



# Unsupervised learning in images and audio to produce neural receptive fields: a primer and accessible notebook

Namratha Urs<sup>1</sup> · Sahar Behpour<sup>2</sup> · Angie Georgaras<sup>3</sup> · Mark V. Albert<sup>1,3,4</sup>

Accepted: 16 July 2021 / Published online: 19 October 2021  
© The Author(s) 2021

## Abstract

Sensory processing relies on efficient computation driven by a combination of low-level unsupervised, statistical structural learning, and high-level task-dependent learning. In the earliest stages of sensory processing, sparse and independent coding strategies are capable of modeling neural processing using the same coding strategy with only a change in the input (e.g., grayscale images, color images, and audio). We present a consolidated review of Independent Component Analysis (ICA) as an efficient neural coding scheme with the ability to model early visual and auditory neural processing. We created a self-contained, accessible Jupyter notebook using Python to demonstrate the efficient coding principle for different modalities following a consistent five-step strategy. For each modality, derived receptive field models from natural and non-natural inputs are contrasted, demonstrating how neural codes are not produced when the inputs sufficiently deviate from those animals were evolved to process. Additionally, the demonstration shows that ICA produces more neurally-appropriate receptive field models than those based on common compression strategies, such as Principal Component Analysis. The five-step strategy not only produces neural-like models but also promotes reuse of code to emphasize the input-agnostic nature where each modality can be modeled with only a change in inputs. This notebook can be used to readily observe the links between unsupervised machine learning strategies and early sensory neuroscience, improving our understanding of flexible data-driven neural development in nature and future applications.

---

✉ Namratha Urs  
namrathaur@my.unt.edu

Sahar Behpour  
sahar.behpour@unt.edu

Angie Georgaras  
aggeorgaras1@gmail.com

Mark V. Albert  
mark.albert@unt.edu

<sup>1</sup> Department of Computer Science and Engineering, University of North Texas, Denton, TX, US

<sup>2</sup> Department of Information Science, University of North Texas, Denton, TX, US

<sup>3</sup> Department of Neuroscience, Loyola University Chicago, Chicago, IL, US

<sup>4</sup> Department of Biomedical Engineering, University of North Texas, Denton, TX, US

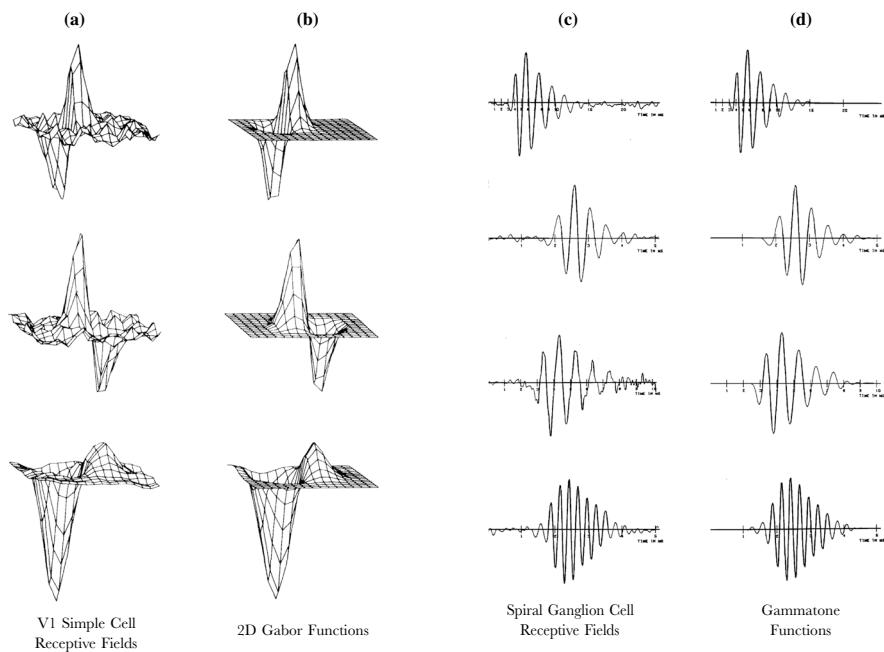
**Keywords** Neural coding · Efficient coding principle · Sensory processing

## 1 Introduction

Bridging the gap between neuroscience and computational approaches presents a mutual benefit to both neuroscientists and computer scientists. The nature of biological systems to perform with high accuracy and extraordinary efficiency in complicated and uncertain environments has led brain-inspired modeling to be a natural frame of reference for advances in Artificial Intelligence (AI) (Fong et al. 2018). Conversely, computational strategies can test and validate intuitions about brain structure and activity by explicitly modeling those intuitions. For example, early visual and auditory neural responses can be predicted using receptive field models based on stimulus–response pairs, but an understanding of the role of that receptive field model as an efficient coding strategy requires using a computational paradigm.

Receptive field models in early sensory neuroscience help understand the response properties of sensory neurons (Sherrington 1906). However, such images explain “what” stimulus drives a particular neuron’s response, but not necessarily “why” neurons would be guided by evolution and adaptation to respond this way. Early measurements of primary visual cortex (V1) simple cell responses to stimuli demonstrate response properties that can be approximated by a 2D Gabor wavelet code (Fig. 1) (Hubel and Wiesel 1962, 1968; Jones and Palmer 1987b), but why such a code among all the alternative coding strategies? The efficient coding hypothesis proposes that the goal of early sensory processing is to reduce redundancy (Barlow 1961; Field 1987). However, several objectives can be formulated from this belief. A sparse coding of grayscale natural images (Olshausen and Field 1996) first demonstrated how these early visual codes can be produced through unsupervised machine learning (Fig. 2). Furthermore, independent coding through Independent Component Analysis (ICA) (Bell and Sejnowski 1997) on natural images created similar receptive fields. In particular, only efficient encoding objectives which are appropriate for neural representations have been found to produce more efficient representations; such representations can be contrasted to compact efficient codes such as PCA or other traditional factor analysis techniques (Field 1994). It is only these neurally-appropriate efficient strategies, such as sparse coding or ICA, applied to natural images that yield filters resembling the 2D Gabor functions seen in early sensory processing.

