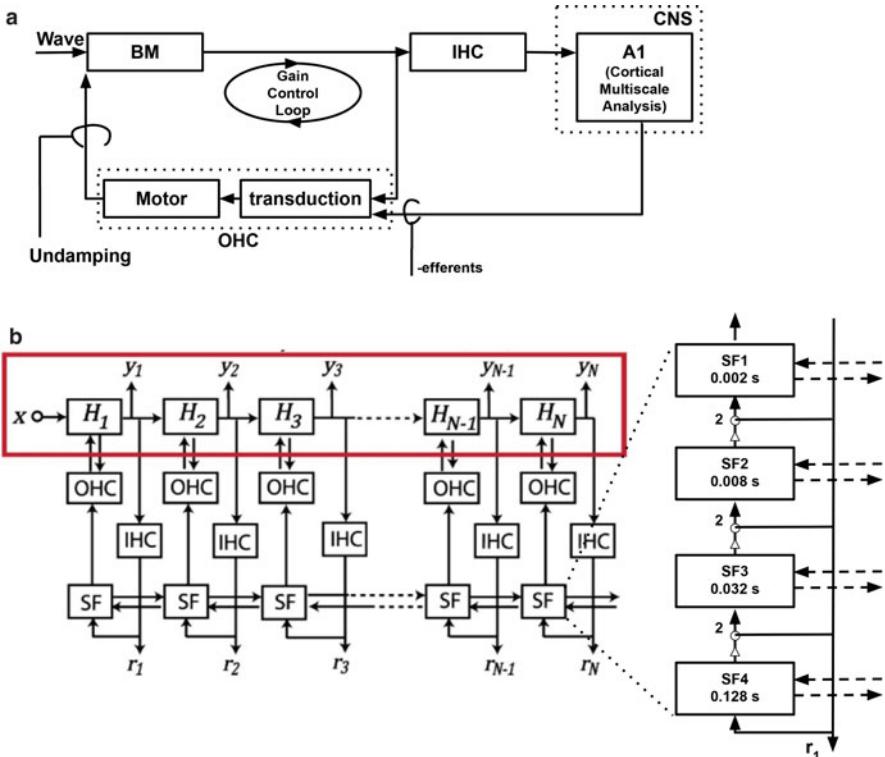


Fig. 2 (a) Visualization of elements of one spatial location of a model of the cochlea involving basilar membrane, IHC, and OHC. (b) Structure of CAR-FAC model

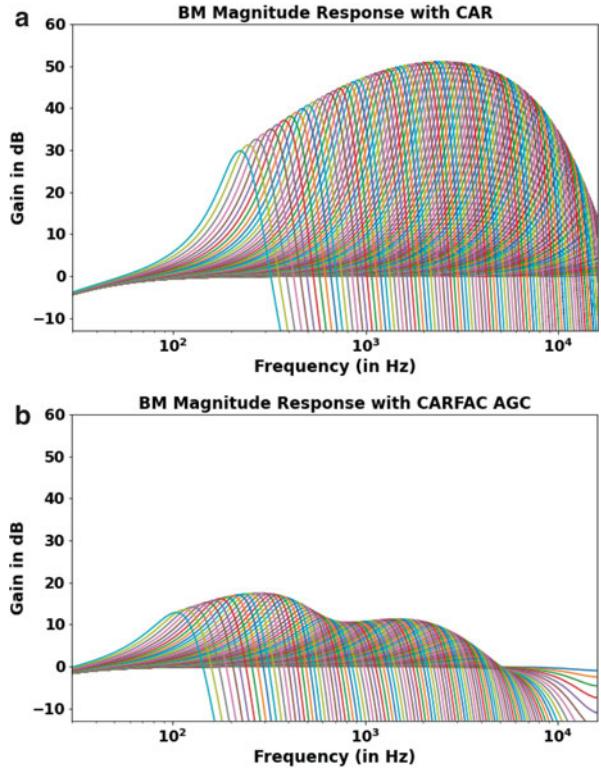
Figure 2b shows the CAR-FAC cochlea model with various blocks, which is discussed in details in the next sections.

2.1.1 Basilar Membrane (BM)

The basilar membrane is similar to the strings on an instrument. It acts as a pseudo-resonant structure [18] which changes in width and stiffness. However, it differs from the parallel strings of a guitar because the basilar membrane is a single structure that has different width, stiffness, mass, damping, and duct dimensions at different points along its length.

Traveling wave characterizes the motion of the basilar membrane [19]. The Characteristic frequency (CF), which is the frequency at which the membrane displays maximum sensitivity to sound vibrations, is determined by the membrane's properties at a given point along its length. The basilar membrane's width is high (0.42–0.65 mm), and stiffness is very low at the apex of the cochlea. The width of the basilar membrane is low (0.08–0.16 mm), and stiffness is very high at the base (near the round and oval windows) [19]. High- and low-frequency sounds localize near the base and apex of the cochlea, respectively.

Fig. 3 Basilar membrane magnitude responses. (a) CAR gain response without OHC (b) CAR-FAC gain response with OHC, which controls the gain dynamically. The pole and zero radius r is dynamically controlled as $r = 1 - d * 2 * \pi * \frac{f}{f_s}$, where $d = 0.2$ and f are the damping factor and center frequency of the filter respectively



In the CAR-FAC (Fig. 3) model, the BM part is modeled using CAR, which is a coupled form of two-pole-two-zero IIR filter. The linearized model of the human cochlea is matched by optimizing the number of filters and their coefficients [20, 21]. The choice of pole frequencies corresponds to equal spacing along the place dimension of the cochlea. This is achieved by utilizing Greenwood function for human cochlea [18]:

$$f = 165.4(10^{2.1x} - 1) \quad (1)$$

The transfer functions of the filter in Z domain is:

$$= g \left[\frac{z^2 + (-2a_0 + hc_0)rz + r^2}{z^2 - 2a_0rz + r^2} \right] \quad (2)$$

The $a_0 = \cos(\theta_R) = \frac{a}{r}$ and $c_0 = \sin(\theta_R) = \frac{c}{r}$ parameters represent the analog pole position in the zero-damping case ($r = 1$). $\theta_R = \frac{f}{f_s}$ is the normalized pole ringing frequency in radians per sample, or pole angle in the z plane and f_s is the sampling frequency. An explicit parameter r can be used to dynamically vary the

pole and zero radius in the z plane (to vary the damping factor). r is controlled by OHC to control the gain.

2.1.2 Outer Hair Cells (OHC)

Outer hair cells (OHC) are acoustical preamplifiers found in mammals to non-linearly amplify less intense sounds as compared to large intense sounds. This helps in converting a wide range of sound intensities to a smaller range of hair displacements.

In CAR-FAC, compressive amplifying wave propagation has been modeled through a Digital outer hair cell (DOHC). The gain control mechanism of DOHC integrates a local instantaneous non-linearity as well as multi-time-scale non-linearity. This filter varies the position of the poles and zeros by changing their radius r in the Z plane. r attains its maximum value in silence and down to a minimum, or passive value is a response to sound. The change in r is controlled by instantaneous non-linearity and multi-scale non-linearity. The instantaneous non-linearity comes from the BM velocity (v), whereas multi-scale non-linearity originates from the DIHC feedback through the AGC loop filter. The change in pole radius is characterized by the following equation: [19]

$$r = r_1 + d_{rz} \times (1 - b) \times NLF(v) \quad (3)$$

where r_1 is the minimum radius corresponding to the maximum damping of the resonator. The coefficient damping controls the damping factor. NLF is the non-linear function of the CAR velocity. The coefficient d_{rz} controls the rate at which the product of CAR velocity and AGC loop variable b affects the damping. With an increase in velocity, the velocity-squared function increases very rapidly and results in the saturation of NLF towards 0.

At low and high levels, OHC results in maximum and negligible undamping, thus yielding linear filter banks in both regions.

2.1.3 Inner Hair Cells (IHC)

Inner hair cells (IHC) are the transducers that help convert the sound-generated motion of the cochlea to neurotransmitter release at synapses, which excite the primary auditory neurons. Oscillatory waves in the fluids of the cochlea generate receptor currents to which IHC responds, similar to the way retinal cone cells respond to light.

Digital inner hair cell (DIHC) models the IHC function. It is modeled as a high-pass filter, a transduction non-linearity unit, a transducer unit, and two low-pass filters [22]. The high-pass filter attenuates the signal below 20 Hz. It is followed by a transduction non-linearity function comprised of a half-wave rectifier and a rational sigmoid function. The half-wave rectified output is given as follows:

$$u = HWR(BM_{hpf} + 0.175) \quad (4)$$

where BM_{hpf} is the high-pass filtered output of the previous stage. The rational sigmoid function is given as follows:

$$n = \frac{u^3}{u^3 + u^2 + 0.1} \quad (5)$$

where n is the transduction non-linearity output. This rational sigmoid function provides a linear response at low amplitudes. Its response saturates at higher amplitudes. The transducer stage is implemented with the following equations:

$$m = 1 - qy = nmq_{new} = (1 - a)q + a(cy) \quad (6)$$

where m is the adaptive gain of its input, which is quickly reduced after signal onset. n and c are constants, and q is the state of the low-pass filter. The output of the transducer is low-pass filtered with two FIR first-order filters (cutoff frequency of about 2 kHz).

2.1.4 Automatic Gain Control (AGC)

The biological cochlea has strong compressive non-linearity over a wide range of frequencies. To properly interpret the cochlea function, this strong non-linearity has been realized as automatic gain control in a digital model. This helps to convert a large dynamic range of input stimuli to a smaller dynamic range of nerve firings. It models the Automatic gain control (AGC) effect on BM vibration through the OHCs.

AGC model in CAR-FAC consists of four-stage one-pole FIR low-pass filter arranged in a parallel-cascaded configuration as shown in Fig. 2b. Their outputs are combined to cover a range of corner frequencies instead of possessing a single corner frequency. The transfer function of these four one-pole filters in parallel is nothing but the addition of the complex transfer function of the individual one-pole filters. Each filter is coupled with left and right neighboring filters and forms a three-stage spatial LPF. Thus, each stage includes a temporal linear LPF with a defined coefficient c_t and a three-tap spatial filter with coefficients $[s_1, 1 - s_1 - s_2, s_2]$. To save computation cost, these filter states are updated at a much lower rate than CAR filters. Their low-frequency gain is varied based on the range of damping factors required in the filters. The output of AGC feeds back to DOHC.

The cochlear models, including the CAR-FAC, are based on primitives that can efficiently be emulated on digital hardware or processor. In [23] and [24], CAR model describing the sound-induced traveling waves phenomena of the basilar membrane on an FPGA has been implemented. It has been extended in [25], which provides details on implementing the FAC functions of the CAR-FAC model on an FPGA. It consists of IHC and OHC algorithm blocks as well as an automatic gain control (AGC) block. There are certain phenomena that are unique to an analog substrate that biology can efficiently exploit for sensing. One such phenomenon is jump-resonance, which could boost the system's performance for machine hearing tasks and will be discussed in the next section.

2.2 Jump-Resonance-Based Auditory Filter Banks

Jump-resonance is a non-linear phenomenon where the output of a system exhibits abrupt jumps and hysteresis when the input frequency is varied [26]. Recent research in understanding active mechanisms in animal and human cochlea has revealed non-linear and hysteretic behavior similar to that of jump-resonance. Moreover, it has been hypothesized in [27] that this non-linear response might be responsible for acute sensitivity and robustness of biological auditory systems. In [28], we demonstrated that jump-resonance that naturally occurs in analog VLSI filters could be exploited to design speaker recognition systems whose performance either match or exceed that of a conventional DSP-based speaker recognition system.

The jump-resonance phenomenon has been observed and studied extensively in non-linear control and in analog filters [30] where jump-resonance leads to a hysteresis behavior when the frequency of the input signal is varied. This is illustrated in Fig. 4a where the output signal magnitude is not only a function of the input signal frequency but also a function of the direction of the frequency sweep. Thus, for frequencies within the hysteresis band defined by $\omega_1 < \omega < \omega_2$, the

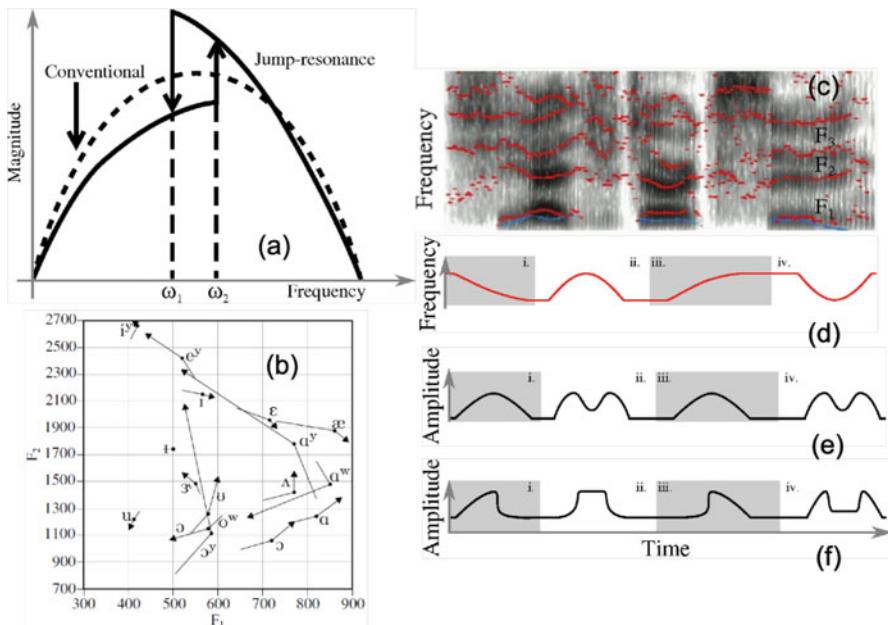


Fig. 4 (a) Jump-resonance hysteresis in auditory filters and its comparison with the response of a conventional auditory filter. (b) Vowel formant trajectories for a female speaker, recreated using data from [29]. (c) Spectrogram of a sample speech utterance showing frequency or format trajectories. (d) Sample format trajectories during a phonetic utterance in the English language. For the sample trajectory: (e) response expected from a conventional band-pass filter and (f) response expected from the filter exhibiting jump-resonance for the sample trajectories

magnitude of the output signal could have two possible magnitudes depending on the frequency trajectory. However, the hysteretic and high-Q response concerning the direction of the frequency sweep could be used as a computational tool for encoding and detecting formant trajectories in speech signals. Formants in speech signals correspond to the resonant frequencies of the vocal tract, in particular, when vowels are pronounced. Figure 4c shows the location of three formant frequencies (F1, F2 and F3) on a spectrogram. The trajectories of formants over time (as shown in Fig. 4c) are particularly relevant for speaker and speech recognition because they are an indicator of the mechanical dynamics of the vocal tract and that these signatures are robust to corruption by ambient noise. For example, Fig. 4b shows example trajectories of the formants F1 and F2 corresponding to different English vowels and corresponding to a female speaker. Figure 4c illustrates the application of jump-resonance using an example of formant trajectories in a spectrogram of a typical speech utterance; and Fig. 4d shows a stylization of some of these trajectories. Also shown in Fig. 4e–f are the comparisons of expected outputs that will be produced by a conventional auditory filter that does not exhibit jump-resonance versus the output produced by a similar filter exhibiting jump-resonance. For instance, the output produced by the conventional filter in regions ii and iv (see Fig. 4e) are identical even though the frequency trajectories in these regions (see Fig. 4d) are different. However, this is not the case for the output produced by the jump-resonance filters (see Fig. 4f). Thus, the illustration shows that jump-resonance could lead to output signatures that are unique to the frequency trajectory and could be useful in providing more discriminatory information to a back-end speech or speaker recognition system. Also, the hysteresis in the filter response caused by the jump-resonance may improve speaker recognition performance, as small variations of spectral content caused by a hoarse voice or other voice modulations don't have to result in large changes in the output. However, from a practical point of view, implementing jump-resonance behavior on a DSP is computationally prohibitive because it requires emulation of chaos and bistability, both of which require high-sampling rates. Fortunately, jump-resonance naturally occurs in continuous-time analog filters and could be easily emulated on dedicated analog VLSI integrated circuits.

