Efficient Coding of Natural Sounds
Modélisation en Neuroscience et ailleurs
Master MVA (2020 - 2021)

Dorian Desblancs
École Normale Supérieure Paris-Saclay
4 Avenue des Sciences, 91190 Gif-sur-Yvette, France
dorian.desblancs@mail.mcgill.ca

Abstract

In this project, we explore the field of efficient coding. We focus on the efficient coding of sounds using independent component analysis (ICA). The methodology used was heavily inspired by Michael S. Lewicki's seminal paper *Efficient Coding of Natural Sounds* [16]. We used four sets of sounds, ordered from least to most periodic: environmental, speech, bird, and synthesizer. Our results demonstrate that the more periodic or harmonic a set of sounds is, the more sinusoidal its sources are. We notice that more random, environmental noises are split into components resembling wavelets. On the contrary, harmonic bird sounds are split into components resembling sinusoids. This work ties into the conclusions drawn by Lewicki. The center frequencies and bandwidths obtained by the set of derived auditory filters are similar to those found in the cochlear nervous system. This further reinforces the idea that information theory is linked to the brain, most notably in the realm of sparse coding.

1 Introduction

Much is known about how we learn. Since the early days of neuroscience, and most notably Donald Hebb's seminal work *The organization of behavior: A neuropsychological theory* [10], computational models describing how mammals acquire information from their surrounding environments have flourished. In the realm of sensorial stimuli, the research is extensive. We know that cells are tuned to environmental features that surround us. In the field of vision, simple cells are tuned to orientations, and spike when they are activated. In the field of audition, cochlear nerve fibers are sharply tuned to particular frequencies. They can be seen as performing a short-term spectral analysis of acoustic signals [23]. However, the reasons behind these tuning remained quite mysterious until the late 1990s.

In the early 90s, the field of computational neuroscience was changed dramatically by the old idea of efficient coding. Efficient coding, presented by Barlow in 1961 [2], hypothesises that the sensory processing processing in the brain should be adapted to natural stimuli. Thirty years later, researchers derived efficient codes for visual images using the novel technique of independent component analysis [9] [4] [21] [17]. They discovered that the statistically independent components present in images were in fact edge filters, akin to those found in the primary visual cortex. This groundbreaking idea was then used for audition in 2002, in the paper *Efficient Coding of Natural Sounds* by Michael S. Lewicki.

In this project, we derived efficient codes for four sets of sounds. We inspired our methodology by that found in the work of Lewicki [16]. This report presents our findings. We will first explore some background information needed for the comprehension of this work. We will then present our methodology and our results. Finally, we will discuss how our work compares to that of Lewicki's, and the next steps we believe that we should take. A link to our code and in-depth results is available at the end of this report.

2 Background Information

2.1 Theory behind Efficient Codes

2.1.1 Barlow's Efficient Coding Hypothesis

The theory behind efficient coding is quite old. The original hypothesis was introduced by Barlow in 1961 [2]. In his work, he treated the mapping between sensory stimuli and neuronal spiking as a communication channel. He postulated that the goal of the nervous system was to maximize the information it was capable of gaining from the environment. In order to do so, the mapping between environmental stimulus and neural coding had to maximize its mutual information (infomax). On the other hand, the capacity of each channel had to be minimized. Why? So as to gain as much information about the environment using as little spikes as possible, hence the idea of efficient codes.

Mathematically, this can be described as follows. We define the entropy of a random variable X with outcomes $x_1, ..., x_n$ as:

$$H(X) = -\sum_{i=1}^{n} p(x_i) logp(x_i) \tag{1}$$

The conditional entropy of X given Y is then defined as:

$$H(X|Y) = -\sum_{y} p(y) \sum_{x} p(x|y) logp(x|y) \tag{2}$$

Both are used to define the mutual information:

$$I(X, Y) = H(Y) - H(Y|X) \tag{3}$$

Note that the entropy describes a random variable's average amount of "uncertainty", or "surprise". Conditional entropy quantifies the amount of information needed to describe the outcome of Y given that the value of X is known. Finally, the mutual information between X and Y describes the amount of information X conveys about Y (or vice versa).