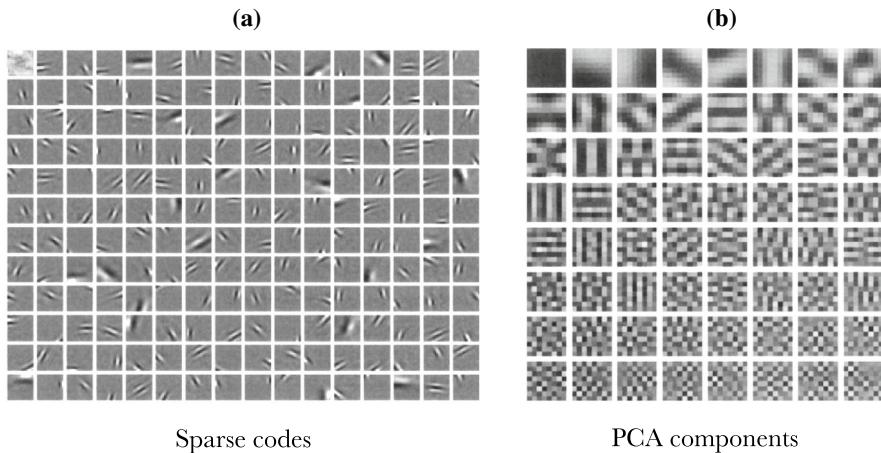
One of the powerful aspects of the efficient coding hypothesis, and its subsequent application to derive neural receptive fields directly from sensory data, is the universal nature across a variety of modalities. Grayscale natural images encoded with a sparse coding or independent coding objective produce grayscale luminance filters, however, animals experience the world also in color, over time, and even binocularly. Uniquely from a computational standpoint, each of these visual modalities can be approached by only a change in input. The application of ICA on natural video sequences results in qualitatively similar spatio-temporal properties to primary visual cortex receptive fields (van Hateren and Ruderman 1998). For example, the derived filters at low spatial frequencies were more sensitive to rapid movement than those at high spatial frequencies, which has been demonstrated in the distribution of spatio-temporal neural receptive fields in animals. Similarly, by applying ICA on color natural images as opposed to grayscale, resulting filters are color selective in similar distributions to what is observed in experimentally measured receptive fields (Fig. 3). There were more achromatic filters which have higher spatial frequencies.



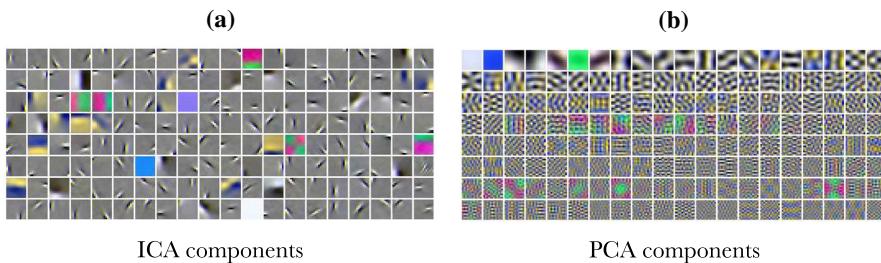
**Fig. 1** Experimentally measured models of visual (*left panel*) and auditory (*right panel*) receptive fields compared to equivalent 2D gabor wavelets and gammatone functions, respectively. Left panel: Simple cell receptive fields of neurons in the primary visual cortex (V1) of cats (**a**) and two-dimensional (2D) Gabor functions (**b**) (Jones and Palmer 1987a). 2D Gabor functions are capable of precisely capturing the spatial aspects of simple cell receptive fields. Right panel: Receptive fields of spiral ganglion cells in the cochlea of cats (**c**) and fit gammatone functions (**d**) (de Boer and de Jongh 1978). Gammatone functions precisely capture the response properties of primary auditory neurons, i.e., aspects of frequency selectivity and temporality

Color opponency also followed a pattern observed in neural receptive fields with distinctly separated red-green, blue-yellow, and bright-dark channels as observed in the distribution of receptive fields representing color (Hoyer and Hyvärinen 2000). Likewise, if binocular images are used as input to ICA, binocular receptive fields are produced (Hoyer and Hyvärinen 2000). The distribution of receptive field properties resembles what is observed in nature, including a variety of filters primarily on one of the two eyes (ocular dominance) as well as a variety of disparity shifts between the left and right eyes, representing the presence of binocular disparity. Through grayscale, video, color, and binocular representations—and potential combinations—efficient coding techniques can derive representations of receptive fields resembling those measured experimentally.

Notably, this flexibility of efficient coding strategies to derive neural receptive fields also extends to auditory processing (Lewicki 2002). Gammatone filters are a parametric model which can be used to characterize the receptive fields of spiral ganglion cells in the cochlea, similar to how 2D Gabor filters resemble V1 receptive fields. By efficiently encoding a variety of natural sounds, ICA can produce linear filters resembling gammatone filters observed in nature (Fig. 4). In this way, the same coding strategy can explain responses in a variety of visual modalities and in the auditory system as well, with only a change in input data.



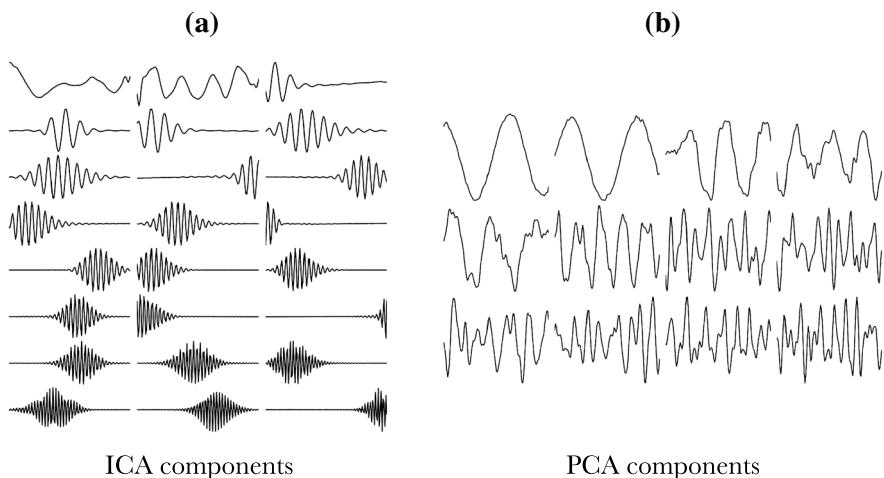
**Fig. 2** Observations of sparse coding and compact coding strategies on grayscale natural scenes. Receptive fields derived from compact coding are not localized and do not resemble the known receptive fields. Sparse coding for natural scenes yields filters that not only resemble simple-cell receptive fields but also develop their characteristic properties (i.e., spatially localized, oriented, and bandpass). **(a)** 192 basis functions as a result of training on  $16 \times 16$  monochromatic image patches from natural scenes (after preprocessing). **(b)** Principal components calculated on  $8 \times 8$  monochromatic image patches extracted from natural scenes (Olshausen and Field 1996)



**Fig. 3** Derived basis functions resulting from efficient coding of color images. The grayscale image model is extended to include colors. **(a)** Independent components (filters) derived from ICA are similar to receptive fields observed in grayscale images, most of which are achromatic, and a few others consist of low spatial frequency red-green and blue-yellow patches. **(b)** 160 principal components of the data bear no resemblance to neural receptive field-like filters (Hoyer and Hyvärinen 2000)

Evolution and adaptation developed sensory systems where the same computation is performed in both visual and auditory systems. Despite the existence of numerous studies about efficient coding applied to natural sensory data, the use of efficient coding to provide a common computational framework across modalities is not prevalent. Here, we further establish the connection between efficient coding strategies and neural receptive fields using a self-contained, easily accessible Jupyter Notebook to enable researchers from both fields.