Figure 5a shows the prototype analog VLSI chipset that has been designed to implement a large array of jump-resonance filters [31]. The parameters of each of the filters (center frequency, quality factor, filter hysteresis) are individually programmable using an external digital interface. Figure 5b and c shows the measured filter responses, which illustrates that the hysteresis levels can be programmed. Figure 5d–g shows the measured response of the filters when the frequency of the input signal is swept according to the trajectory as shown in Fig. 5d and e. As predicted in Fig. 5f, the filter responses can discriminate between the two frequency trajectories validating the frequency encoding properties of the jump-resonance filters.

Table 1 shows the results from a preliminary experiment where the outputs of 11 jump-resonance filters were used as features to train a speaker recognition system. A Support vector machine (SVM) was used as a back-end recognizer [32, 33], and

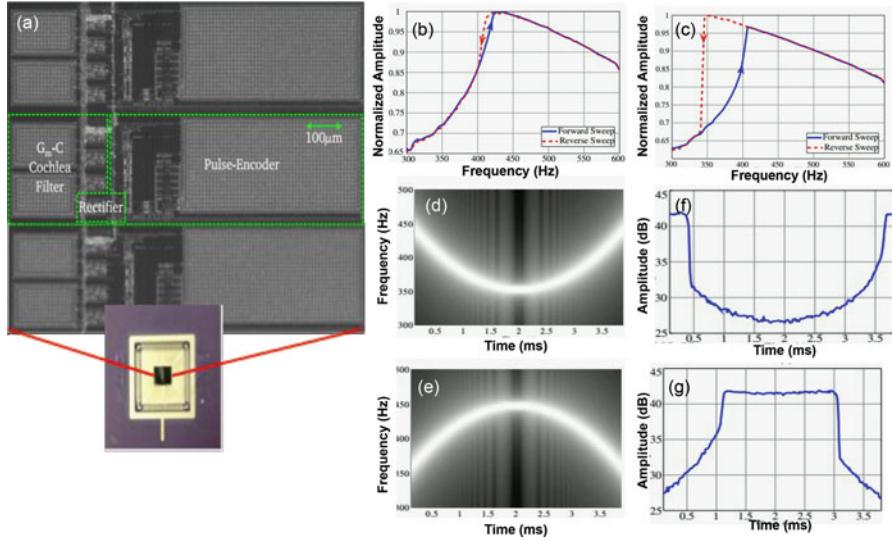


Fig. 5 (a) Micrograph of analog VLSI chipset with 100 programmable jump-resonance filter banks; (b)–(c) measured response of a single jump-resonance filter illustrating that the hysteresis level can be tuned; (d)–(e) time-frequency map of the input to the jump-resonance and (f)–(g) corresponding measured response of the filter

Table 1 Comparison of EER and PD for YOHO

Neuromorphic VLSI (analog)		Jump-resonance (analog)		MFCC (software)	
EER	PD	EER	PD	EER	PD
0.0201	0.9563	0.0014	1.000	0.0017	1.000

the training/test samples were obtained from the YOHO speech corpus. The result shows that the performance of analog VLSI jump-resonance features is similar or better than DSP-based MFCC features. For comparison purposes, the Equal error rate (EER) corresponding to the state-of-the-art conventional neuromorphic VLSI features is reported, which shows the result to be an order of magnitude lower. The power dissipation of the analog VLSI chipset is less than 100 nW [34] which is three orders of magnitude more energy-efficient than a DSP implementation of MFCC feature extraction algorithm.

As shown in Fig. 1, the first few neurosensory layers comprise neural circuitry that can extract discriminating features from the cochlear interface. In literature, there are two approaches to design neuromorphic feature extractors: (a) features that mimic the observation from neurobiological experiments and (b) features that are learned based on machine learning primitives. In the next section, we describe two examples of such feature extractors.

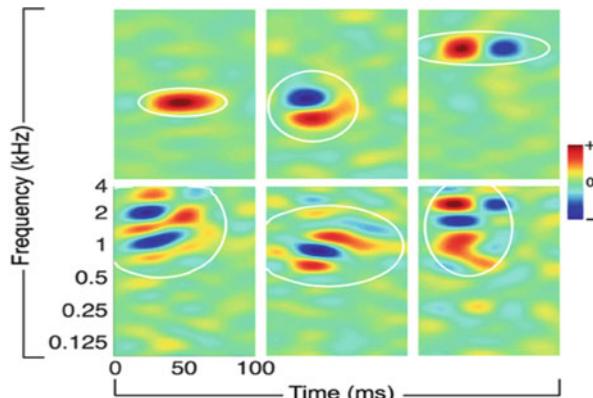
2.3 Biological Feature Extraction (Spectro-temporal Receptive Field (STRF)-Based Approach)

Analysis of sound signals inside the cochlea and its spectral decomposition is only the first stage of a series of complex transformations to the sound as it travels through the complete auditory pathway. After the cochlear analysis, the neural signals from the cochlea reach the primary auditory cortex via the thalamus and midbrain, where they undergo a multi-resolution spectro-temporal analysis [35]. Based on the physiological experiments on the mammalian auditory cortex, scientists have characterized the spatial organization and response properties of the cortical cells. Since the organization of the auditory cortex is similar in mammals, these physiological findings provide insight into the representation of sound features in the human auditory cortex.

The primary auditory cortex (also known as A1) extracts the temporal and spectral modulation content from the cochlear output. It has been found that the cells in the primary auditory cortex respond selectively to different rates of temporal modulation and different scales of spectral modulation [35]. The way a neuron in the A1 region responds to these rates and scales of modulation is specified by its spectro-temporal receptive fields (STRFs). Figure 6 shows the STRFs of six neurons measured from the auditory cortex of a ferret [36]. Along the ordinate, STRF depicts the range of frequencies to which the neuron responds. It is evident that some neurons are responsive to a wide range of frequencies, while some others are responsive to a narrow range. Similarly, along its abscissa, STRF depicts the response dynamics to an impulse delivered at each frequency. Some responses fade rapidly, while some others stay twice as long as others. For certain neurons, sensitivities towards a particular rate and scale are combined, and their STRF takes more complex forms with different orientations.

Along with the cortical model, STRFs can be used to mimic the human auditory pathway up to the A1 and to perform any audio-related tasks such as speaker

Fig. 6 STRFs recorded from the primary auditory cortex of ferret



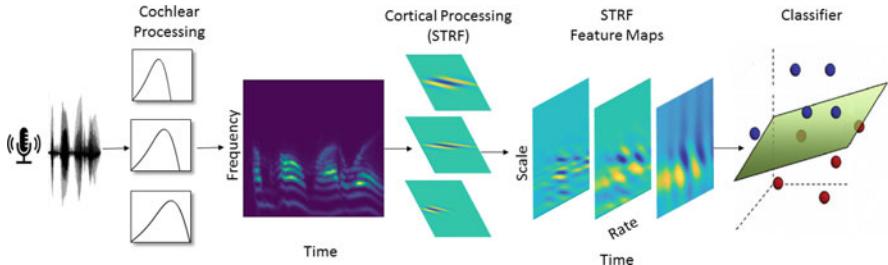


Fig. 7 Application of STRFs in a sound classification task. The audio samples are used to generate the spectrogram using the cochlea model. The output of the cochlea model is convolved by the STRF filters to generate the cortical feature maps. The features are used by a simple classifier network to differentiate between various sound signals

recognition, sound segregation, denoising, voice activity detection, etc. Figure 7 shows how STRFs can be used for an auditory classification task.

2.4 Statistical Learning-Based Feature Extraction

In this section, we will discuss the data-driven statistical learning-based feature extraction technique. Even though incorporation of “smart” signal acquisition could potentially enhance the fidelity of the separated auditory signal, in real-world environments, the quality of separation could significantly vary, as is observed in many ICA-based applications [37]. Therefore, the feature extraction module becomes critical in achieving robust recognition performance. For the proposed end-to-end learning approach, we expect the feature extraction module to demonstrate three key properties: (a) the ability to extract salient auditory features that are robust to ambient noise and interference; (b) computational efficiency and the ability to extract features in real time; and (c) the ability to guide the signal acquisition process to facilitate more robust separation of auditory sources. We will first describe our previous work related to the property (a) and defer (b) and (c) as a part of the proposed research.

Over four decades, the auditory features used in most state-of-the-art recognition systems have relied on spectral-based techniques, which inherently assume the linearity of the speech production mechanism. Some examples of spectral-based features include Mel-frequency cepstral coefficients (MFCC), perceptual linear prediction (PLP), cochlear features, and linear prediction (LP) coefficients. MFCC features, for example, are calculated using a discrete cosine transform on the smoothed power spectrum, and PLP, similar to MFCCs, are based on human auditory models. Even though these features are computationally efficient and deliver acceptable accuracy under controlled recording conditions, their performance degrades significantly when subjected to noise present in practical environments. This performance degradation is primarily attributed to an unavoid-

able mismatch between training and recognition conditions, especially when the characteristics of all noise sources are not known in advance. Even though the effect of mismatch can be alleviated by adaptation of the statistical model used in the target recognition, the overall system performance is still limited by the quality of speech features. Also, for the proposed end-to-end learning, it is beneficial to extract and adapt parameters of the feature extraction so that the process can be coupled with the signal acquisition.

A promising alternative approach to extracting noise-robust speech features involves the use of data-driven statistical learning techniques that do not make strict assumptions about the spectral properties of the speech signal. Examples include kernel-based techniques [38] that operate under the premise that robustness in the speech signal is encoded in high-dimensional temporal and spectral manifolds that remain invariant in the presence of ambient noise. The objective of such a feature extraction procedure is to identify the parameters of the noise-invariant manifold. The procedure used in [38] requires solving a quadratic optimization problem for each frame of speech, making the data-driven approach computationally intensive. Also, due to its semi-parametric nature, the methods proposed in [38] do not incorporate demonstrably useful a priori information revealed by neurobiological and psychoacoustical studies [39,40]. More recently, it has been demonstrated that cortical neurons use the highly efficient and sparse encoding of auditory signals [41] over spectral and temporal receptive fields. The study [41] shows that auditory signals can be sparsely represented by a group of basis functions that are functionally similar to gammatone functions, which, in turn, are equivalent to time-domain representations of human cochlea filters that are also used in psychoacoustical studies. Other neurobiological studies [35] have proposed a hierarchical auditory processing model consisting of spectro-temporal receptive fields (STRFs) [35] that capture information embedded in different frequency, spectral, and temporal scales. The results from many of these recent neurobiological and psychoacoustical studies are being incorporated in small-scale speech recognition systems. The result has opened the possibility that a hierarchical auditory recognition model could be constructed similar to the popular HMAX model used in real-time vision-based recognition systems [43,44,45]. The central challenge in developing such a model is determining a computationally efficient architecture that can combine the auditory receptive fields at different spectral and temporal scales.