Barlow proposed that the nervous system was to minimize the following quantity, a slight alteration of redundancy:

$$R = 1 - \frac{H(Y)}{C} \tag{4}$$

where C defines the channel capacity. He proposed $H(Y)$ define the nervous system's response entropy, while $H(Y|X)$ define its noise entropy. As such, by minimizing its redundancy, the nervous system would be able to efficiently code the de-noised sensorial stimuli it was exposed to. This idea was further explored by Atick and Redlich in 1990 [1]. It also led to a large number of experimental studies validating his theory, the first one being Laughlin's 1981 study of fruit flies [14]. The mathematical foundations of Barlow's work rely on Claude Shannon's landmark work [24] in information theory.

2.1.2 Independent Component Analysis (ICA)

Independent component analysis (ICA) is a method used for splitting a signal into its multivariate sub components. This technique is especially popular in the realms of imaging and sound, in which the algorithm is used derive the independent components that constitute these. Why independent? Because ICA assumes that the sub-components are non-Gaussian. As such, they are statistically independent and can form sparse codes for a signal using a natural basis.

In our case, we assume a linear noiseless ICA. As such, the model can be written as:

$$x = \sum_{k=1}^{n} a_k s_k \tag{5}$$

where $x = (x_1, ..., x_m)^T$ is the observed mixture of signals. It is represented by the basis vectors $s_k = (s_1, ..., a_k)^T$, which define the signal sub-components, and the coefficients a_i, which define the mixing matrix coefficients. ICA is hugely popular in the domain of blind source separation. For sound, this entails splitting a sound into the sources that compose it. The most famous of these problems is referred to as the cocktail party problem [12] [3].

In the realm of neuroscience, ICA is a popular method for deriving efficient codes. The components it finds are statistically independent, and minimize redundancy. As such, they can be used to study the characteristics of naturally occurring phenomenon. In imaging, the extracted sources resemble localized edge detectors [17].These were found to greatly resemble the neuronal tuning found in the visual system cells (Hubel and Wisel [11]). As such, the sources derived by ICA are believed to mimic the nervous system cell's tunings. The sources extracted for sound are to be presented later in this report.

2.2 Time-Frequency Signal Representations

The Fourier and Wavelet transforms are arguably the two most important signal transformations to date. The Fourier transform extracts the frequency information present across an entire signal. It contains this information in terms of the magnitude and phase angle of any frequency component. It however does not contain any information regarding the temporal characteristics of a signal (such as when a certain frequency appeared). As such, signals are thought of in terms of sinusoids, or smooth periodic oscillations. The Wavelet transform was introduced to solve this problem. It trades off some of the frequency domain resolution present in the Fourier transform for a greater temporal resolution. As such, the signals are thought of as wavelets. The mathematical details of both transforms and their applications to sound and imaging can be found in [18] [27].

3 Methodology

3.1 Signal Decomposition Methods

Based upon the theory behind efficient coding, the auditory coding model assumed is [16]:

$$a_i(t) = \sum_{k=0}^{N-1} x(t-k)h_i(k) \tag{6}$$

The signal over a time window of N samples is encoded in the responses $a_1(t), ..., a_M(t)$. The goal is to derive the set of filters $h_1(t), ..., h_M(t)$, which correspond to the signal decomposition elements found by independent component analysis. In our case, we adapted N such that it corresponds to approximately 0.05 seconds of our test signals (next section). We used a steady value of $M = 128$ sources. Finally for each test data set, 256 signals were used to find the sources. This was done in order to extract sources that were representative of a particular set of sounds.

We implemented the FastICA algorithm proposed by Oja et al. [13] for our independent component analysis. We used an approximation of negentropy as measure of non-Gaussianity. This approximation was defined as:

$$J(y) = [E(G(y)) - E(G(V))]^2 \tag{7}$$

We chose the following function for G:

$$G(y) = ye^{-\frac{y^2}{2}} \tag{8}$$

whose derivative is:

$$G'(y) = (1-y^2)e^{-\frac{y^2}{2}} \tag{9}$$

We tested multiple functions for G, but found the one defined above to work best experimentally. The Fast ICA algorithm is then defined by the following steps:

1. Choose a random weight vector w

2. Let $w^+ = E[xg(w^T x)] - E[g'(w^T x)]w$
3. Let $w = \frac{w^+}{||w^+||}$
4. If not converged, go back to 2.

Note that we pre-processed the data by performing the standard centering and whitening steps. The algorithm was run for 20000 iterations, which we found to be ample time for convergence. For more information, we encourage the reader to read about the algorithm in Oja and Hyvarinen's classic book, *Independent Component Analysis: Algorithms and Applications* [13].