In the available notebook, we demonstrate how the same efficient coding scheme can be used to model early sensory processing regardless of the modality (e.g., grayscale images, color images, and audio). We employ unsupervised machine learning techniques,



**Fig. 4** Auditory filters resulting from the efficient coding of natural sounds. **(a)** A representative subset of ICA-derived filters for speech in increasing order of peak resonance frequency. This representation is found to be between that of environmental sounds and animal vocalizations since speech contains both harmonic and anharmonic sounds. Efficient coding applied to a sound ensemble of environmental sounds and animal vocalizations in a 2:1 proportion yielded similar filters (like those for speech). **(b)** Representative subset of PCA-derived filters, in decreasing order of captured variance, for the same (2:1) ensemble. These filters are not localized in time with only the largest components being sinusoidal, thereby having very little relevance to physiological filters as found in the auditory system (Lewicki 2002)

specifically Independent Component Analysis (ICA) and Principal Component Analysis (PCA), to simulate a neural efficient coding objective and contrast it with a non-neural efficient coding objective, respectively. Similarly, linear filters are generated using both natural and non-natural input data to demonstrate that neural codes are a result of a neural coding objective and an appropriate natural data set, similar in structure to the sensory data that animals have evolved and adapted to process.

## 2 Neural efficient coding objectives

In a statistical sense, neural efficient coding entails transforming multivariate input data into a new, efficient representation that is able to reconstruct as much information as possible from the essential structure of the data. A code is said to be “efficient” when the new representation of the data also satisfies additional criteria beyond reconstructing a signal, such as reducing the size or statistical redundancy of the representation. From a computational standpoint, such representations can be derived by learning from raw data without respect to the tasks of the system, commonly known as unsupervised machine learning. This section provides an overview of unsupervised learning strategies that have been used in the context of neural coding; however, we begin this overview with a clear contrast to compact coding which is a common objective in applied efficient coding but does not represent the goals of most neural codes.

## 2.1 Compact coding, for comparison

Compact coding removes redundancy in the input data by reducing its dimensionality in a manner that yields minimal information loss. The input data is transformed into a representation whose dimensionality is less than the original data. For example, with binary data, the goal would be to reduce the numbers of 0's and 1's to represent the original data—a common goal in applied computing. Such compact codes can be obtained from a Principal Component Analysis (PCA) of the data. PCA is a highly versatile, unsupervised learning technique with a variance maximization criterion where the goal is to find potentially relevant factors, i.e., linear combinations of features, that best explains the variance in the data. In other words, PCA seeks to find the “hidden factors”, also known as latent variables, which would allow us to predict the feature values for individual samples if these factors were known to us. Mathematically, PCA learns a small set of components to represent the input data meaningfully, albeit these components can represent only a subset of the inputs due to the reduced dimensionality. Due to its ability to uncover structures inherent in data, PCA has been a common technique for applications such as image compression, noise reduction, visualization, and feature engineering tasks for supervised machine learning, however, as we will see the goals of compact coding differ from the objectives of most neural codes.

## 2.2 Sparse coding

The objective behind sparse coding is to represent information with as few simultaneously active neurons as possible in a large population. In a binary coding scheme, the goal would be to reduce the number of 1's in the code, rather than both 0's and 1's as in compact coding. This is justified biologically in part as neural spiking is metabolically expensive. Unlike compact codes, sparse codes are capable of producing a number of components greater than the dimensionality of the data to effectively capture higher-order statistics inherent in the data.

## 2.3 Independent coding

One important goal in encoding schemes is to identify the underlying cause or latent variables that account for variability present in the data. While compact codes such as PCA attempt this, minimizing the size of the representation creates constraints, such as forced orthogonality, that reduce the interpretability of the components and introduce high-order statistical dependence. However, unsupervised learning objectives that attempt to maximize statistical independence can be more successful and create interpretable and useful components. Independent codes can be produced using the unsupervised learning technique of Independent Component Analysis (ICA) (Comon 1994). ICA was originally developed to address the blind source separation problem (Jutten and Herault 1991) and has been particularly useful for problems with linear mixing, such as the classic cocktail party problem (Bronkhorst 2015; Cherry 1953; Haykin and Chen 2005). ICA creates components through linear combinations of features with responses that are maximally statistically independent under a specific set of assumptions. Notably, as will be discussed, ICA also commonly produces codes that are sparse, depending on the data.

## 2.4 Slow feature analysis/temporal coherence

Statistical regularities in sensory input arise as a consequence of the persistence of objects around us. Individual pixels change drastically over short time spans with typical variations such as lighting, translation, rotation, etc., however, our internal representation of the world does not vary as dramatically or quickly. A reasonable objective for coding our natural sensory experience is to bias toward more stable representations to match this reality. Since invariant aspects are critical for survival, finding meaningful representations that are not influenced by fast-changing, irrelevant information becomes crucial. Slow Feature Analysis (SFA) is an unsupervised learning algorithm with a goal to maximize the invariance in the representation over time by extracting those components that vary slowly over time from multivariate data (Wiskott and Sejnowski 2002). The resulting filters derived by SFA resemble the simple cell responses of neurons, thereby suggesting its comparability with ICA. Additionally, SFA-derived filters exhibit interesting non-linear response properties such as direction selectivity and inhibition, which are similar to the response behavior of complex cells in V1 (Berkes and Wiskott 2005). Further, under temporal constraints, SFA shares common properties with ICA (Blaschke et al. 2006).

## 2.5 Why is sparse coding more neurally appropriate?

Empirically, natural images and sounds contain many statistical dependencies beyond linear correlations, and compact coding strategies, such as PCA, do not adequately account for that higher-order statistical structure (Field 1994). PCA is limited to deriving components by maximizing the variance and successively removing the maximum variance component by forced orthogonality, but there are other useful metrics for identifying latent variables. Two latent underlying variables may be moderately correlated but PCA would be unable to capture the two latent variables without additional steps. Due to orthogonality between components and earlier components capturing the most information, interpreting and utilizing the later PCA components becomes a challenge. Additionally, these orthogonal components identified by PCA can be highly statistically dependent despite zero correlation. These concerns suggest that compact codes, such as from PCA, may not be as useful for capturing low-level statistical redundancy.

On the contrary, encoding information with sparsity brings several advantages. Individual neuronal firing is metabolically expensive, albeit common and typical. Task-level neuronal engagement becomes critical in analyzing the encoding strategy adopted by the primary visual cortex. With less than 1% of concurrently active neurons, representations that use fewer active neurons to encode sensory information become essential (Lennie 2003). Sparser codes lead to activation of a minimal number of neurons at a time which lowers energy consumption and improves metabolic efficiency while still yielding a reliable representation of the signal.

Empirical demonstrations of sparse and independent coding have been successful in creating neural receptive fields on natural images and sounds (Field 1994; Lewicki 2002). Sparse representations have succinctly accounted for receptive field properties and exhibited a higher degree of statistical independence (Olshausen and Field 1996). Resembling 2D Gabor filters, the derived sparse codes were found to be selective to location, orientation, and spatial frequency identical to the response properties of simple cell receptive fields.