SPARK-based auditory feature extraction In [46] we proposed to use the previously reported hierarchical auditory feature extraction model using an RKHS-based statistical learning approach. The model called SPparse Auditory Reproducing Kernel (SPARK), summarized in Fig. 8, consists of two signal-processing layers. The first layer computes the similarities between the sample speech signal and different sets A_1 to A_M of pre-computed gammatone basis functions. Each set is comprised of time-delayed versions of a gammatone function emulating an auditory phase-sensitive receptive field. An example of the gammatone receptive field is shown in Fig. 8 (bottom). In the frequency domain, this field is similar to the cochlea filter banks. The second layer of the SPARK model implements a winner-take-all

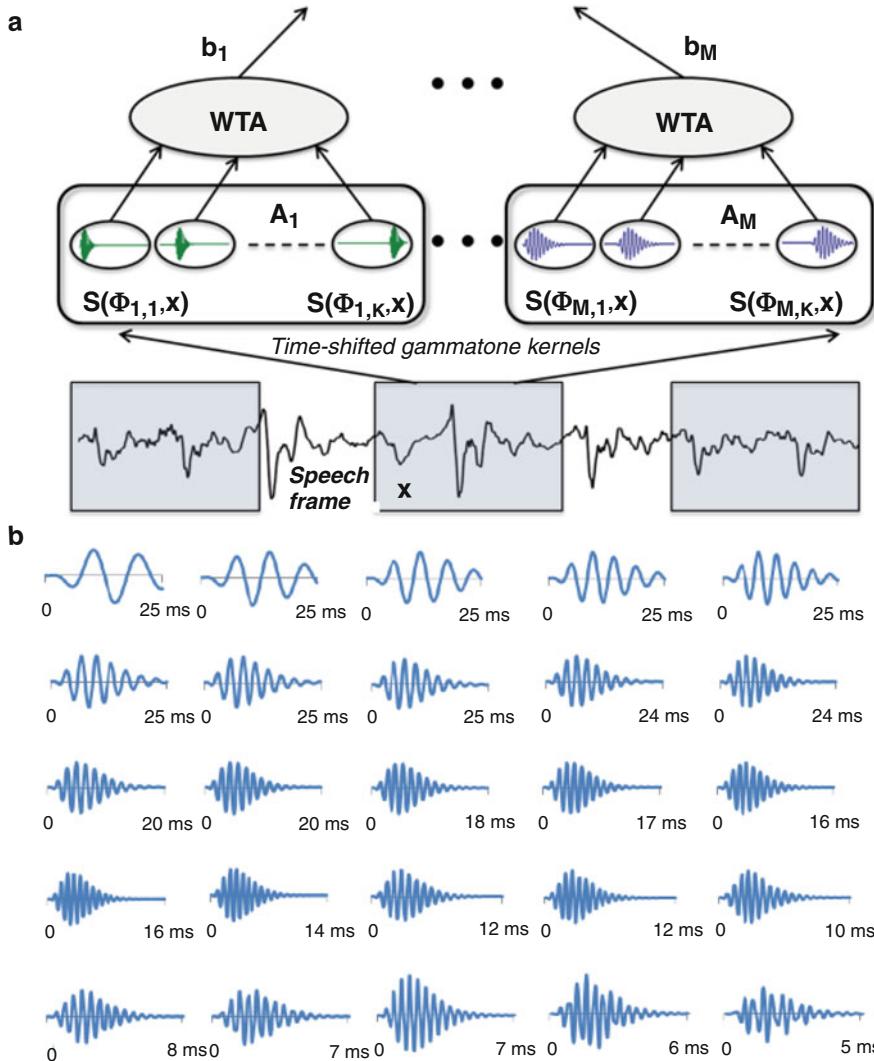


Fig. 8 (Top) Hierarchical architecture for SPARK auditory feature extraction. (Bottom) Time-domain representation of gammatone basis which will be used for CLU feature extraction. The gammatone filters take the form $\varphi_m(t) = a_m t^{k-1} \exp^{-2\pi\beta_m t} \cos(2\pi f_m t + \varphi_m)$ where m denotes the index of different gammatone basis

(WTA) function, which selects the largest similarity metric from each set A_1 to A_M (Fig. 8). The WTA-based selection is similar to the popular HMAX algorithm [47] used in vision systems, where edge-sensitive receptor functions are used to compute similarity measures.

As shown in Fig. 8 (top), the similarity function $s(., .)$ is computed between a frame of speech signal x and a set of physiologically inspired gammatone functions $\varphi_m(t)$. The similarity functions are first determined by minimizing the following objective function with respect to the similarity functions:

$$\min_s \left[\lambda \|s\|_2^2 + \|\psi(\mathbf{x}) - f\|_2^2 \right] \quad (7)$$

where $\|\cdot\|$ denotes a functional norm and $f = \sum_{m=1}^M \sum_{l=1}^L s(\varphi_{l,m}, x) \Psi(\varphi_{l,m})$ where $\Psi(.)$ is a non-linear mapping that transforms the speech signal into a higher-dimensional space. The first part of the objective function acts as a regularizer, which penalizes large norms of similarity functions, thus favoring similarity measures that are smooth (or penalizes high-frequency components of the similarity function). The second part of the cost function is the least-square error function computed between the speech vector and the reconstructed waveform f . The hyper-parameter λ controls the trade-off between achieving a lower reconstruction error and obtaining a smoother similarity function. This trade-off is similar to the conventional bias-variance trade-off encountered in any functional regression procedure [48, 49]. In [50], we showed that the solution to (7) yields the following form of the similarity function:

$$s(\varphi_{l,m}, \mathbf{x}) = (\mathbf{K} + \lambda \mathbf{I})^{-1} K(\varphi_{l,m}, \mathbf{x}) \quad (8)$$

where $k = \Psi(.)^T \Psi(.)$ is a reproducing kernel satisfying Mercer's criteria. The use of reproducing kernels avoids this “curse of dimensionality” by avoiding direct inner-product computation. Note that the matrix inverse in equation 8 is independent of the input speech vector x and hence can be pre-computed and stored. Thus, the computation of the SPARK similarity functions involves computing kernels and a matrix-vector multiplication, which can be made computationally efficient. An important consequence of projecting the speech signal onto a gammatone function space (emulating the auditory STRFs) is that the highest scores (in L_2 norm) in the similarity metric vector s will capture the salient, higher-order, and spectro-temporal aspects of the speech signal. On the other hand, the low-energy components of $s(., .)$ will also capture similarities to noise and channel artifacts. Feature pooling of the similarity scores in the SPARK model serves two purposes. First, it introduces competitive masking, where only the largest similarity score is chosen. This function emulates the local competitive behavior, which has been observed in auditory receptive fields. The second purpose of feature pooling is to introduce a compressive weighting function (similar to psychoacoustical responses), which enhances the resolution at low similarity scores and reduces the resolution at the high similarity.

Figure 9 (left) shows the spectrograms of utterance “one” under clean and noisy (subway recording) conditions. Figure 9 (middle) shows the similarity metric vector computed using 25 ms of speech segments shifted by 10 ms over clean and noisy speech utterances. Similarly Fig. 9 (right) shows the pooled vectors \mathbf{b} for the same utterances.

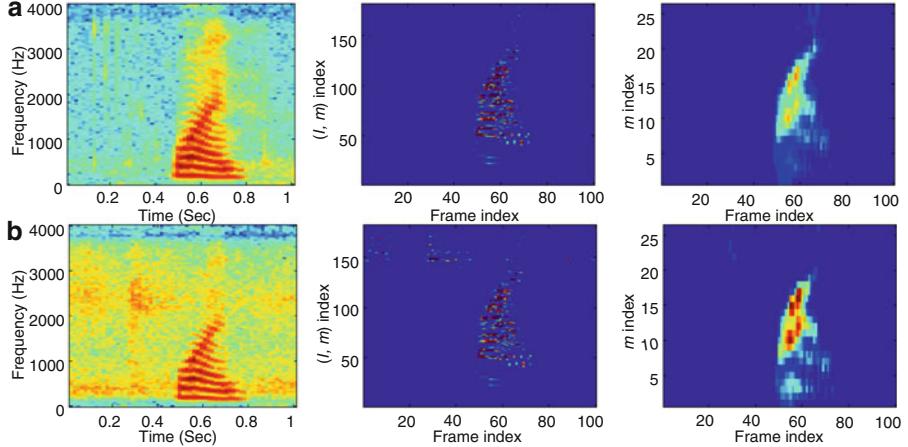


Fig. 9 (Left) Spectrogram corresponding to utterance “one” under clean conditions (top) and under noisy conditions (bottom) – subway noise at 10 dB SNR; (middle) similarity scores based on equation 7 for clean and noisy utterance; (right) SPARK feature obtained after pooling the similarity scores according to Fig. 8

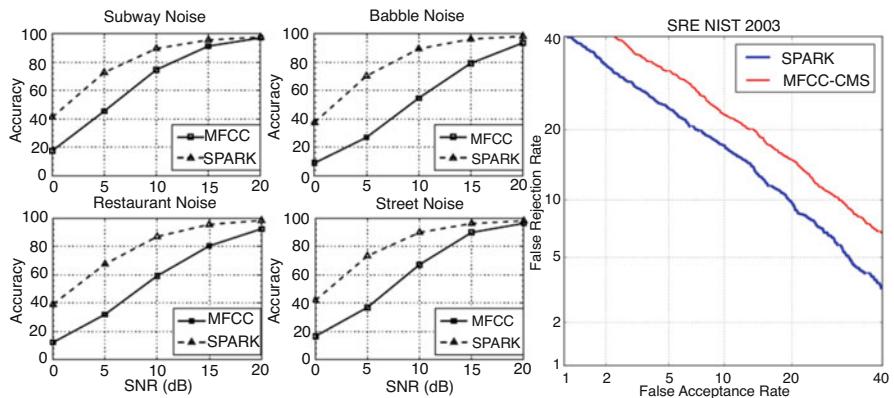


Fig. 10 (Left and middle) Comparison of speech recognition accuracy obtained using SPARK features and the ETSI STQ WI007 DSR front-end (MFCC) evaluated for the AURORA2 corpus under different recording conditions and noise levels. (Right) Comparison of speaker recognition accuracy evaluated using the NIST 2003 SRE dataset

In [46], the accuracy of the SPARK-based recognition system has been compared against the baseline speech features extracted using the ETSI STQ WI007 DSR front-end [51]. The basic ETSI front-end generates the 39-dimensional MFCC features. Figure 10 (left and middle) compares the word recognition rate obtained by the SPARK and basic ETSI based recognizers. The results show that the SPARK-based recognition system consistently outperforms the benchmark at all Signal-to-noise ratio (SNR) levels. The average relative word-accuracy improvement was

found to be 33%, 36%, and 27% for set A, set B, and set C of the AURORA2 dataset. The SPARK features were also evaluated for the task of speaker recognition using the NIST 2003 SRE dataset. Figure 10 (right) compares the average receiver operating curve (ROC) with an equivalent recognizer using the MFCC (with cepstral mean subtraction) features. The results again show an improved recognition performance compared to the state of the art.

3 Case Study II: Neuromorphic Olfaction

In this section, we will discuss the fundamental sensing principles employed by the biological olfactory system and how they have inspired parallel engineering approaches [54, 55, 56]. It might be worth pointing out that engineered electronic noses that combine arrays of chemical transducers and signal processing algorithms [57, 58, 59] still lack capabilities exhibited by even simpler organisms such as insects. While application in several domains, including homeland security, medical diagnostics, and environmental monitoring, exists, this potential is yet to be realized and presents a challenge as well as a future opportunity for creating a neuromorphic sensory system. We will focus on reviewing those biological principles that have been mimicked for processing chemosensor arrays. A number of sensing technologies [59] have been used for transducing information about chemical cues. It is worth pointing out that the molecular features that are sensed by olfactory receptors remain poorly understood. Therefore, the chemical transducers used in electrical or optical noses do not necessarily contain the same information as the biological sensor array. Nevertheless, they can be designed to provide a combinatorial input that is qualitatively similar, thereby allowing subsequent processing using various bio-inspired approaches.

3.1 Basic Building Blocks of the Biological Olfactory System

This section gives details of the biological olfactory system. The various components of the olfactory system are biological transducers/arrays that sense the cues in the environment, gaining control to normalize the large dynamic range of encountered stimulus and the circuitry involved in the reduction of overlapping neural representations. These components handle multiple odors encountered at the same time and the generation of sparse odor codes.

3.1.1 Biological Transducers and Sensor Arrays

A quick comparison of the organization of the olfactory system indicates that most animals employ striking similar features to interact with chemical cues in their environment [61] (Fig. 11). First, a large number of sensors that vary both in their selectivity and sensitivity [62, 63] and multiple copies of each sensor type are employed in the sensor array (located at the nasal epithelium invertebrates and the antenna for invertebrates). Second, this input is quickly sorted such that sensors

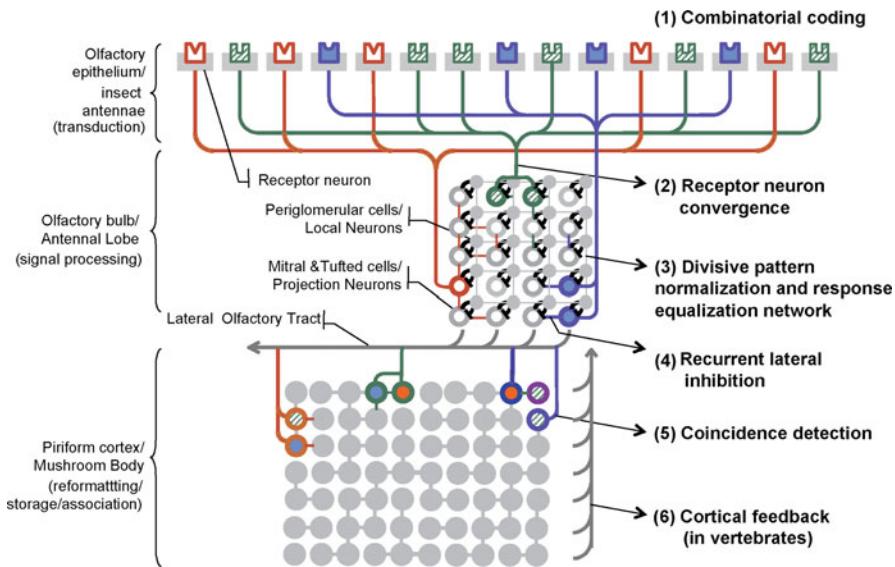


Fig. 11 Different anatomical stages and their putative computational roles in vertebrate/invertebrate olfactory pathways are identified. (Adapted from [60])

of the same type appear to be grouped in spherical structures of neurophil called the glomeruli [63]. Here there are some differences between species. In some animals, such as the fly and the mouse main olfactory bulb, sensors of the same type send projections to one or two glomeruli in the next anatomical stage (called the olfactory bulb in vertebrates or antennal lobe in insects), while in other models, such as locusts [64] and mouse accessory olfactory bulb [64], each sensor projects to multiple glomeruli. Irrespective of the precise wiring scheme with which the sensors are connected to the following circuits, this results in a massive reorganization of the sensory input from biological chemical transducers. This reorganization helps achieve multiple computational functions, including dimensionality reduction (several hundreds of thousands of sensors to few tens or hundreds of glomeruli). Since sensors of the same type are grouped, the uncorrelated fluctuations or noise can be canceled out using this input sorting scheme [66]. This improves the signal-to-noise ratio by a factor of \sqrt{n} , where n is the number of sensors that are grouped. This mechanism can cause the detection threshold of the overall system to become lower than that of each individual transducer/sensor [67].

The information from reorganized sensors is then processed in the vertebrate olfactory bulb/invertebrate antennal lobe by a neural circuit comprising excitatory and inhibitory elements. In the vertebrate system, the excitatory neurons are mitral/tufted cells, and inhibitory neurons are periglomerular cells and granule cells. In invertebrate models, they are referred to as principal/projection neurons (excitatory) and local neurons (inhibitory). Note that the local neurons, at least in invertebrates, have been found to be spiking (i.e., digital in fruit flies and honey

bees) or non-spiking (i.e., analog in locusts) depending on the insect model. Only the output of the excitatory neurons is projected onto higher centers in both cases. Most importantly, several signal processing tasks are carried out through interactions between the neurons present at this anatomical level.