Finally, notice that Lewicki's [16] methodology for independent component analysis is quite different from ours. We took the liberty to propose our own approach. The differences in methodology may however be the reason behind some of our differing results (see Discussion).

3.2 Test Data

We extracted sources from four data sets. The first one was the PolandNFC data set [5]. This data set regroups 4000 recordings of nocturnal birds during migration. The audio snippets are centered around the bird sounds. There is however lots of noise in these recordings, due to the weather and surrounding environment. We tried to use this data set as a testing ground for harmonic, animal sounds.

The second data set is the TIMIT [28] data set. This data set regroups 2343 audio files that contain speech recordings. These recordings are a mix of male and female voices. Note that this data set was also used in Lewicki's [16] original paper.

The third data set used was the TUT Acoustic Scenes [19] data set. This data set, popular in the field of automatic acoustic scene recognition, contains a diverse array of recordings that are from restaurants, beaches, nature... It was used to extract the sources behind environmental sounds.

Finally, we also tested our methodology on a set of synthesizer sounds from the Website MusicRadar (`https://www.musicradar.com/`). These were used to test our methodology on less noisy, highly periodic signals.

For each data set, we extracted approximately 0.05 seconds worth of samples. These were randomly selected from a randomly selected subset of 256 signals. We also used a high-pass filter to remove frequencies whose period was longer than the number of samples extracted per signal.

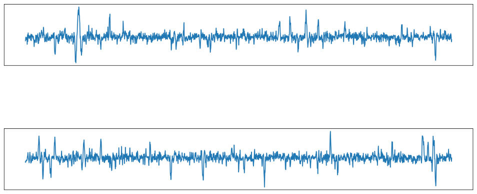
4 Results

The following page displays some of our results. Note that the amplitude and time-frequency (spectrogram) representations of every single auditory source obtained from efficient coding can be found in the repository listed at the end of the report.

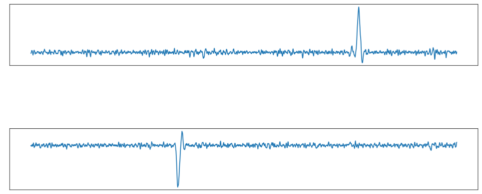
We limited this report to displaying two auditory components that were representative of the 128 extracted using our ICA implementation. Since the sounds used for each data set are chosen at random, there may be a little variability regarding the extracted sources (when running the code). This variability was found to be minor, however. We display both the amplitude spectrum and spectrograms of each sample audio code. We find that the spectrogram representation allows us to further visualize the center frequency, frequency bandwidth, and periodicity of each derived auditory filter.

Now, when looking at our results, one can immediately notice that the components obtained from our environmental sounds are very different from the rest. They are extremely localized in time and frequency and aperiodic. These results are similar to those obtained by Lewicki [16]. In his paper, he describes filters that resemble a wavelet. We obtain a similar result. This is due to the fact that environmental sounds are incredibly variable in their characteristics, and occur very randomly. As such, sinusoid-shaped components are rarely, if ever, extracted. On the contrary, the components tend to resemble spikes in the amplitude spectrum.

We do however observe that as our signals become more harmonic, so do the extracted components. Theoretically, in our data sets, the second least periodic type of signal is speech. Speech can be both periodic and aperiodic. This is due to the fact that some sounds are voiced, and therefore periodic (this is notably the case of vowels in English), while other are unvoiced, and aperiodic (consonants in English). As such, our components resemble a mix of sinusoids and wavelets. Both the spectrogram and amplitude representations support this claim.