Independent coding through ICA also results in linear codes that resemble neural receptive fields in the primary visual cortex (Bell and Sejnowski 1997). Notably, these receptive fields yield sparse neural responses as expected due to the similar receptive field profiles in sparse codes. ICA and sparse coding are considered to be equivalent with sparse sources; the demonstration that follows assumes sparse sources. More technically, ICA yields a model similar to sparse coding only with a super-Gaussian prior since super-Gaussian distribution is sparse. ICA produces components that are not required to be orthogonal and does not have a strict ordering as in the case of PCA. Additionally, the resulting sparse responses enable the reduction of the high metabolic cost associated with the spiking activity of a single neuron. ICA can be used to represent the goals of the first linear stage of visual and auditory processing in the brain. With the statistical independence assumption, ICA itself is a linear modeling strategy, however, there are a number of non-linear encoding strategies related to ICA; for example, topographic independent component analysis (Hyvärinen et al. 2001; Hyvärinen and Hoyer 2001).

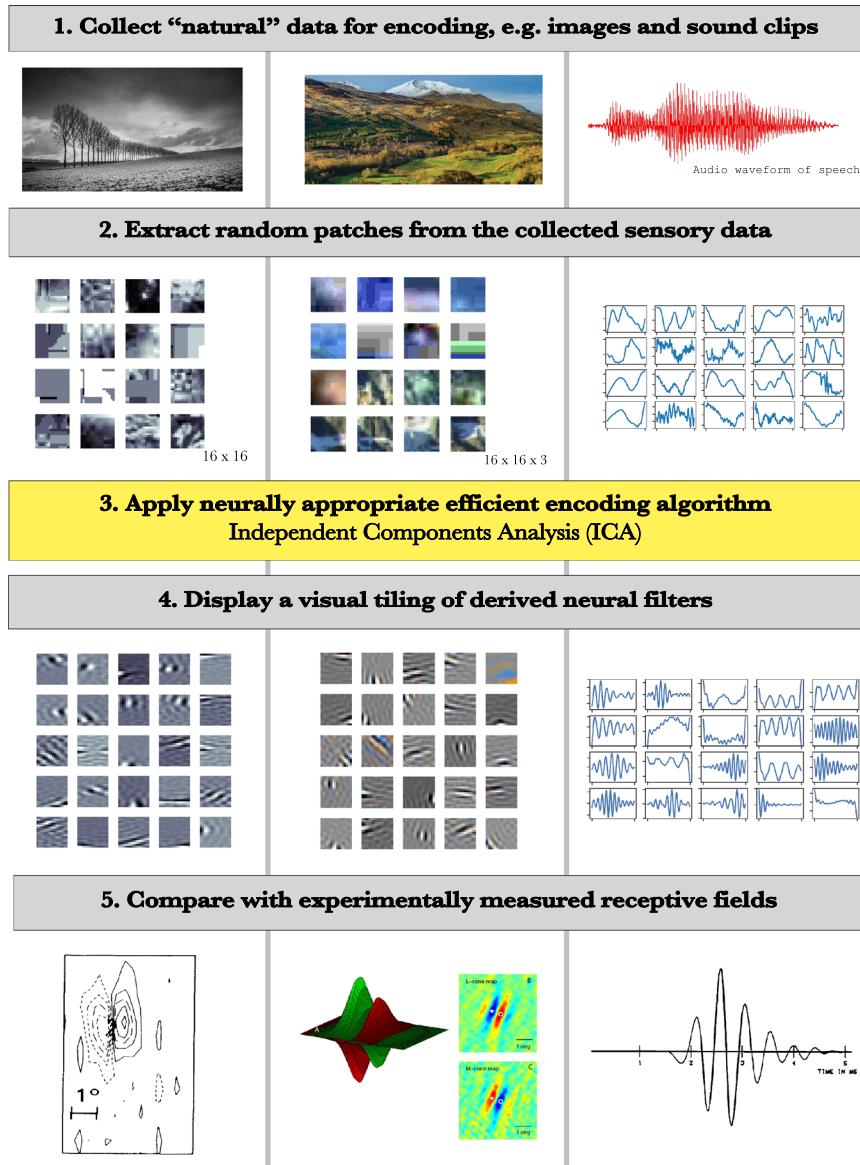
In practice, the efficient coding strategies which create V1-like receptive fields may have differing objective functions, but the end result produces filters which satisfy the goals of the other objectives. For this reason, we are not suggesting one objective above the other objectives, but rather use one of these objectives, namely ICA, as a stand-in for neural efficient coding objectives. We also contrast this objective with common efficient coding objectives that are non-neural, namely PCA, to emphasize that the precise concept of “efficiency” is critical in relating to neural coding.

### 3 Steps of efficient coding

We have made the following demonstration of the efficient coding principle accessible through a self-contained, publicly available Jupyter Notebook. In this notebook we model the sensory processing of visual and auditory modalities; specifically, grayscale images, color images, and audio. Since the efficient coding hypothesis utilizes the same algorithm regardless of the input, the computational strategy for efficient encoding remains identical irrespective of the modality being modeled. This strategy is formulated as a five-step procedure (Fig. 5) and is described below. The notebook demonstration is designed for anyone to gain direct, introductory experience in neural efficient coding.

#### 1. Collection of sensory data

As a first step, we collect data pertaining to different sensory modalities, i.e., visual and auditory. Further, for each modality, we collect natural and non-natural inputs to demonstrate the impact of the data on the presence or absence of neural codes as observed in animals. In the context of this work, the term *natural* refers to stimuli that occur in our environment and also share similar statistical properties with each other. Natural scenes are images of the visual environment in which the artifacts of civilization do not appear (Olshausen and Field 2000). For example, visual scenes such as rocks, trees, mountains, bushes, prairies, flowers, and water are considered natural. Similarly, a bird’s song, rustling leaves, and human speech are examples of natural sounds and portray the characteristics of being harmonic, anharmonic, or both, respectively. Interestingly, images of human-made structures such as buildings and man-made sounds do have similar underlying statistical structures but do not qualify as being natural since our



**Fig. 5** A five-step modality-agnostic computational strategy to model efficient coding with only a change in inputs. (1) Collect “natural” sensory data (grayscale images, color images, and audio). (2) Extract random patches from the data. (3) Apply a neurally appropriate encoding algorithm, i.e., ICA. (4) Visually tile the derived filters from the algorithm. (5) Compare the derived encodings with their corresponding experimentally measured receptive fields: grayscale (Jones and Palmer 1987a), color (Johnson et al. 2008; Shapley and Hawken 2011), and audio (de Boer and de Jongh 1978)

definition of “natural” is based on the statistical properties that leads to the robustness of data and not strictly defined by the statistics inherent in the data itself.

On the other hand, non-natural inputs are stimuli that our sensory system does not commonly observe from the environment, such as psychedelic visuals and white noise. Note that the concept of non-natural generally does not exist but is a construct we have used to refer to the category of inputs that are not considered natural.