3.1.2 Gain Control and Response Equalization

Chemical cues encountered in natural environments can vary by enormous amounts (a banana on a plant or underneath your nose). To deal with the challenges in encountering a stimulus over an extensive dynamic range, the concentration information is often compressed in a signal processing operation typically referred to as the gain control. This is true for many sensory system [68] and is also the case in olfaction. It has been shown in fruit flies, for example, the overall inhibition received by the second-order projection neurons in the antennal lobe scale with the total sensory neuron input received [69]. This is succinctly captured in the following equation:

$$R = \gamma \frac{I^n}{\sigma^n + I^n} \quad (9)$$

where I is the activity of the sensory neuron and the remaining variables determine the overall shape of the I/O curve (Fig. 12). Furthermore, this sigmoidal input-output relationship can further be modulated by the presence of masking/competing odorants that activate other projection neurons in the same neural circuit. The revised relationship that takes into account the total activation of all other sensory neurons to the antennae lobe neurons (Σ) is captured as follows:

$$R = \gamma \frac{I^n}{\sigma^n + \Sigma^n + I^n} \quad (10)$$

While this form of divisive normalization can help saturate the responses evoked by odorants at higher intensities, a different mechanism would be necessary to

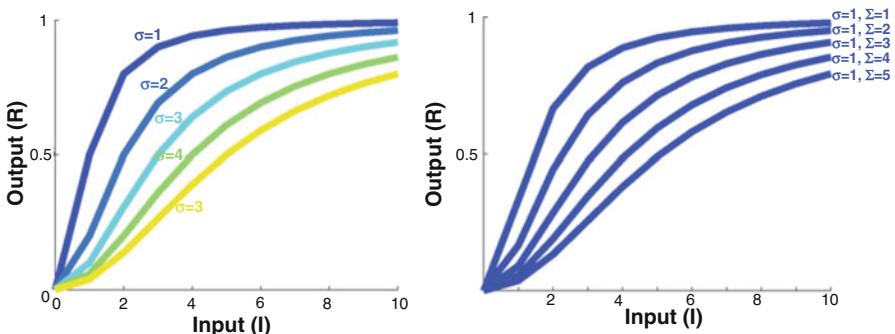


Fig. 12 Input-output relationships following equations 9 and 10. For both plots $n = 1$

boost responses at lower intensities. In vertebrate olfactory systems (and possibly in invertebrates as well [70]), such a mechanism does seem to exist [71]. Notably, both signal manipulations seem to be achieved through local neurons that mediate gain control through GABAergic chemical synapses to saturate responses at high intensities and excitatory gap junctions that amplify weaker responses evoked by odorants at lower intensities. Overall this computation has been suggested to help maintain the representation of odorants relatively stable over a large intensity range.

3.1.3 Contrast Enhancement

The next computation of considerable importance in olfaction is the ability of the circuits to reduce overlap between neural representations. Note that the sensory neurons employed in the periphery can have broad specificities and generate a highly overlapping combinatorial response to any odorant encountered. Therefore, the following circuits in the antennal lobe/olfactory bulb are tasked with the role overlap and increasing specificity of the responses elicited by different odorants.

One potential way the olfactory neural circuits downstream to sensory neurons are thought to make representations of different odorants more specific is through the refinement of responses elicited by an odorant over time. It has been shown in zebrafish that the initial responses elicited are highly correlated between different stimuli. But these initial overlapping response profiles are refined such that more odor-specific features are extracted, and the neural representation becomes identity-specific with time [72]. Intuitively, for the problem of recognizing a fruit (say lemon), the response refinement would progress in the following fashion: odor present → fruit odor → citrus fruit → lemon (Fig. 13). As can be noted, the initial responses would be common across a group of odorants (i.e., fruity or citrus), thereby allowing identification of odor groups (i.e., solves the clustering problem). At the same time, the later responses would be odor-specific and allow solving the recognition problem. This computation has also been referred to as de-correlation over time.

In direct contrast to this approach, a recent work in mice has suggested that only the early responses may matter for recognition of an odorant [73]. According to this theory, the most sensitive sensory neurons should have the fastest response to the stimulus and would be consistently recruited over a wide intensity range. Therefore, the combination of early responders would represent the stimulus identity robustly and in an intensity-invariant fashion. Such an approach would be consistent with behavioral evidence that reveals odor recognition happens rapidly within a few hundred milliseconds after the onset of a stimulus.

Irrespective of whether the stimulus-evoked responses that evolve over time in neural circuits are relevant to behavioral readout or not, both these schemes provide novel approaches to encode a sensory stimulus that could be mimicked in their engineered counterparts.

Note that contrast enhancement could also occur between stimuli. Recent work on locusts reveals that when encountered in a non-overlapping sequence, the response to the second odorant in that sequence varies in a stimulus history-dependent manner [74]. The neurons that responded to both the first and second

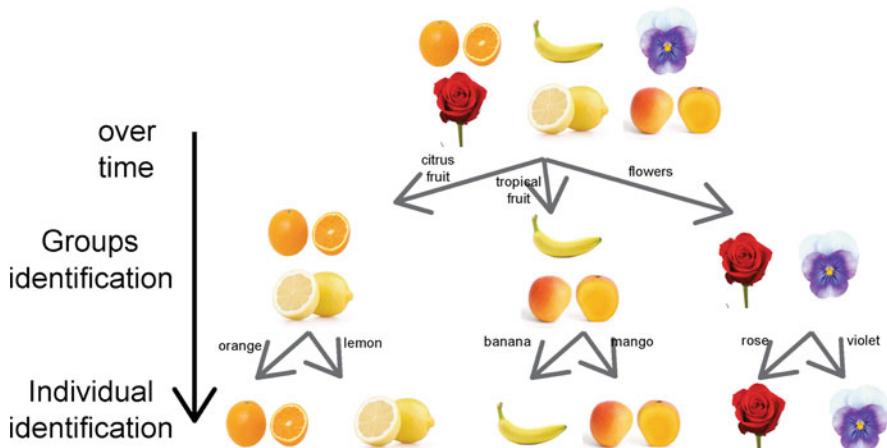


Fig. 13 Extracting stimulus-specific features over time. Neural responses at different points in time correspond to each level of this decision tree. Top levels correspond to earlier response phases, and lower levels are reached through refinement over time. Also, note that top levels uncover odor groups or clusters, and the bottom level allows precise recognition

stimuli are suppressed, and only the unique responders that represent the second stimulus in the sequence robustly encode for that odorant. This could be viewed as a novelty contrast enhancement computation.

3.1.4 Background Invariance

Recognizing a freshly encountered stimulus independent of other competing/distracting stimuli that are already present in an environment is a computation that must be performed by all sensory systems. This capability has been reported in both invertebrate [75] and vertebrate olfaction [76]. In the insect olfactory system, it was reported that within a few hundred milliseconds, the neural circuits in the antennal lobe are ready and can robustly track any new stimulus encountered. Neural responses evoked by the same odorant in different backgrounds pattern-match and allow for background-invariant recognition of a stimulus.

Furthermore, for robust odor recognition, a very simple framework was proposed [74]. In this approach, solitary encounters of an odorant without other competing cues would activate a specific subset of olfactory neurons. Encounters of the same stimulus in different backgrounds or in different sequences (i.e., varying stimulus history) would activate only a variable subset of the neurons that are usually activated in solitary encounters. However, the odor identity could still be recovered by decoding information from a flexible subset of neurons. A simple linear perceptron-like classifier with a tunable threshold would be sufficient in such a scenario to robustly recognize the odorant. Such a computation can also be thought of as an OR-of-ANDs logic operation (Fig. 14).

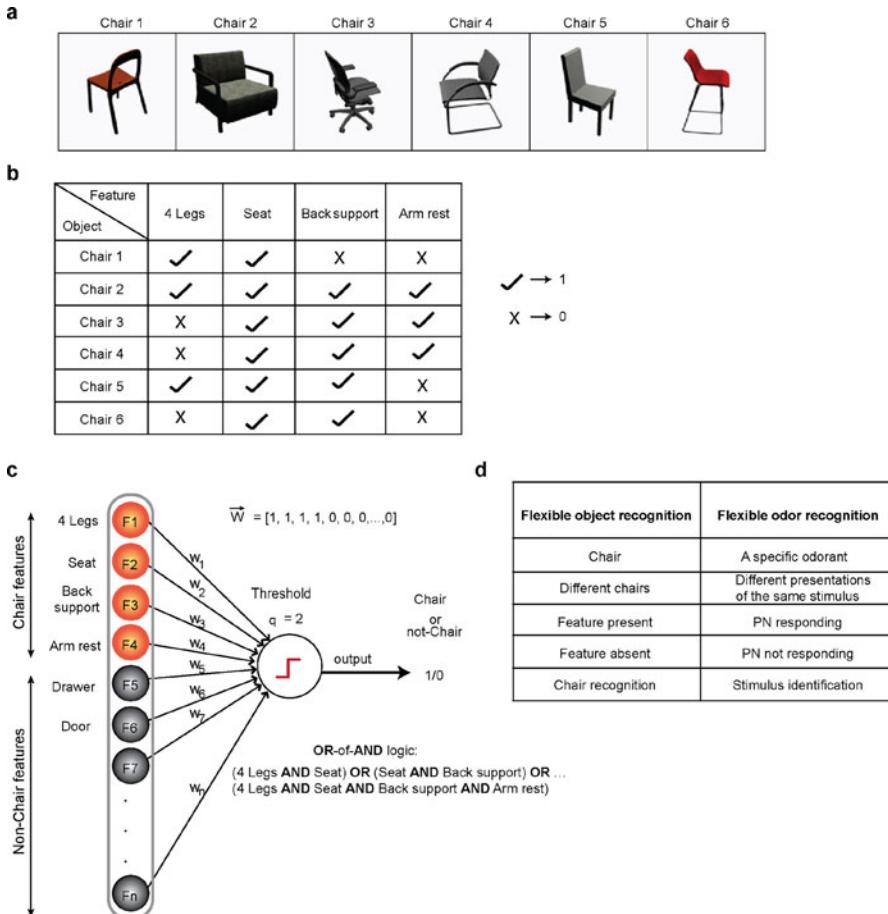


Fig. 14 A simple illustration of a flexible decoding approach allowed robust recognition of odorants in an invertebrate olfactory system. Each feature corresponds to the activation of a single projection neuron in the antennal lobe. Each object represents an odor stimulus. Variations across different chairs are used to indicate the variability encountered when the same stimulus was received in different stimulus sequences

3.1.5 Handling Chemical Mixtures

Most naturally occurring olfactory stimuli are complex mixtures. However, unlike vision and audition, handling multiple stimuli that are encountered simultaneously (i.e., odor mixtures) poses a difficult challenge for the biological olfactory system. Several studies have investigated how binary mixture that constitutes two independent components are processed. Results from those studies indicate that such binary mixtures can be a linear combination of the component odor responses [77] or can result in neural representations that switch between different attractor states with proximity to the dominant component in the mixture [78]. It is well known

that odor mixtures with more than three components are processed as whole odor objects that cannot be further segmented [79]. In this sense, olfaction is regarded as a holistic sense. Note that this processing limitation only applies to temporally synchronous mixtures where the components are encountered near simultaneously. For temporally asynchronous mixtures, as noted earlier, the olfactory system rapidly adapts to the first cue and maintains sensitivity to process novel stimuli encountered.

3.1.6 Sparse Decoders

Odor-evoked responses in the early processing stages in the olfactory system are dense, and the information about the odorant is usually distributed across a large combination of neurons. However, as the responses are transmitted to higher centers (mushroom body in insects and piriform cortex in vertebrates), the activity becomes sparse both in space and time. How are these sparse odor codes generated? A very simple mechanism: each decoding neuron in the higher center receives input from a random combination of neurons in the encoding layer, and a suitably chosen threshold can generate sparse neural responses [80]. How these neural responses mediate learning and memory, as they have been hypothesized to, remains to be understood.

Based on the current understanding of how signals are received and processed in biology, an overall schematic of a neuromorphic approach for machine olfaction is shown in Fig. 15.

4 Biological Signal Processing Principles

Sections 2 and 3 discussed neuromorphic cochlea model and olfactory models, respectively, derived from the underlying physics of biological cochlea and olfactory systems. This section has two different parts that discuss how biological signal processing principles are mimicked or can be used in building neuromorphic sensors and systems. All the principles described in these sections are implemented by the sensory neuron layer referred in Fig. 1

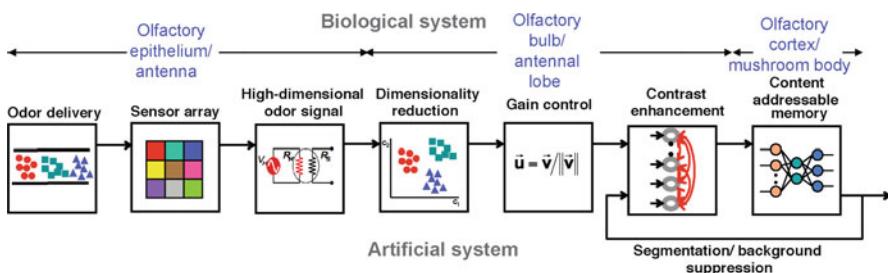


Fig. 15 A neuromorphic architecture for machine olfaction [56, 60]. Each component represents sensing or a signal processing module in biological olfaction that has inspired parallel approaches for analyzing responses of chemical sensor arrays

4.1 Noise-Shaping in Integrate and Fire Spiking Neural Networks

Even though several spiking neural network architectures have been studied in literature [81], one of the popular models of neurons is based on the integrate and fire model. In this model, a neuron receives an external stimulus $x_i(t)$ (which is typically a current) along with pulses from the neighboring neurons $d(t)$ (known as pre-synaptic spikes), which are then integrated, and the resulting intrinsic voltage $V_i(t)$ is compared against a random threshold. When this intrinsic voltage exceeds the threshold, the neuron produces a spike, after which the intrinsic voltage is reset to a random parameter that is below the threshold voltage. This section illustrates the concept of noise-shaping using a mathematical model of an integrate-and-fire spiking neural network. Consider a neuronal network consisting of N integrate-and-fire neurons. Each neuron is characterized by its intrinsic voltage $V_i(t)$, $i = 1, \dots, N$ and fires whenever V_i exceeds a threshold V^{th} . Between consecutive firings, the dynamics for V_i are given by the coupled differential equation 14

$$\frac{dV_i}{dt} = -\frac{V_i}{\tau_m} - \sum_{j=1}^N \sum_m K_{ij} d(t - t_j^m) + x_i(t) \quad (11)$$

where $t_j^m, m = 1, 2, \dots$ is the set of firing times of the j th neuron and τ_m denotes the time constant of the neuron capturing the “leaky” nature of integration. The parameter set K_{ij} denotes the synaptic weights between the i th and j th neuron and also denotes the set of learning parameters for this integrate-and-fire neural network. To show how the synaptic weights K_{ij} influence noise-shaping, consider two specific cases: (a) when $K^{ij} = 0$ implying there is no coupling between the neurons and each neurons fires independently of the other and (b) when $K_{ij} = K$ implying that the coupling between the neurons is inhibitory and is constant. For this simple experiment, the set of parameters τ_m is set to 1 ms, and N is set to 50 neurons. For the case when the input $x_i(t)$ is constant, the raster plots indicating the firing of the 50 neurons are shown in Fig. 16a for the uncoupled case and in Fig. 16b for the coupled case. The bottom of the plot (the blue band in Fig. 16a and 16b) shows the firing pattern of the neuronal population, which has been obtained by combining the firings of all the neurons. It can be seen that for the uncoupled case (Fig. 16a), the population firing show clustered behavior where multiple neurons fire in close proximity, whereas for the coupled case (Fig. 16b), the firing rates are respectively uniform indicating that the inhibitory coupling reduces the correlation between the neuronal firings.