For the purposes of our demonstration, we collected a small sample of high resolution grayscale and color images as natural visual scenes while considering psychedelic images as non-natural visual scenes. With respect to the auditory stimuli, we used a recorded version of human speech and a dog barking as natural sounds whereas we used white noise recordings as non-natural sounds. Of the many non-natural sounds possible, we selected white noise recordings for the purposes of this demonstration; colored noise is another type. Regardless of the type of non-natural auditory stimulus chosen, it is not the pairwise correlations that are important but rather the higher order statistics in the data. Colored noise does not yield Gabor-like filters with ICA, however, other non-natural patterns can. For example, amorphous blob-like patterns that resemble spontaneous neural activity in the developing visual system (Albert et al. 2008). Although the study demonstrates ICA-like results on images, the same principle applies for auditory stimuli.

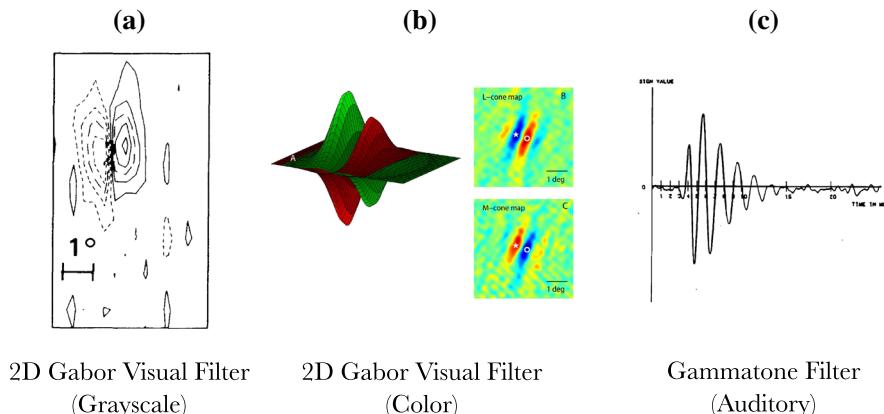
## 2. Extraction of random samples (patches)

Upon gathering data for each modality and before applying an encoding algorithm, the sensory data is preprocessed to extract smaller subsamples. For each modality, samples are randomly extracted across the dataset with a number of samples per image. Image patches and sound samples are extracted, and multidimensional samples, such as 2D image patches, or 3D image patches with color layers, are flattened into 1D vector representations to create a single `samples x features` matrix. We use 100K and 500K samples for our experiments with each modality. For the visual modality, patch widths of size 8x8 pixels and 16x16 pixels are used for both grayscale and color images. Additionally, we also specify channel information for color images (8x8x3 and 16x16x3).

Each of these pixel patches are then reshaped to a 64- or 256-dimensional vector for each grayscale image patch (alternatively, a 192- or 768-dimensional vector for each color image patch). These smaller patch sizes were chosen to keep the required computations fast and efficient to run the Jupyter Notebook with minimal memory usage on various computer platforms. Images were normalized to zero mean and unit variance before extracting pixel patches. Blank patches, as a result of random sampling of patches, were discarded. Extracted image patch samples were also normalized to zero mean and unit variance. For the audio modality, we extract 100K and 500K smaller sound clips of 100 dimensions from a sampling frequency of 44.1 kHz with downsampling at a rate of 3:1, therefore the sound clips represent approximately 7ms in length.

## 3. Application of encoding algorithms

To contrast neurally with non-neural efficient codes, we applied two unsupervised machine learning algorithms. Specifically, we use the FastICA algorithm (Hyvärinen 1999) to perform Independent Component Analysis (ICA) and Principal Component Analysis (PCA) to model the efficient coding of sensory data. The implementations of FastICA and PCA are available in scikit-learn, a machine learning library for Python (<https://www.scikit-learn.org>). We varied the number of components for ICA and PCA with the optimal value for the number of components determined on an ad hoc basis.



**Fig. 6** Experimentally measured filters from physiology, corresponding to simple cell receptive fields and spiral ganglion cell receptive fields, for visual and auditory modalities, respectively. (a) 2D Gabor filter for grayscale vision (Jones and Palmer 1987a). (b) 2D Gabor filter for color vision (Johnson et al. 2008; Shapley and Hawken 2011). (c) Gammatone filter for auditory signals (de Boer and de Jongh 1978)

#### 4. Display of resulting filters

The encoding algorithm, when applied to the collected data, yields filters. The goal of this step is to display these filters for visual inspection. In the visual tiling, the rows and columns represent the derived Gabor-like and gammatone-like filters (Fig. 5, step 4). Irrespective of the modality, the (Python) code for displaying the original extracted patches is reused for visually portraying the derived filters.

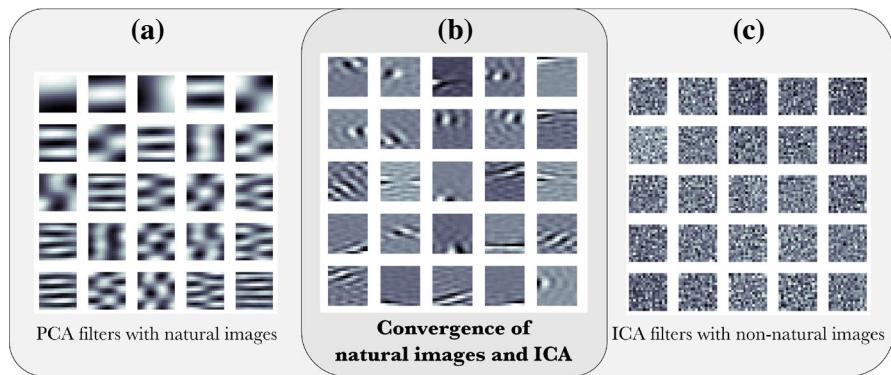
#### 5. Comparison with physiological filters

In the last step, we perform a visual comparison of the derived filters against experimentally measured receptive fields from physiology (Fig. 5, step 5). The physiological standards for receptive fields (see Fig. 6) are obtained from prior experimental neuroscience research measuring neural receptive fields. Receptive fields of simple cells in the primary visual cortex resembling 2D Gabor wavelets were found for grayscale images (Jones and Palmer 1987a). Similar 2D Gabor filters with additional red-green, yellow-blue opponents were observed for color images (Johnson et al. 2008; Shapley and Hawken 2011). Auditory receptive fields resembling gammatone filters were recorded from spiral ganglion cell axons that make up the auditory nerve (de Boer and de Jongh 1978).

## 4 Neural filters produced from natural scenes and sounds and neural efficient coding objectives

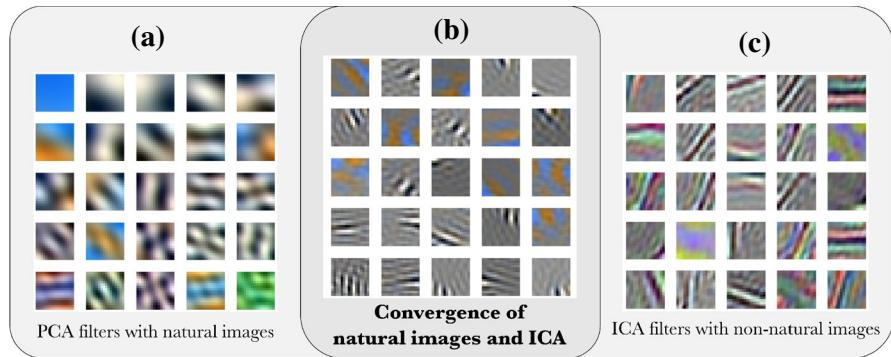
Figures 7, 8, and 9 illustrate the filters derived from applying ICA and PCA to natural and non-natural data from visual and auditory modalities. Upon visual comparison with physiological receptive fields (see Fig. 5, step 5), we observe that ICA-encoded filters qualitatively resemble experimentally measured physiological receptive fields. For natural scenes, ICA produces Gabor-like filters comparable with the neural receptive fields of V1 simple

### Grayscale Vision



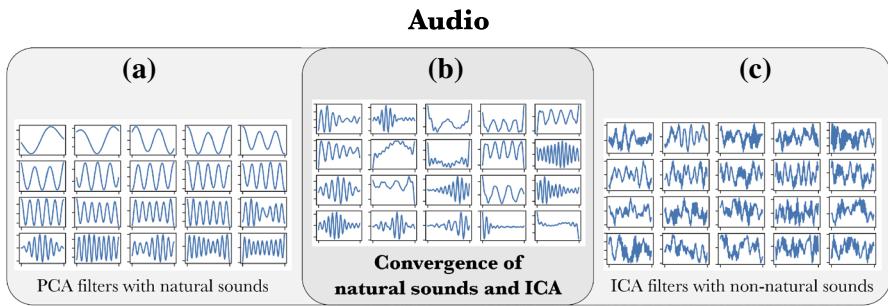
**Fig. 7** ICA- and PCA-derived visual filters for grayscale vision with natural and non-natural scenes (images). (a) PCA-encoded filters for natural grayscale images yield neurally inappropriate filters. (b) Efficient coding of natural grayscale images using ICA produces neurally appropriate filters. (c) Efficient codes of non-natural grayscale images do not produce neurally appropriate filters with ICA

### Color Vision

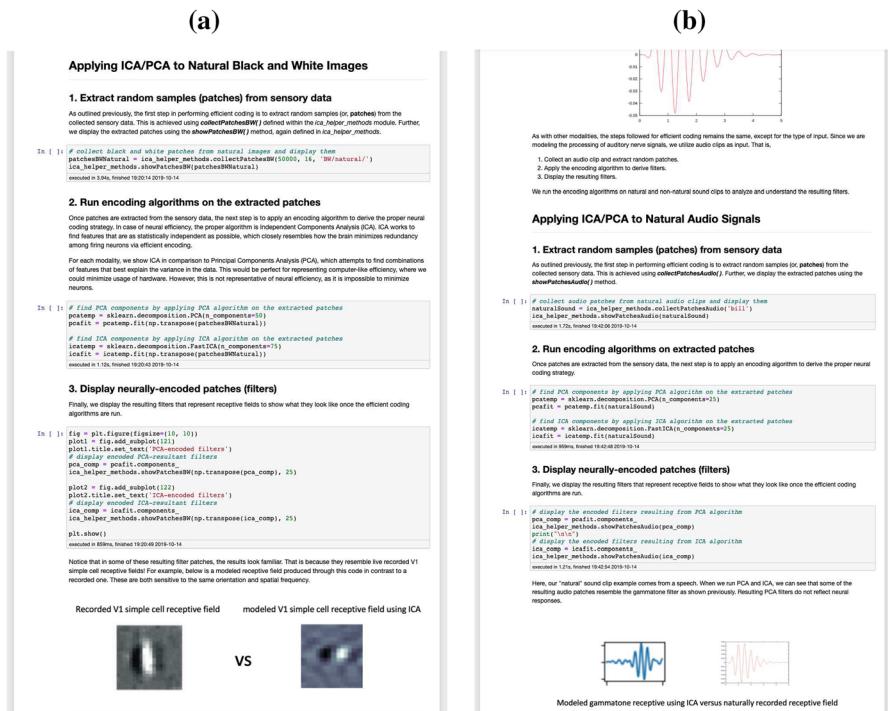


**Fig. 8** ICA- and PCA-derived visual filters for color vision with natural and non-natural scenes (images). (a) PCA-encoded filters for natural color images yield neurally inappropriate filters. (b) Efficient coding of natural color images using ICA produces neurally appropriate filters. (c) Efficient codes of non-natural color images do not produce neurally appropriate filters with ICA

cells. For natural sounds, ICA yields filters similar to gammatone filters found in the auditory system. In contrast, PCA fails to produce models analogous to the empirical filters in physiology for natural inputs. We also observe that ICA-modeled filters, when applied to non-natural inputs, do not exhibit neural-like properties, like those from natural inputs. These observations suggest that ICA is more capable of producing neural codes than PCA. Further, the same five-step coding strategy has been demonstrated for both modalities, with the only change being the inputs passed to the unsupervised learning algorithm as shown in Fig. 10.



**Fig. 9** ICA- and PCA-derived auditory filters with natural and non-natural audio signals. (a) PCA-encoded filters for natural sounds yield neurally inappropriate filters. (b) Efficient coding of natural sounds using ICA produces neurally appropriate filters. (c) Efficient codes of non-natural sounds do not produce neurally appropriate filters with ICA



**Fig. 10** The self-contained, accessible Jupyter Notebook demonstrating the efficient coding principle using the five-step computational strategy with unsupervised learning. (a) Step-by-step application of PCA and ICA to natural grayscale images (vision). (b) The same stepwise application of PCA and ICA to natural sounds (audio). Irrespective of the modality, the same five-step coding strategy is used to efficiently code the inputs, with the only change being the inputs passed to the unsupervised learning algorithm

## 5 Discussion

With a systematic demonstration of neural efficient coding for different modalities, notebook users are readily able to observe that natural scenes and sounds have sufficient statistics to create receptive fields resembling those in the early visual and auditory systems. Similarly, the concept of efficiency is necessary and must match neural coding objectives—sparse or independent coding rather than compact coding, for example. Across all modalities for natural inputs, ICA-encoded filters (illustrated in Figs. 7, 8, and 9 as the convergence of natural inputs and ICA across all modalities) closely resemble experimentally measured receptive fields from physiology (Fig. 6). On the contrary, PCA-encoded filters did not produce neural-like receptive field models.

However, this notebook stresses the need for not only proper coding objectives but also appropriate input data such as natural scenes and natural sounds. For example, ICA-encoded filters from non-natural inputs are not comparable with physiologically measured receptive fields, while those made with natural inputs are comparable. This is understandable as “natural” scenes and sounds are more closely related in statistical structure to the images and sounds that animals have evolved and adapted to over time. In terms of the effect of parameters in the code, the size of the pixel patches or the length of the audio snippets is less significant to the running time of the code than the number of ICA components selected since dimensionality reduction is performed internally (data whitening step of FastICA). The amount of data required to produce quality filters increases substantially as the number of ICA dimensions increases. This in conjunction with the running time of the code is a limiting factor for readily accessible demonstrations.