The inhibitory coupling of neurons improves the signal-to-noise ratio of the incoming signal by employing the noise-shaping techniques, which are commonly used in sigma-delta analog-to-digital converters. The noise-shaping technique alters the spectral shape of the noise power by pushing it to high-frequency bands [82,83]. To understand the implication of the inhibitory coupling for noise-shaping, a sinusoidal input $x_i(t) = A_0 \sin(2\pi f_0 t)$ at frequency $f_0 = 1$ kHz is applied to all

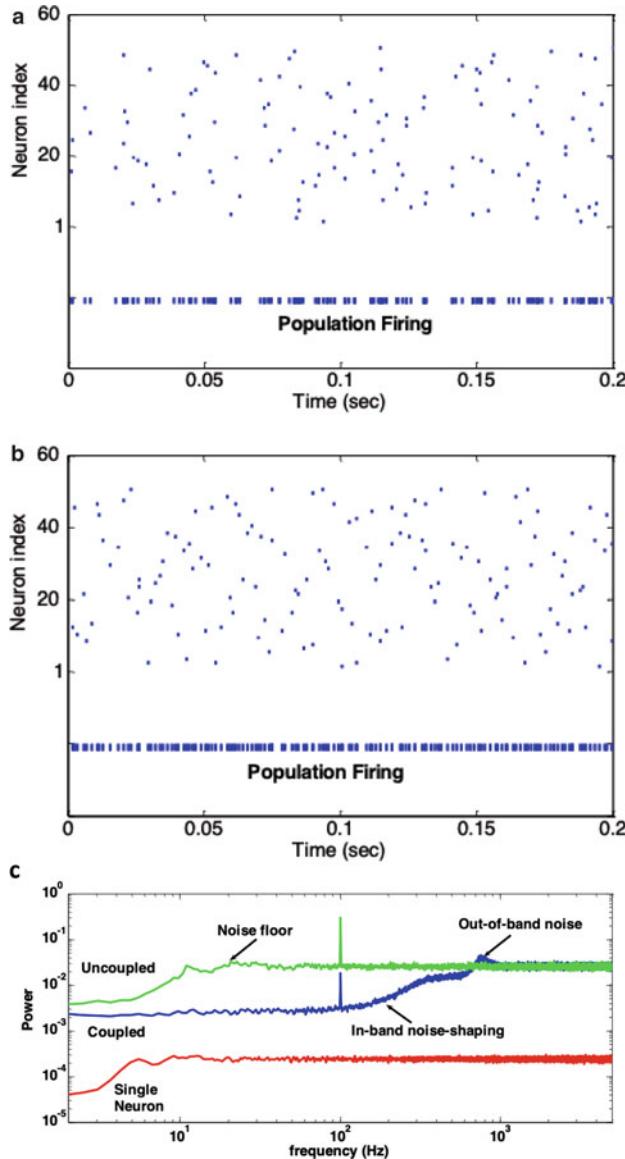


Fig. 16 Raster plot corresponding to the (a) uncoupled and (b) coupled case along with their spectra (c)

the neurons, and the population firing rates are analyzed in the frequency domain using a short Fourier transform. Figure 16c shows a comparison of the power spectrum obtained for a coupled, an uncoupled, and a single neuron. The spectrum corresponding to a single neuron shows that it cannot track the input signal since

its bandwidth (1 KHz) is much larger than the firing rate of the neuron, whereas, for the uncoupled/coupled case, the input signal can be easily seen. For the uncoupled case, the noise floor, however, is flat, whereas, for the coupled case, the noise from the signal band is shifted outside (shown in Fig. 16c). The shaping of the in-band noise-floor enhances the signal-to-noise ratio of the network, and it was shown that for a large network, the improvement is directly proportional to the number of neurons. **However, exploiting the full potential of noise-shaping in a spiking neural network requires the appropriate choice of the learning parameters K_{ij} (excitatory and inhibitory) that can take advantage of the spectral properties of the input signal.** Even though several learning algorithms have been proposed for spiking neural network for supervised and unsupervised learning, none of them exploit the spectral properties of the input signal. Also, in a usual setting, the neural network could receive diverse input signals (rather than a common input), which motivates investigation into a generic framework of spiking neural network learning algorithms combining noise-shaping principles.

4.2 Adaptive Analog-to-Digital Converter Based on Spiking Neural Networks

The mathematical model corresponding to the integrate-and-fire network given in equation 11 can be approximated in its vector differential form as $\frac{dV}{dt} = -\mathbf{K}(t)\mathbf{d}(t) + \mathbf{x}(t)$, where $\mathbf{K}(t)$ denotes the synaptic weight matrix. We have also ignored the leakage term in this formulation since silicon implementation has superior integration properties than biological neurons. Assuming that the inverse of the parameter \mathbf{K}^{-1} exists and varies slowly with respect to time (this can be guaranteed by imposing mild structural and convergence constraints), the network differential equation can be expressed as $\frac{d\mathbf{K}^{-1}\mathbf{V}}{dt} = -\mathbf{d}(t) + \mathbf{K}^{-1}(t)\mathbf{x}(t)$ which after a change of variables $\mathbf{w}(t) = \mathbf{K}^{-1}(t)\mathbf{v}(t)$, $\mathbf{A}(t) = \mathbf{K}^{-1}(t)$ leads to

$$\frac{d\mathbf{w}}{dt} = -\mathbf{d}(t) + \mathbf{A}(t)\mathbf{x}(t) \quad (12)$$

Equation 12 can then be expressed as the time-domain evolution of a min-max optimization function given by

$$\max_{\mathbf{A}} \min_{\mathbf{w}} \|\mathbf{w}\|_1 - \mathbf{w}^T \mathbf{A} \mathbf{x} \quad (13)$$

where the vector $w \in R^M$ denotes an internal state vector of the network. The L1 norm in the objective function (11) denotes a regularization factor [84, 85] that penalizes large excursion of the vector \mathbf{w} . The minimization step in (11) will ensure that the state vector \mathbf{w} correlates with the transformed input signal $\mathbf{A}^T \mathbf{x}$ (tracking step), and the maximization step in (11) will be used to learn the parameters of the linear transformation \mathbf{A} such that it minimizes the correlation (de-correlation step).

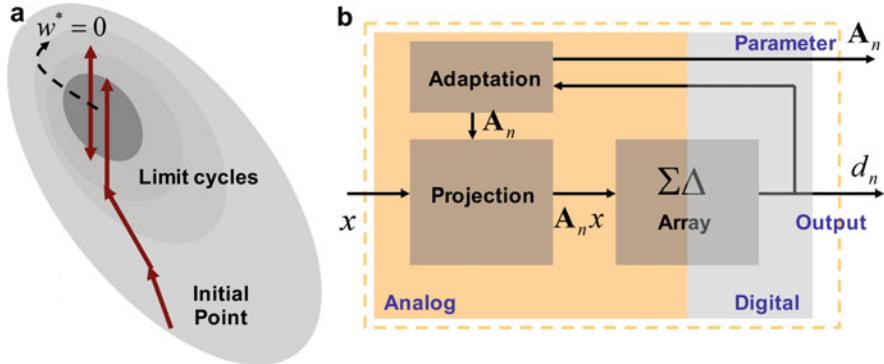


Fig. 17 (a) Concept behind on-line delta-sigma learning where the objective is to induce bounded limit cycles about an optimal point. The statistical property of the oscillations will encode information embedded in the high-dimensional manifold. (b) The architecture of delta-sigma neural network

The approach bears similarities with game-theoretic techniques where tracking and de-correlation have been formulated as conflicting objectives.

The uniqueness of the formulation as compared to other neural network learning algorithms is the use of $L1$ regularization in Equation 13 for generating bounded delta-sigma limit cycles. This is illustrated in Fig. 17a which shows a two-dimensional contour for the optimization function in Eq. 13. For a bounded transform $\|\mathbf{A}\|_1 \leq \mathbf{M}$, the solution to Eq. 13 is well defined and is given by $\mathbf{w}^* = 0$ (see Fig. 17a). In the proposed learning approach, only the path to the final solution \mathbf{w}^* and statistical property of the limit cycles about the solution are of importance since it encodes the temporal dynamics of the spike train (digital pulses). The gradient descent step for the inner-loop minimization in Eq. 13 is given by

$$\mathbf{w}_n = \mathbf{w}_{n-1} + (\mathbf{A}_{n-1}\mathbf{x} - \mathbf{d}_n) \quad (14)$$

where $\mathbf{d}_n = sgn(\mathbf{w}_{n-1})$ is the binary valued spike output obtained using the $L1$ norm of the state vector w . Equation 14 is equivalent to a first-order delta-sigma modulation and hence inherits its noise-shaping characteristics. As the recursion (14) progresses, a binary limit cycle about the solution w^* is generated which is shown in Fig. 17a. However, these limit cycles are bounded and $\|\mathbf{w}_n\|_1 \leq 2\mathbf{M}$ for all n , and the asymptotic behavior of recursion (14) can be expressed as

$$E_n\{\mathbf{d}_n\} \xrightarrow{n \rightarrow \infty} \mathbf{A}_\infty \mathbf{x} \quad (15)$$

where the expectation operator $E_n\{\cdot\}$ is with respect to time instants n . Equation 15 shows that the mean of the binary sequence \mathbf{d}_n asymptotically approaches the desired transformation where the linear transform is denoted by \mathbf{A}_∞ . Hence, the binary sequence \mathbf{d}_n asymptotically encodes the transformed signal with infinite

precision. Updates for matrix \mathbf{A} are determined by the maximization step of Eq. 13 and is given by a gradient ascent as

$$\mathbf{A}_n = \mathbf{A}_{n-1} - 2^{-P} \mathbf{d}_n \Psi(\mathbf{x})^T \quad (16)$$

where $\Psi : M \rightarrow M$ being any monotonic function and P being an update parameter that determines the resolution of the manifold. If the transformation \mathbf{A} is bounded, recursion (16) asymptotically leads to

$$E_n\{\mathbf{d}_n \Psi(x_n)\} \xrightarrow{n \rightarrow \infty} 0 \quad (17)$$

which shows that the learning algorithm converges to a manifold with parameters \mathbf{A}_∞ such that the binary sequence produced by recursion (14) is orthogonal to a set of functions Ψ . Equations 14 and 16 will be the basis of delta-sigma learning algorithms proposed in this work. **Update given by Equation 16 bears similarity with gradient-based rules used in neural network algorithms except for binary sequences \mathbf{d}_n , which in spike-based learning will embed noise-shaping characteristics.** The proposed formulation leads to a high-dimensional analog-to-digital converter whose architecture is shown in Fig. 5b. It consists an analog projection A followed by a delta-sigma array and an adaptation unit which implement the recursions (14) and (16).

4.3 Signal De-Correlation and Network Stability

The basic learning algorithm given by Equations 14 and 16 has been validated in [53] using multi-channel neural data recorded from the dorsal cochlear nucleus in adult guinea pigs. For this preliminary experiment, eight channels of neural data were used, and the manifold parameter matrix \mathbf{A} was constrained to a lower-triangular form. The data were recorded at a sampling rate of 20 KHz and at a resolution of 16 bits. Figure 18a shows a clip of multi-channel recording for duration of 0.5 s. It can be seen from the highlighted portion of Fig. 18a that the data exhibits a high degree of cross-channel correlation. Therefore, localization of individual spikes across the channel is difficult.