One of the primary outcomes of this work is the availability of a self-contained, accessible notebook demonstrating neural efficient coding as a form of unsupervised learning. Although there have been previous studies related to efficient coding, this work provides an integrated, easy-to-follow notebook of the tools and techniques discussed here. Despite the distinction in neuroscience and computational curricula of different modalities, our notebook brings them together in a systematic fashion. The produced notebook uses the same five-step efficient coding strategy to model the neural receptive fields, emphasizing that each modality can be modeled with only a change in inputs (Fig. 10). Additionally, this notebook serves as an educational medium illustrating the power of computational principles like efficient coding to a broader audience of neuroscientists.

Through our work, we exemplify ICA as a good representative for creating efficient, neural-like representations of sensory data. Besides computational efficiency, the neuronal plausibility of ICA from a biological standpoint is of equal importance. For natural images, ICA yields neural-like filters that exhibit the same properties as the receptive fields of V1 simple cells. However, the algorithmic implementations of ICA can vary, thereby influencing their biological plausibility. For instance, the learning rule in the infomax network is highly non-local since neurons rely on the feedback information from neurons in the output layer, resulting in a biologically implausible system (Bell and Sejnowski 1997). More biologically plausible mechanisms have been proposed which suggest ICA-like learning in the brain. Some of the earliest methods introduced a local algorithm where each neuron utilizes the connection information local to itself (Cichocki et al. 1999; Földiák 1990; Linsker 1997). Another mechanism involved a model that uses spiking neurons and intrinsic plasticity to maximize information transmission (Savin et al. 2010). A more recent improvement towards biological plausibility

has been a learning rule, called Error-Gated Hebbian Rule or EGHR, that requires only synaptic-level local information (Isomura and Toyoizumi 2016). In spite of such biologically realistic learning improvements, there is no clear consensus on the brain's encoding strategy(ies) or how these unfold over development (Avitan and Goodhill 2018).

Although ICA is highlighted as a way to efficiently encode sensory data, compact coding can also be essential for the brain. Especially when sensory data needs to be compressed, even in the early processing stage of the brain, a dimensionality reduction technique such as PCA becomes useful. For instance, PCA-like learning becomes crucial in object perception since visual inputs are extremely high-dimensional (DiCarlo et al. 2012). Another example is the cocktail party problem (Cherry 1953) where data pertaining to the speaker gets compressed into fewer dimensions due to the structural differences between the eye and the brain. However, the non-locality aspect of PCA algorithms (Oja 1989), like that of ICA, constrained the understanding of neuronal mechanisms that might be responsible for PCA-like learning. One such local learning rule called EGHR- $\beta$  has been proposed to perform PCA and ICA simultaneously using a single-layer feedforward neural network (Isomura and Toyoizumi 2018).  $\beta$  is an interpolation parameter taking on a value of either zero or one. While  $\beta=0$  enables separation of independent sources as per the ICA rule (Bell and Sejnowski 1997) regardless of the dimensionality of input and output neurons,  $\beta=1$  allows extraction of the subspace containing the principal components along the lines of PCA rule (Oja 1989). The aspect of locality is vital to the biological plausibility of a neuronal mechanism that performs neuronally plausible efficient coding (ICA) and dimensionality reduction (PCA).

Certain limitations have been identified through this work which can be addressed as part of future work. For instance, the demonstration of efficient coding using unsupervised learning to create receptive field models has been carried out using a relatively small data set of images from the internet which can bias the results. Further, our model of evaluation has been a mere visual comparison with physiologically-measured neural filters; an empirical or statistical approach to evaluate the derived filters would provide a stronger correlation with physiology. Additionally, for demonstration purposes and simplicity, we only use grayscale, color, and audio, however, video and binocular modalities will also be added in future versions of the notebook. Such additions can further emphasize the principle that the same encoding objective can model neural receptive fields with only a change in inputs.

Though the emphasis of our work has been the unsupervised aspect of learning, the ultimate role of these encodings, in both nature and computational applications, is to improve task-oriented behavior. In this regard, from an applied computational perspective, the influence of unsupervised learning during pre-training for deep learning based vision tasks sounds promising for better generalization from the training data (Erhan et al. 2010). Further, a biologically plausible implementation of ICA-like learning in a neural network has been proposed demonstrating system robustness with respect to the parameters analogous to how biological networks function (Gerhard et al. 2009). Additionally, we explored the combination of innate learning hypotheses (Albert et al. 2008) and efficient coding using ICA on images of spontaneous activity patterns (Behpour et al. 2020). ICA was found to produce filters similar to those produced for natural images which further suggests the usefulness of ICA during model training for vision tasks. Another possible direction is the use of efficiently coded ICA filters as pre-trained features in the early layers of a deep learning model. This is clear as many deep learning convolutional neural network strategies produce linear filters in the first layer of processing that moderately resemble the neural filters

described here, further complicating the potential for identifying a single computational objective for these neurons.

## 6 Conclusion

This work presents a parsimonious view of the connection between neurally appropriate efficient coding of the natural environment and developing sensory systems. By building a self-contained Jupyter Notebook, we demonstrated the efficient coding principle in a systematic way for different visual and auditory modalities. Our experiments support that independent, sparse coding objectives, such as ICA, create filters that are more similar to physiological receptive fields than compact codes, such as PCA. Thus, with a change in the inputs, the same five-step computational strategy can be used to model early sensory processing regardless of the modality (i.e., grayscale images, color images, and audio).

The Jupyter Notebook is intended for introductory computational neuroscience research and general outreach to understand the power of unsupervised learning principles, such as the efficient coding principle, to those with general neuroscience interests. This consolidated review illustrates the power of computational principles like efficient coding and can be utilized by those interested in efficient coding or neuroscience regardless of one's programming knowledge. Understanding the principle of efficient coding on early visual and auditory systems could provide insights into the more complex sensory systems such as olfaction and somatosensation. Integrating prior works of using efficient coding principle for different sensory modalities, our objective is to make this demonstration accessible to facilitate future research on multimodal integration.

The Jupyter Notebook and the documentation concerning its environment setup is publicly available at <https://www.biomed-ai.com/apps>.

**Acknowledgements** This section is not applicable to this article.

**Author contributions** Namratha Urs, Sahar Behpour, and Angie Georgaras analyzed the data. Namratha Urs and Mark V. Albert designed the study. Namratha Urs and Sahar Behpour wrote the paper. Namratha Urs prepared the manuscript. Ryan Moye and Mark V. Albert reviewed the manuscript.

**Funding** This work was supported by startup funds to author Mark V. Albert to direct the Biomedical Artificial Intelligence lab at the University of North Texas.

## Declarations

**Conflict of interest** The authors have no conflicts of interest to declare that are relevant to the content of this article. Additionally, the authors have no relevant financial or non-financial interests to disclose.

**Availability of data and material (data transparency)** All the relevant data and instructional material used in this work are available at <https://www.biomed-ai.com/apps>.

**Code availability (software application or custom code)** The source code and the Jupyter Notebook is available at <https://www.biomed-ai.com/apps>.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article

are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Albert MV, Schnabel A, Field DJ (2008) Innate visual learning through spontaneous activity patterns. *PLoS Comput Biol* 4(8):e1000137
- Avitan L, Goodhill GJ (2018) code under construction: neural coding over development. *Trends Neurosci* 41(9):599–609
- Barlow HB (1961) Possible principles underlying the transformation of sensory messages. *Sensory Commun* 1:217–234. <https://doi.org/10.7551/mitpress/9780262518420.003.0013>
- Behpour S, Urs N, and Albert MV (2020). Towards an "Innate Learning" Efficient Coding Model using Spontaneous Neural Activity [Poster presentation]. In CMD-IT/ACM Richard Tapia Celebration of Diversity in Computing, Dallas, Texas, United States.
- Bell AJ, Sejnowski TJ (1997) The "independent components" of natural scenes are edge filters. *Vision Res* 37(23):3327–3338. [https://doi.org/10.1016/S0042-6989\(97\)00121-1](https://doi.org/10.1016/S0042-6989(97)00121-1)
- Berkes P, Wiskott L (2005) Slow feature analysis yields a rich repertoire of complex cell properties. *J vis* 5(6):579–602
- Blaschke T, Berkes P, Wiskott L (2006) What is the relation between slow feature analysis and independent component analysis? *Neural Comput* 18(10):2495–2508
- Bronkhorst AW (2015) The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Atten Percept Psychophys* 77(5):1465–1487
- Cherry EC (1953) Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am* 25(5):975–979
- Cichocki A, Karhunen J, Kasprzak W, Vigário R (1999) Neural networks for blind separation with unknown number of sources. *Neurocomputing* 24(1):55–93
- Comon P (1994) Independent component analysis, A new concept? *Signal Process* 36(3):287–314. [https://doi.org/10.1016/0165-1684\(94\)90029-9](https://doi.org/10.1016/0165-1684(94)90029-9)
- de Boer E, de Jongh HR (1978) On cochlear encoding: potentialities and limitations of the reverse-correlation technique. *J Acoust Soc Am* 63(1):115–135. <https://doi.org/10.1121/1.381704>
- DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? *Neuron* 73(3):415–434
- Erhan D, Courville A, Bengio Y, and Vincent P (2010). Why Does Unsupervised Pre-training Help Deep Learning? (Y. W. Teh & M. Titterington (eds.); Vol. 9, pp. 201–208). JMLR Workshop and Conference Proceedings. <http://proceedings.mlr.press/v9/erhan10a.html>
- Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am. A, Opt Image Science* 4(12):2379–2394. <https://doi.org/10.1364/josaa.4.002379>
- Field DJ (1994) What is the goal of sensory coding? *Neural Comput* 6(4):559–601. <https://doi.org/10.1162/neco.1994.6.4.559>
- Földiák P (1990) Forming sparse representations by local anti-Hebbian learning. *Biol Cybern* 64(2):165–170
- Fong RC, Scheirer WJ, Cox DD (2018) Using human brain activity to guide machine learning. *Sci Rep* 8(1):5397. <https://doi.org/10.1038/s41598-018-23618-6>
- Gerhard F, Savin C, and Triesch J (2009). A robust biologically plausible implementation of ICA-like learning. *ESANNN*. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.227.3442&rep=rep1&type=pdf>
- Haykin S, Chen Z (2005) The cocktail party problem. *Neural Comput* 17(9):1875–1902
- Hoyer PO, and Hyvärinen A (2000). Independent component analysis applied to feature extraction from colour and stereo images. *Network*, 11(3): 191–210. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1104668/>
- Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160:106–154. <https://doi.org/10.1113/jphysiol.1962.sp006837>
- Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195(1):215–243. <https://doi.org/10.1113/jphysiol.1968.sp008455>
- Hyvärinen A (1999) Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Netw*/publ IEEE Neural Netw Council 10(3):626–634

- Hyvärinen A, Hoyer PO (2001) Topographic independent component analysis as a model of V1 organization and receptive fields. *Neurocomputing* 38–40:1307–1315
- Hyvärinen A, Hoyer PO, Inki M (2001) Topographic independent component analysis. *Neural Comput* 13(7):1527–1558
- Isomura T, Toyoizumi T (2016) A local learning rule for independent component analysis. *Sci Rep* 6:28073
- Isomura T, Toyoizumi T (2018) Error-gated hebbian rule: a local learning rule for principal and independent component analysis. *Sci Rep* 8(1):1835
- Johnson EN, Hawken MJ, Shapley R (2008) The orientation selectivity of color-responsive neurons in macaque V1. *J Neurosci: off J Soc Neurosci* 28(32):8096–8106. <https://doi.org/10.1523/JNEURONSCI.1404-08.2008>
- Jones JP, Palmer LA (1987a) An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol* 58(6):1233–1258. <https://doi.org/10.1152/jn.1987.58.6.1233>
- Jones JP, Palmer LA (1987b) The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J Neurophysiol* 58(6):1187–1211. <https://doi.org/10.1152/jn.1987.58.6.1187>
- Jutten C, Herault J (1991) Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Process* 24(1):1–10. [https://doi.org/10.1016/0165-1684\(91\)90079-X](https://doi.org/10.1016/0165-1684(91)90079-X)
- Lennie P (2003) The cost of cortical computation. *Curr Biol: CB* 13(6):493–497. [https://doi.org/10.1016/s0960-9822\(03\)00135-0](https://doi.org/10.1016/s0960-9822(03)00135-0)
- Lewicki MS (2002) Efficient coding of natural sounds. *Nat Neurosci* 5(4):356–363. <https://doi.org/10.1038/nn831>
- Linsker R (1997) A local learning rule that enables information maximization for arbitrary input distributions. *Neural Comput* 9(8):1661–1665
- Oja E (1989) Neural networks, principal components, and subspaces. *Int J Neural Syst* 01(01):61–68
- Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381(6583):607–609. <https://doi.org/10.1038/381607a0>
- Olshausen BA, and Field DJ (2000). Vision and the coding of natural images: the human brain may hold the secrets to the best image-compression algorithms. *American Scientist* 88(3): 238–245. <http://www.jstor.org/stable/27858027>
- Savin C, Joshi P, Triesch J (2010) Independent component analysis in spiking neurons. *PLoS Comput Biol* 6(4):e1000757
- Shapley R, Hawken MJ (2011) Color in the cortex: single- and double-opponent cells. *Vision Res* 51(7):701–717. <https://doi.org/10.1016/j.visres.2011.02.012>
- Sherrington CS (1906) Observations on the scratch-reflex in the spinal dog. *J Physiol* 34(1–2):1–50. <https://doi.org/10.1113/jphysiol.1906.sp001139>
- van Hateren JH, Ruderman DL (1998) Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proceed Biol Sci/r Soc* 265(1412):2315–2320. <https://doi.org/10.1098/rspb.1998.0577>
- Wiskott L, Sejnowski TJ (2002) Slow feature analysis: unsupervised learning of invariances. *Neural Comput* 14(4):715–770. <https://doi.org/10.1162/089976602317318938>

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