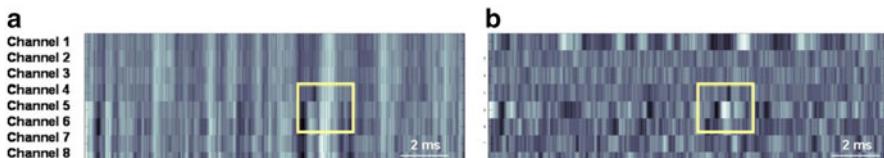
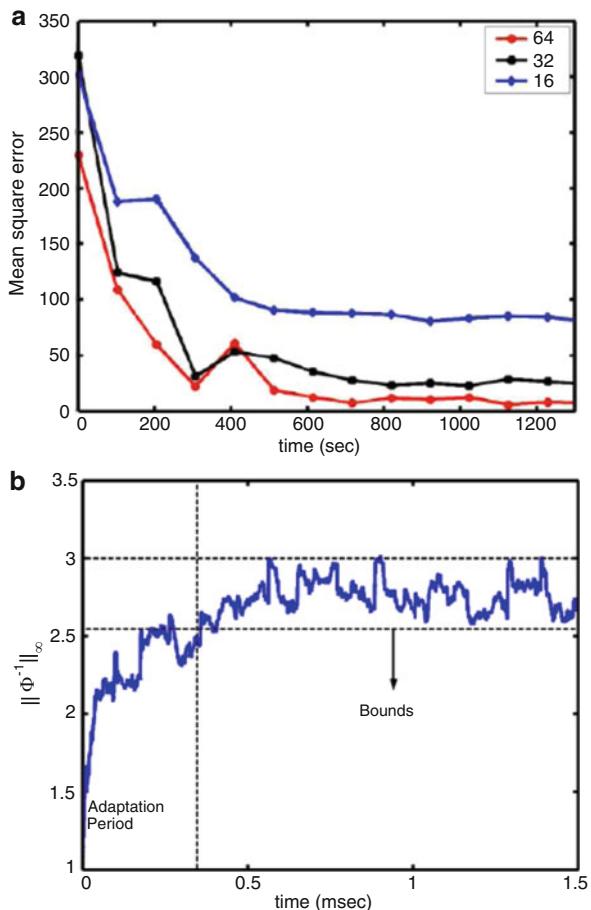


Fig. 18 Preliminary results using a recorded eight-channel neuronal data from the dorsal cochlear nucleus of a guinea pig [53]: (a) the x-axis denotes time, and the gray scale represents the amplitude of the input signal. (b) Output produced by the learning algorithm

The recursions (14) and (16) were applied to the eight-dimensional inputs \mathbf{x} , and a running average is computed using the digital sequence produced by recursion (14) and is plotted in Fig. 18b. It can be seen that the on-line learning algorithm determines the spatial distribution of neuronal signal across channels and, in the process, localizes the spikes in the multi-channel data. The convergence of the recursions (14) and (16) can be verified by observing the norm of the parameter matrix as shown in Fig. 19b. During the learning (adaptation) period, the algorithm learns the spatial distribution of the neural signals and then subsequently tracks any drift in spatial distribution. To verify that the learned transformation is information-preserving, the original signal is reconstructed using the parameter matrix as $\hat{\mathbf{x}} = \mathbf{A}^{-1} \frac{1}{N} \sum_{n=1}^N \mathbf{d}_n$. Figure 19a shows plots of the mean square error between the original input signal and the reconstructed signal, showing that the error asymptotically decreases with final reconstruction error determined by the size of the reconstruction window N (oversampling rate). Thus, the recursions (14) and (16)

Fig. 19 (a) Mean square error between reconstructed multi-channel data and original neural data. (b) Norm of the parameter matrix demonstrating bounded limit cycles



are information-preserving and de-correlated the high-dimensional input signal in real time.

The salient outcomes of this preliminary study were: (a) Within the constraints imposed on the matrix (**A**), delta-sigma learner can estimate the parameters of the manifold, such that it captures low-dimensional information embedded in high-dimensional neural data. (b) After the learner has estimated the model parameters based on input data, it tracks drift in the statistical distribution in the input signal and accordingly updates the network parameters. (c) The performance of learning is dependent on the oversampling rate (ratio of recursion rate to Nyquist rate) but asymptotically converges to a constant error. Prior knowledge is embedded into the structure of the parameter matrix **A** which also affects the speed of the delta-sigma learning algorithm.

4.4 Neuromorphic Olfaction: Dimensionality Reduction

The sorting of olfactory sensory neuron output wires (i.e., their axons) such that sensors of the same type send input to the same region has inspired the development of a dimensionality reduction approach [60, 86, 87]. In this approach, the response of a chemosensor to a panel of odorants was used to define its selectivity profile. Chemosensors with similar selectivity profiles or vectors were clustered (i.e., group similar sensors) using a topology-preserving self-organized maps. This resulted in a 2D map of response profiles (Fig. 20). As the information from multiple redundant sensors was integrated, the SNR and, therefore, odor separability improved as a result of reorganizing sensor input in this fashion [87].

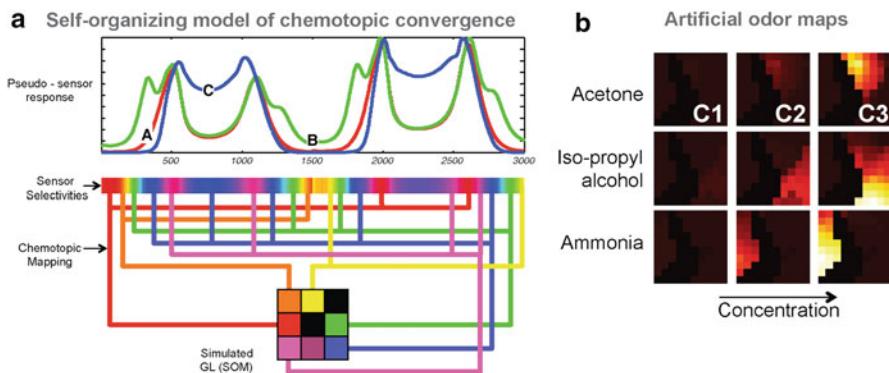


Fig. 20 (a) A schematic of the approach followed to mimic clustering of sensory neuron axons at the input of the olfactory bulb [60, 86, 87]. Responses of similar sensors, i.e., have similar responses to the odor panel (aligned selectivity vector), were grouped using a self-organizing map to create a 2D spatial image for each odorant. (b) Artificial odor maps created from an array of chemosensors are shown. Each row corresponds to a single odorant, and each column represents a particular concentration of that odorant

4.5 Neuromorphic Olfaction: Gain Control

Neuronal circuits at the input of the olfactory bulb have inspired dynamic, recurrent neural networks to perform gain control computation. To model the periglomerular circuits in the olfactory bulb, a model of lateral shunting inhibition was proposed [88]. In this model, the activity of neuron x_i was modeled as follows:

$$\frac{dx_j^0(t)}{dt} = \underbrace{-Dx_j^0(t)}_{\text{neuron dynamics}} + \underbrace{(B - x_j^0(t))G_j^o}_{\text{self excitation}} - \underbrace{x_j^0(t) \sum_{k \neq i} c_{ki} G_k^o}_{\text{shunting lateral inhibition}} \quad (18)$$

where D relates inversely to decay constant, B represents saturation response level, G_i represents the input received by the i th neuron, and the matrix C represents the connectivity matrix (with c_{ki} being the connection weights between the k th and the i th neurons). It can be easily shown that the steady-state response of neuron x_i becomes

$$x_i^0 = \frac{BG_i^0}{D + G_i^0 + \sum_{k \neq i} c_{ki} G_k^0} \quad (19)$$

For parameters $B = 1$, $D = 0$ and $c_{ki} = 1 \forall k, i$ this converges to the L1-norm of the input received. Representative results from this model are shown in Fig. 22. Note that the responses of a chemiresistive sensory array to three odorants each delivered at three different concentrations are shown. As can be noted, when lateral shunting inhibition is removed ($c_{ki} = 0 \forall k, i$), the variations due to concentrations become dominant (Fig. 21a). With global shunting inhibition ($c_{ki} = 1 \forall k, i$), the concentration information is squashed (Fig. 21c). Notably, by controlling the extent or spread of lateral inhibition, the amount of variation introduced by odor intensity-related information can be controlled (Fig. 21b).

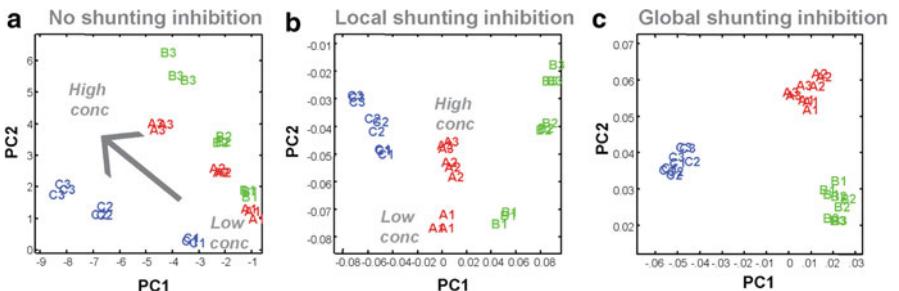


Fig. 21 Results from a shunting inhibition model for gain control [88]. Multivariate sensor responses were used as inputs to a network of neurons modeled using Eq. 18. The steady-state network activity is shown for three odorants each presented at three different concentrations. **(a)** Output of a network of neurons with no lateral shunting inhibition ($c_{ki} = 0 \forall k, i$) is shown. **(b)** Output of the neural network with local shunting inhibition is shown. **(c)** Output of the neural network model with global shunting inhibition ($c_{ki} = 1 \forall k, i$) is shown

4.6 Neuromorphic Olfaction: Contrast Enhancement

Similarly, circuits at the output of the olfactory bulb involving principal mitral/tufted cells and inhibitory granule cells have been modeled to enhance separability between odor signals [89]. Unlike the shunting model with multiplicative interaction between neurons in the circuit, here the lateral interactions were modeled as additive connections as follows:

$$\frac{dm_j(t)}{dt} = \underbrace{-\frac{m_j(t)}{\tau_j}}_{\text{neuron dynamics}} + \underbrace{\sum_{k=1}^M L_{kj}\phi(m_k(t))}_{\text{lateral inhibition}} + \underbrace{G_j}_{\text{ORN input}} \quad (20)$$

where τ_j is the time constant for decay of response of neuron j , L_{ki} indicates connection strength between neurons k and i , $\phi()$ is a logistic function, and G_j is the input received by the neuron. Note that the connectivity matrix was modeled to mediate center surround-type interactions.

The time evolution of the neural responses in this network to different odorants revealed an initial transient phase that settled into a steady-state or a fixed-point attractor very similar to the results reported *in vivo* (Fig. 22) [90]. Notably, similar

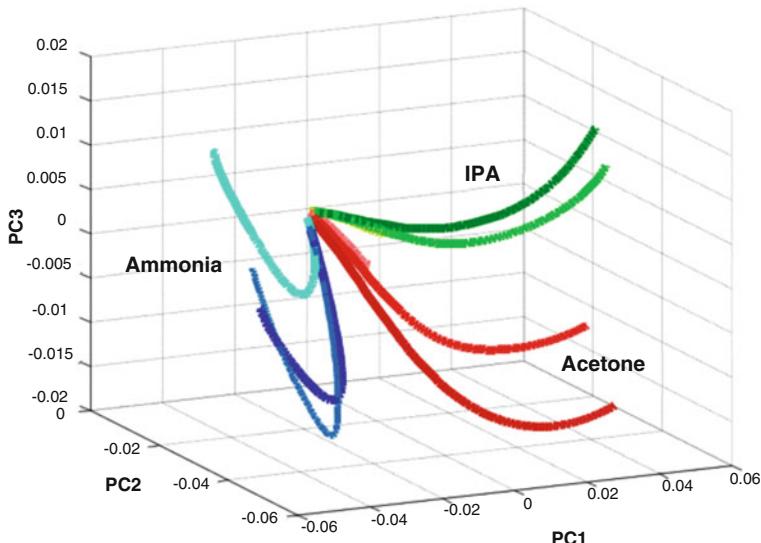


Fig. 22 Neural responses from a recurrent neural network model with additive lateral interactions (Eq. 20) are shown after PCA dimensionality reduction. Responses over time are connected to create a stimulus-evoked response trajectory. Nine trajectories are shown corresponding to sensor responses to three different odorants at three different intensities [89]. Note that the direction of the trajectories encodes stimulus identity information in an intensity-invariant fashion

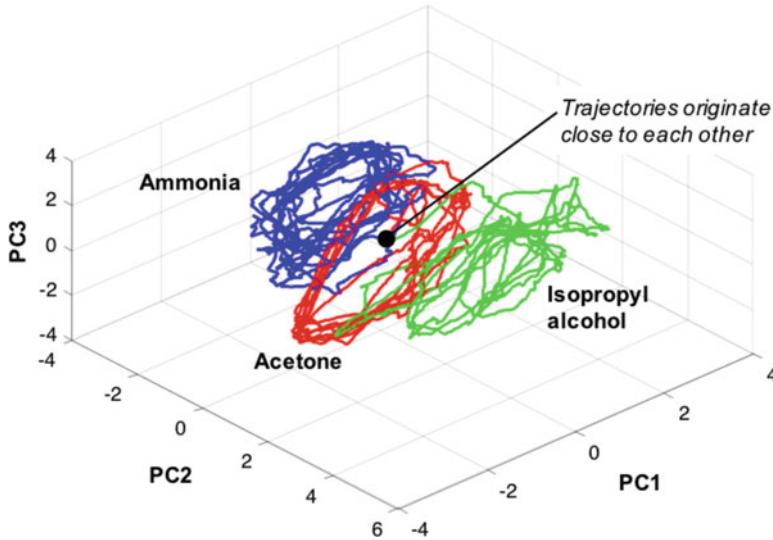


Fig. 23 Response dynamics of neural networks with short-term adaptation are shown after PCA dimensionality reduction. Neural responses that are temporally contiguous are connected to create a stimulus-evoked response trajectory and reveal how activity in the neural network evolves. Note that each odorant is now encoded by a limit-cycle attractor [92]

results were also obtained when the analog neurons in Equation 20 were replaced by integrate-and-fire spiking neurons [91].

A further extension to this model that also included a short-term response adaptation was proposed. Inclusion of adaptation term altered the odor-evoked attractors from a fixed-point dynamics to a limit-cycle one (Fig. 23).

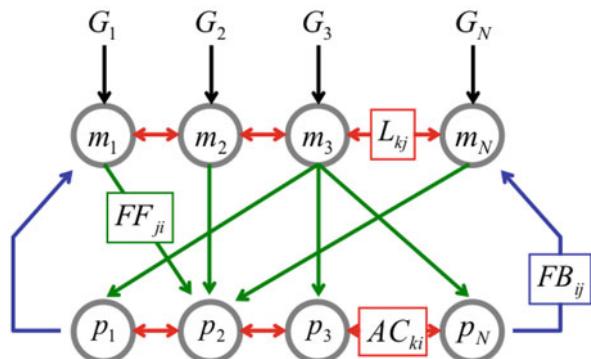
4.7 Neuromorphic Olfaction: Mixture Segmentation and Background Suppression

To process more complex signals that arise from multiple competing odorants encountered simultaneously (i.e., odor mixtures), a highly integrated model of olfactory bulb and olfactory cortex was proposed (Equations 21, 22) [60, 93] (Fig. 24).

$$\frac{dM}{dt} = \underbrace{-\frac{1}{\tau} + L\phi(M) + G}_{\text{same as before}} + \underbrace{FB\phi(P)}_{\text{Cortex-to-bulb-feedback}} \quad (21)$$

$$\frac{dM}{dt} = \underbrace{-\frac{1}{\lambda} + AC\phi(P)}_{\text{Cortico-cortical connections}} + \underbrace{FF\phi(P)}_{\text{Feedforward Connections}} \quad (22)$$

Fig. 24 A schematic of the olfactory bulb-olfactory cortex model. Note that the model is driven by inputs from a chemosensor array



$$\begin{aligned} \mathbf{A} &= [1, 0, 0, 1, 1, 0] \\ \mathbf{B} &= [0, 1, 1, 1, 0, 0] \\ \mathbf{AB} &= [0.8, 0.5, 0.5, 0.6, 0.8, 0] \end{aligned}$$

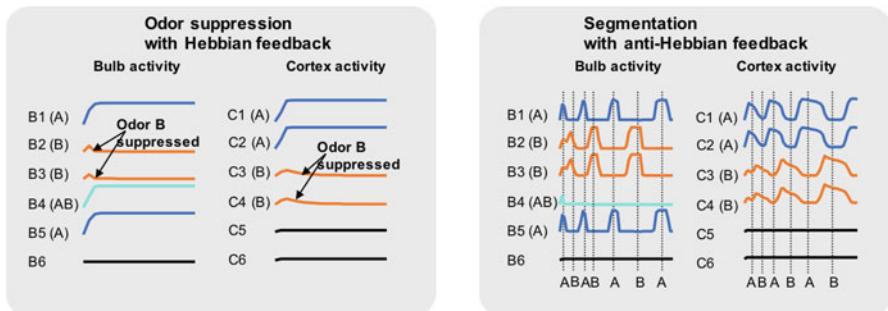


Fig. 25 Simulated patterns generated by two odorants A and B, and their binary mixtures are shown on the top panel. The bulb neural response and the cortical activity when Hebbian feedback connections were used are shown on the left panel. Note that the mixture response is dominated by the odorant A pattern, and therefore the weaker component B is completely suppressed in the bulb and cortex. Similar results using anti-Hebbian feedback is shown on the right panel. Note that the activity oscillates between activation that corresponded to odorant A (bulb and cortical neurons 1 and 2) and switched to activation that corresponded to odorant B (bulb and cortical neurons 3 and 4)

Cortical neurons in this model combined neural responses from the bulb (FF) in operation typically referred to as coincidence detection. The Association connection (AC) between pyramidal cortical neurons in the cortex was learned using Hebbian learning and allows for pattern completion of incomplete input from the bulb. The feedback connections (FB) from the cortex to the olfactory bulb were established using either Hebbian or anti-Hebbian learning. Note that Hebbian feedback allowed the strongest component in the odor mixture to drive the olfactory bulb and olfactory cortex neural responses (i.e., winner takes all) (Fig. 25; left panel). However, anti-Hebbian feedback suppressed the input activated in the cortex and allowed each

component of a mixture to drive the model response and, therefore, be recognized sequentially over time (Fig. 25; right panel).

4.8 Neuromorphic Olfaction: De-correlation of Signals

As mentioned earlier, evidence suggests that in certain models of biological olfaction, neural responses are refined over time such that features that are common across groups of chemicals are extracted first and odor-specific response attributes get emphasized later [94]. Inspired by this approach, analytical techniques have been proposed that can perform similar computations but to process responses from a chemosensory array (Fig. 26). This divide-and-conquer strategy was demonstrated to have several advantages. First, since only those features that have the highest SNR were utilized for making categorical decisions at each decision node, this approach was shown to be less sensitive to noise that creeps in as the chemosensors age over time. Second, since the nodes that are in the higher levels are common to multiple analyses, it provides a generalizable approach where some predictions about an odorant not used during the training phase can still be made.

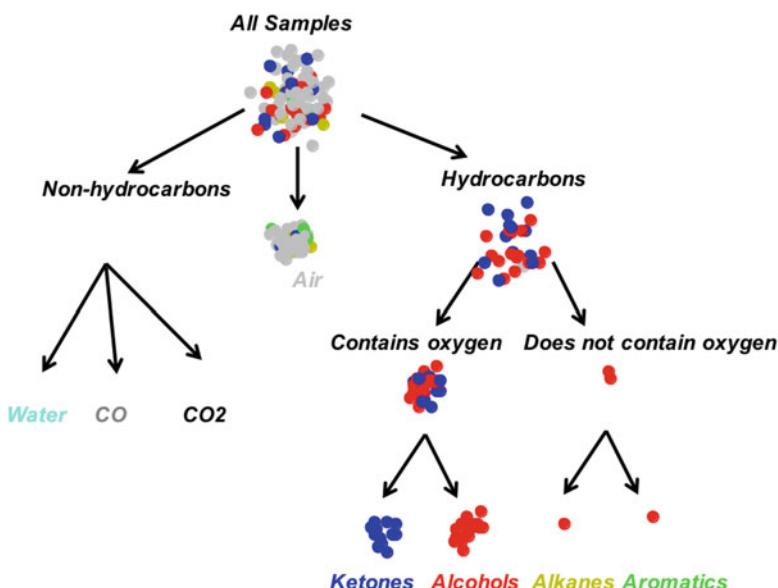


Fig. 26 Refinement of information to allow identification of generic chemical categories first, followed by more precise recognition later. Note that at each step, only those features that provide the best separation between the branching options are used. This makes this approach less sensitive to noise caused by the aging of sensors. More importantly, this approach allows generalization to chemicals not used during the training phase [94]

5 Discussion and Future Direction

Audition and olfaction are sensory modalities that have highly divergent requirements. Audition is fast and can encode more information in small time windows, while olfaction is relatively slow and is low-bandwidth. Audition is highly robust, and the recognition performance gracefully declines in noisy environments. On the other hand, olfaction is highly sensitive to changes in ambient conditions. Separating auditory signals from multiple sources (i.e., the cocktail party problem) is a feat of pattern recognition achieved daily. In contrast, odor mixtures with more than three components are difficult to segment. So, the two systems chosen for this book chapter may appear to have less in common. However, as we have discussed so far when you “look under the hood,” the computational processing principles in the two sensory systems have several things in common.

Even more, parallels can be observed when a dynamical systems perspective is followed to understand the encoding of sounds and odorants. A recent study compared what neural response features evoked by an odor and a monotone have in common [95]. It was shown that a sensory stimulus, be it an odorant or a monotone, evoked a highly transient response (i.e., ON transient) that was information-rich and activated a larger number of neurons. When the stimulus was sustained, again in both systems, the activity converged onto stable patterns of neural firings distributed in a subset of neurons that were active during the ON-transient period. This epoch is referred to as the “fixed-point” activity. The ensemble neural responses remained in this fixed-point activity until the stimulus was removed. Following stimulus termination, another round of dynamic activity (i.e., OFF transient) was observed. Intriguingly, most neurons that were active during the ON-transient phase were silenced, and many neurons silenced during the ON-transient phase became active following stimulus termination. Therefore, both stimulus presence and its absences (at least immediate following its termination) were actively encoded by two nearly orthogonal sets of neurons in both sensory systems. Whether such dynamical coding principles provide an advantage in realizing neuromorphic computational architectures still needs to be explored.

Particularly in olfaction technologies, it has been more than three decades since the concept of an electronic nose was first proposed [96]. Despite a growing need for a portable, noninvasive chemical sensing tool in several application domains (such as security, environmental monitoring, and medical diagnostics), this demand has not been met. The capabilities of the state-of-the-art chemical sensing technologies still pale in comparison with their biological counterparts. This raises a fundamental question, why can’t electronic noses as good as biological noses or even an insect antenna be engineered?

There are two main reasons that could contribute to this difference in capabilities: sensors and computations. Even a relatively simple invertebrate olfactory system has 50–100 different proteins or sensors to transduce chemical information. There is still a lack of a rich repertoire of materials that can function as chemical sensors in electronic noses. Therefore, a key area of improvement is in developing diverse

sensing materials that can provide a high-dimensional and information-rich sensory input to subsequent pattern recognition engine. Additionally, keeping these sensors stable and viable for long periods of operation (days/weeks/months/years) is a key challenge that must be met to ensure that recognition performance does not degrade over time (a task that is easier said than done!).

Second, the neural circuits that process information from olfactory receptor neurons are highly conserved across divergent species [97]. This possibly indicates that the subsequent information processing schemes have been highly optimized, and a conserved set of signal processing motifs may be used to manipulate and extract information about chemical cues in the organisms' surroundings. While these circuits are still being understood, modeling and developing neuromorphic solutions inspired by them could potentially diminish this gap in performance.

Finally, until such time when a rich repertoire of chemical transducers and highly optimized neuromorphic processing modules for chemosensory signals are available, a potential stop-gap solution that exists involves developing hybrid systems that are part-engineering and part-biology. The key idea here would let biology solve the difficult task of transducing chemical information into an electrical signal using its exquisite sensor array. Once the signal has been received and processed by the olfactory neural networks, these signals can be tapped using electrodes implanted into the brain of the organism to realize an “biosensor” [98]. The pattern of neural signals extracted can be used to determine the identity of the odorant/chemical that is in the vicinity of the organism. Such an approach can be regarded as a sophisticated version of the “canary in the coal mine” approach but one that takes advantage of the modern tools that enable sophisticated neural readouts for chemical discrimination.

Overall, in this chapter, we have discussed audition and olfaction pathways with several important neuromorphic signal processing techniques such as jump-resonance, noise shaping, dimensionality reduction, adaptive gain control, and so on. However, these case studies illustrate a few specific examples. The future challenges lie in extending these neuromorphic principles to build more complex perceptual tasks. In this regard, the futuristic goal would be to apply these concepts to emerging devices like memristor that store and process data simultaneously just like the brain does. This is also known as in-memory computing, unlike von Neuman architecture, where the logic and memory blocks are located separately.

Acknowledgments The authors would like to thank the Scheme for Promotion of Academic and Research Collaboration (SPARC), MHRD, Govt. of India, for funding (SPARC/2018-2019/P606/SL) this work. The authors acknowledge the Brain, Computation and Learning workshop at the Indian Institute of Science, India, where much of this work has been done.

References

1. Arisoy, F.D., et al.: Bioinspired photocatalytic shark-skin surfaces with antibacterial and antifouling activity via nanoimprint lithography. *ACS Appl. Mater. Interfaces* (2018). <https://doi.org/10.1021/acsmami.8b05066>
2. Barth, F.G.: Spider strain detection. In: Barth, F.G., Humphrey, J.A.C., Srinivasan, M.V. (eds.) *Frontiers in Sensing: From Biology to Engineering*, pp. 251–273. Springer, New York/Wien (2012). ISBN: 9783211997482
3. Miles, R.N., Hoy, R.R.: The development of a biologically-inspired directional microphone for hearing aids. *Audiol. Neuro-Otol.* **11**(2):86–94 (2006)
4. Sarpeshkar, R.: Neuromorphic and biomimetic engineering systems. *McGraw-Hill Yearbook of Science & Technology* 2009, pp. 250–252. McGraw-Hill, New York (2009)
5. Barth, F.G.: Spider mechanoreceptors. *Curr. Opin. Neurobiol.* **14**(4):415–422 (2004). <https://doi.org/10.1016/j.conb.2004.07.005>
6. Lichtsteiner, P., Delbrück, T.: 64 × 64 event-driven logarithmic temporal derivative silicon retina. In: *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, pp. 157–160 (2005)
7. Lichtsteiner, P., Posch, C., Delbrück, T.: A 128 × 128 120 dB 30 mW asynchronous vision sensor that responds to relative intensity change. In: *IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 2060–2069 (2006)
8. Lichtsteiner, P.: An AER temporal contrast vision sensor. Ph. D. Thesis, ETH Zurich, Department of Physics (D-PHYS), Zurich (2006))
9. Mar, D.J., Chow, C.C., Gerstner, W., Adams, R.W., Collins, J.J.: Noise shaping in populations of coupled model neurons. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10450–10455 (1999)
10. Linden, J.F., Liu, R.C., Sahani, M., Schreiner, C.E., Merzenich, M.M.: Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. *J. Neurophysiol.* **90**(4):2660–2675 (2003)
11. Aono, K., Shaga, R., Chakrabarty, S.: Exploiting jump-resonance hysteresis in silicon cochlea for extracting speaker discriminative formant trajectories. *IEEE Trans. Biomed. Circuits Syst.* **7**(4):389–400 (2013)
12. Alتوé, A., Pulkki, V.: Transmission line cochlear models: improved accuracy and efficiency. *J. Acoust. Soc. Am.* **136**:302–308 (2014). <https://doi.org/10.1121/1.4896416>
13. Glasberg, B.R., Moore, B.C., Nimmo-smith, I.: Comparison of auditory filter shapes derived with three different maskers maskers. *J. Acoust. Soc. Am.* **75**:536 (1984). <https://doi.org/10.1121/1.390487>
14. Patterson, R.D., Unoki, M., Irino, T.: Extending the domain of center frequencies for the compressive gammachirp auditory filter. *J. Acoust. Soc. Am.* **114**, 1529–1542 (2003). <https://doi.org/10.1121/1.1600720>
15. Lyon, R.F., Katsiamis, A.G., Drakakis, E.M.: History and future of auditory filter models. In: *IEEE International Symposium on Circuits and Systems (ISCAS)* (Paris), pp. 3809–3812 (2010). <https://doi.org/10.1109/ISCAS.2010.5537724>
16. Wang, S., Koickal, T.J., Hamilton, A., Cheung, R., Smith, L.S.: A bio-realistic analog CMOS cochlea filter with high tunability and ultra steep roll-off. *IEEE Trans. Biomed. Circ. Syst.* **9**, 297–311 (2015). <https://doi.org/10.1109/TBCAS.2014.2328321>
17. Lyon, R.F.: Human and Machine Hearing -Extracting Meaning from Sound. Cambridge University Press, Mountain View (2017)
18. Xu, Y., Thakur, C.S., Singh, R.K., Hamilton, T.J., Wang, R.M., van Schaik, A.: A FPGA implementation of the CAR-FAC cochlear model. *Front. Neurosci.* **12**:198 (2018)
19. Lyon Richard, F., Moore, Brian CJ.: Human and Machine Hearing: Extracting Meaning from SoundHuman and Machine Learning: Extracting Meaning from Sound, 567pp. Cambridge University Press, Cambridge (2017). Price:\$ 54.99. ISBN: 978-1-107-007536. *Acoust. Soc. Am. J.* **144**:567 (2018)

20. Xu Y., Thakur, C.S., Hamilton, T.J., Wang, R., van Schaik, A.: An FPGA implementation of the CAR-FAC cochlear model. *Front. Neurosci.* **12**:198 (2018)
21. Thakur, C.S., et al.: FPGA implementation of the CAR model of cochlea. In: IEEE International Symposium on Circuits and Systems (ISCAS) (2014)
22. Lyon, R.F.: Human and Machine Hearing -Extracting Meaning from Sound. Cambridge University Press, Mountain View (2017)
23. Thakur, C.S., Hamilton, T.J., Tapson, J., van Schaik, A., Lyon, R.F.: FPGA implementation of the CAR model of the cochlea. In: 2014 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1853–1856 (2014)
24. Thakur, C.S., Hamilton, T.J., Tapson, J., van Schaik, A., Lyon, R.F.: Live demonstration: FPGA implementation of the CAR model of the cochlea. In: 2014 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1853–1856 (2014)
25. Xu, Y., Thakur, C.S., et al.: Electronic cochlea: car-FAC model on FPGA. In: IEEE Biomedical Circuits and Systems Conference (BioCAS 2016) (2016)
26. Fukuma, A., Matsubara, M.: Jump resonance criteria of nonlinear control systems. *IEEE Trans. Autom. Control* **11**(4):699–706 (1966)
27. Martin, P., Mehta, A.D., Hudspeth, A.J.: Negative hair-bundle stiffness betrays a mechanism for mechanical amplification by the hair cell. *Proc. Natl. Acad. Sci.* **97**(22):12026–12031 (2000)
28. Aono, K., Shaga, R.K., Chakrabarty, S.: Exploiting jump-resonance hysteresis in silicon auditory front-ends for extracting speaker discriminative formant trajectories. *Biomed. Circuits Syst. IEEE Trans.* **PP**(99):1 (2013)
29. Aono, K., Shaga, R.K., Chakrabarty, S.: Exploiting jump-resonance hysteresis in silicon cochlea for formant trajectory encoding. In: 2012 IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS), pp. 85–88, 5–8 Aug 2012
30. Moni, R.S., Rao, K.R.: Jump-phenomenon in active-RC filters. *IEEE Trans. Circuits Syst.* **29**(1):54–55 (1982)
31. Aono, K., Shaga, R., Chakrabarty, S.: Exploiting jump-resonance hysteresis in silicon cochlea for extracting speaker discriminative formant trajectories. *IEEE Trans. Biomed. Circuits Syst.* **7**(4):389–400 (2013)
32. Chakrabarty, S., Cauwenberghs, G.: Gini-support vector machine: quadratic entropy based multi-class probability regression. *J. Mach. Learn. Res.* **8**:813–839 (2007)
33. Fazel, A., Chakrabarty, S.: An overview of statistical pattern recognition techniques for speaker verification. *IEEE Circuits Syst. Mag.* **11**(2):62–81 (2011)
34. Aono, K., Shaga, R., Chakrabarty, S.: Exploiting jump-resonance hysteresis in silicon cochlea for extracting speaker discriminative formant trajectories. *IEEE Trans. Biomed. Circuits Syst.* **7**(4):389–400 (2013)
35. Chi, T., Ru, P., Shamma, S.A.: Multiresolution spectrotemporal analysis of complex sounds. *J. Acoust. Soc. Am.* **118**(2):887–906 (2005). <https://doi.org/10.1121/1.1945807>
36. Patil, K., et al.: Music in our ears: the biological bases of musical timbre perception. *PLoS Comput. Biol.* **8**(11):e1002759 (2012)
37. Amari, S., Cichocki, A., Yang, H.H.: A new learning algorithm for blind signal separation. In: Advanced Neural Information Processing Systems (NIPS), vol. 8, pp. 757–763. MIT Press, Cambridge, MA (1996)
38. Chakrabarty, S., Deng, Y., Cauwenberghs, G.: Robust speech feature extraction by growth transformation in reproducing Kernel Hilbert space. *IEEE Trans. Speech Lang. Acoust.* **15**(6):1842–1849 (2007)
39. Patterson, R., Moore, B.: Auditory filters and excitation patterns as representations of frequency resolution. *Freq. Sel. Hearing* **363**:123–177 (1986)
40. Smith, E.C., Lewicki, M.S.: Efficient auditory coding. *Nature* **439**:978–982 (2006)
41. Chi, T., Ru, P., Shamma, S.: Multiresolution spectrotemporal analysis of complex sounds. *J. Acoust. Soc. Am.* **118**:887–906 (2005)
42. Chi, T., Ru, P., Shamma, S.: Multiresolution spectrotemporal analysis of complex sounds. *J. Acoust. Soc. Am.* **118**:887–906 (2005)

43. Wessinger, C.M., VanMeter, J., Tian, B., Lare, J.V., Pekar, J., Rauschecker, J.P.: Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J. Cogn. Neurosci.* **13**:1–7 (2001)
44. Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J.T., Hickok, G.: Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cereb. Cortex* **20**:2486–2495 (2010)
45. Boemio, A., Fromm, S., Braun, A., Poeppel, D.: Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.* **8**:389–395 (2005)
46. Fazel, A., Chakrabarty, S.: Sparse auditory reproducing kernel (SPARK) features for noise-robust speech recognition. *IEEE Trans. Audio Speech Lang. Process.* **20**(4) (2012). <https://doi.org/10.1109/TASL.2011.2179294>
47. Riesenhuber, M., Poggio, T.: Hierarchical models of object recognition in cortex. *Nature Neurosci.* **2**:1019–1025 (1999)
48. Wahba, G.: Splines Models for Observational Data. Series in Applied Mathematics, vol. 59. SIAM, Philadelphia (1990)
49. Girosi, F., Jones, M., Poggio, T.: Regularization theory and neural networks architectures. *Neural Comput.* **7**:219–269 (1995)
50. Fazel, A., Chakrabarty, S.: Sparse auditory reproducing kernel (SPARK) features for noise-robust speech recognition. *IEEE Trans. Audio Speech Lang. Process.* **20**(4) (2012). <https://doi.org/10.1109/TASL.2011.2179294>
51. Hirsch, H.G., Pearce, D.: The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In: Proceedings of ASR, pp. 181–188 (2000)
52. Mar, D.J., Chow, C.C., Gerstner, W., Adams, R.W., Collins, J.J.: Noise shaping in populations of coupled model neurons. *Proc. Natl. Acad. Sci. U. S. A.* **96**:10450–10455 (1999)
53. Gore, A., Chakrabarty, S.: Large Margin Analog-to-digital converters with applications in Neural Prosthetics. In: Advances in Neural Information Processing Systems (NIPS) (2006)
54. Pearce, T.C.: Computational parallels between the biological olfactory pathway and its analogue ‘the electronic nose’: Part I. Biological olfaction. *Biosystems* **41**:43–67 (1997)
55. Pearce, T.C.: Computational parallels between the biological olfactory pathway and its analogue ‘the electronic nose’: Part II. Sensor-based machine olfaction. *Biosystems* **41**:69–90 (1997)
56. Raman, B., Stopfer, M., Semancik, S.: Mimicking biological design and computing principles in artificial olfaction. *ACS Chem. Neurosci.* **2**:487–499 (2011)
57. Persaud, K., Dodd, G.: Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. *Nature* **299**:352–355 (1982)
58. Nagle, T., Schiffman, S.S., Gutierrez-Osuna, R.: The how and why of electronic noses. *IEEE Spectr.* **35**(9):22–34 (1998)
59. Pearce, T.C., et al.: Handbook of Machine Olfaction: Electronic Nose Technology. Wiley, Wiley-VCH Verlag GmbH & Co. KGaA (2006)
60. Raman, B.: Sensor-based machine olfaction with neuromorphic models of the olfactory system. Ph. D. Dissertation, Texas A&M University, College Station (2005)
61. Ache, B.W., Young, J.M.: Olfaction: diverse species, conserved principles. *Neuron* **48**:417–430 (2005)
62. Hallem, E.A., Carlson, J.R.: Coding of odors by a receptor repertoire. *Cell* **125**:143–160 (2006)
63. Uchida, N., Poo, C., Haddad, R.: Coding and transformations in the olfactory system. *Ann. Rev. Neurosci.* **37**:363–385 (2014)
64. Farivar, S.: Cytoarchitecture of the locust olfactory system. In: Biology. California Institute of Technology (2005). <https://doi.org/10.7907/4Y60-KH68>
65. Belluscio, L., et al.: A map of pheromone receptor activation in the mammalian brain. *Cell* **97**(2):209–220 (1999)
66. Laurent, G.: A systems perspective on olfactory coding. *Science* **286**:723–728 (1999)

67. Pearce, T.C., et al.: Robust stimulus encoding in olfactory processing: hyperacuity and efficient signal transmission. In: *Emergent Neural Computational Architectures Based on Neuroscience*, Springer, Berlin, Heidelberg, pp. 461–479 (2001)
68. Carandini, M., Heeger, D.J.: Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**:51–62
69. Olsen, S.R., Bhandawat, V., Wilson, R.I.: Divisive normalization in olfactory population codes. *Neuron* **66**:287–299 (2010)
70. Yakus, E., Wilson, R.: Electrical coupling between olfactory glomeruli. *Neuron* **67**:1034–1047 (2010)
71. Zhu, P., Frank, T., Friedrich, R.W.: Equalization of odor representations by a network of electrically coupled inhibitory interneurons. *Nat. Neurosci.* **16**(11):1678–1686 (2013)
72. Friedrich, R.W., Laurent, G.: Dynamic optimization of odor representations by slow temporal patterning of mitral cell activity. *Science* **291**:889–894 (2001)
73. Wilson, C.D., et al.: A primacy code for odor identity. *Nat. Commun.* **8**:1477 (2017)
74. Nizampatnam, S., et al.: Dynamic contrast enhancement and flexible odor codes. *Nat. Commun.* **9**:3062 (2018)
75. Saha, D., et al.: A spatiotemporal coding mechanism for background-invariant odor recognition. *Nat. Neurosci.* **16**:1830–1839 (2013)
76. Rokni, D., et al.: An olfactory cocktail party: figure-ground segregation of odorants in rodents. *Nat. Neurosci.* **17**:1225–1232 (2014)
77. Khan, A.G., Thattai, M., Bhalla, U.S.: Odor representations in the rat olfactory bulb change smoothly with morphing stimuli. *Neuron* **57**(4):571–585 (2008)
78. Niessing, J., Friedrich, R.W.: Olfactory pattern classification by discrete neuronal network states. *Nature* **465**:47–52 (2010)
79. Liang, D.G.: Perception of odor mixtures. In: Doty, R.L., (ed.) *The Handbook of Olfaction and Gustation*, pp. 283–298. Marcel Dekker, New York (1995)
80. Jortner, R.A., Farivar, S., Laurent, G.: A simple connectivity scheme for sparse coding in an olfactory system. *J. Neurosci.* **14**(7):1659–1669 (2007)
81. Gerstner, W.: Population dynamics of spiking neurons: fast transients, asynchronous states, and locking. *Neural Comput.* **12**(1):43–89 (2000)
82. Chacron, M.J., Lindner, B., Longtin, A.: Noise shaping by interval correlations increases information transfer. *Phys. Rev. Lett.* **92**:080601 (2004)
83. Wiesenfeld, K., Satija, I.: Noise tolerance of frequency-locked dynamics. *Phys. Rev. B* **36**:2483–2492 (1987)
84. Gore, A., Chakrabarty, S.: Min-max optimization framework for designing sigma-delta learners: theory and hardware. *IEEE Trans. Circuits Syst. I* **57**(3):604–617 (2010). <https://doi.org/10.1109/TCSI.2009.2025002>
85. Chakrabarty, S.: Multiple-input multiple-output analog-to-digital converter US Patent no: 7,479,911, Issued 20 Jan 2009
86. Perera, A., et al.: A dimensionality-reduction technique inspired by receptor convergence in the olfactory system. *Sens. Actuat B: Chem.* **116**:17–22 (2006)
87. Raman, B., et al.: Processing of chemical sensor array with a biologically inspired model of olfactory coding. *IEEE Trans. Neural Netw.* **17**:1015–1024 (2006)
88. Raman, B., Gutierrez-Osuna, R.: Concentration normalization with a model of gain control in the olfactory bulb. *Sens. Actuat. B: Chem.* **116**:36–42 (2006)
89. Raman, B., Yamanaka, T., Gutierrez-Osuna, R.: Contrast enhancement of gas sensor array patterns with a neurodynamics model of the olfactory bulb. *Sens. Actuat. B: Chem.* **119**:547–555 (2006)
90. Stopfer, M., et al.: Odor identity vs. intensity coding in an olfactory system. *Neuron* **39**(6):991–1004 (2003)
91. Raman, B., Gutierrez-Osuna, R.: Chemosensory processing in a spiking model of the olfactory bulb: chemotopic convergence and center surround inhibition. In: *Proceedings of Advances in Neural Information Processing Systems (NIPS 2004)*, Vancouver, vol. 17 (2004)
92. Raman, B., et al.: Sensor-based machine olfaction with a neurodynamics model of the olfactory bulb. In: *IEEE Intelligence Robots and Systems*. IEEE, Sendai (2004)

93. Raman, B., Gutierrez-Osuna, R.: Mixture segmentation and background suppression in chemosensor arrays with a model of olfactory bulb-cortex interaction. In: International Joint Conference on Neural Networks. IEEE, Montreal (2005)
94. Raman, B., et al.: A bioinspired methodology for artificial olfaction. *Anal. Chem.* **80**:8364–8371 (2008)
95. Saha, D., et al.: Engaging and disengaging recurrent inhibition coincides with sensing and unsensing of a sensory stimulus. *Nat. Commun.* **8**:15413 (2017)
96. Persaud, K., Dodd, G.: Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. *Nature* **299**:352–355 (1982)
97. Ache, B.W., Young, J.M.: Olfaction: diverse species, conserved principles. *Neuron* **48**:417–430 (2005)
98. Saha, D., et al.: Explosive sensing with insect-based biorobots. *Biosens. Bioelectron.* **X** **6**:100050 (2020)