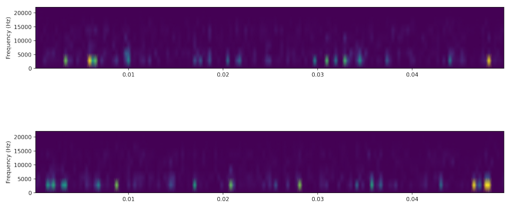


(a) Bird Sounds

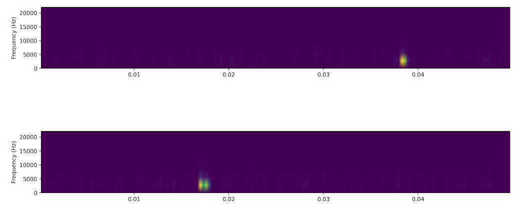


(b) Environmental Sounds

Figure 1: Sample Audio Components derived from Efficient Coding

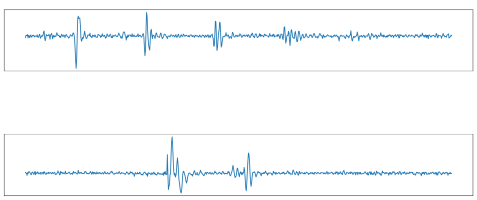


(a) Bird Sounds

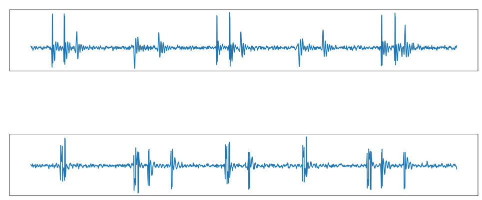


(b) Environmental Sounds

Figure 2: Spectrogram of Sample Audio Components

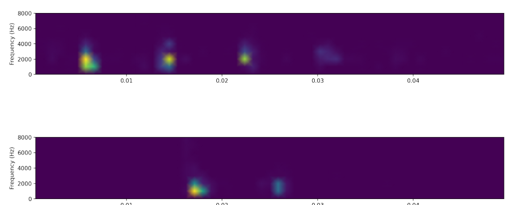


(a) Speech Sounds

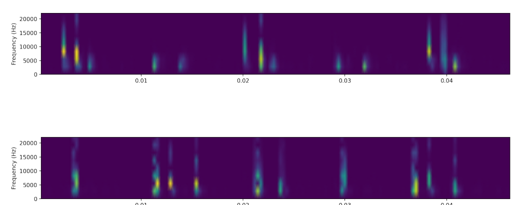


(b) Synthesizer Sounds

Figure 3: Sample Audio Components derived from Efficient Coding



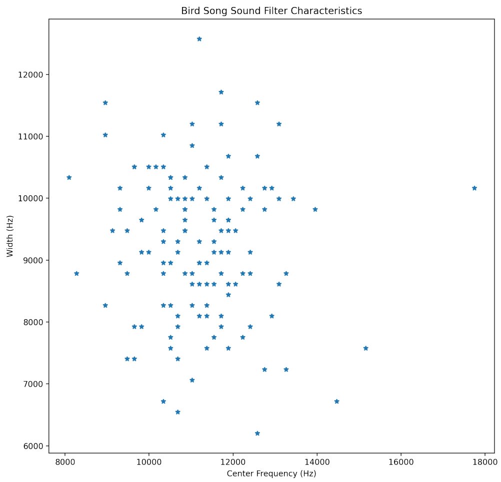
(a) Speech Sounds



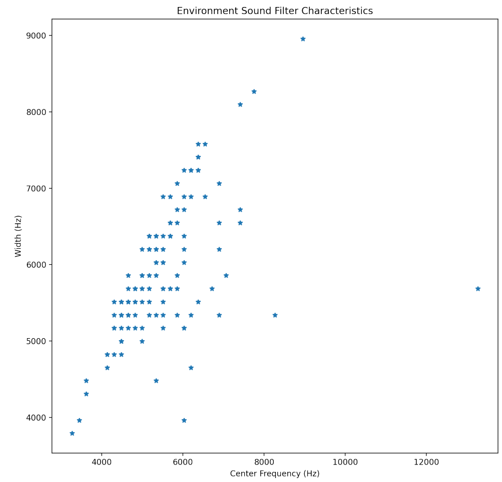
(b) Synthesizer Sounds

Figure 4: Spectrogram of Sample Audio Components

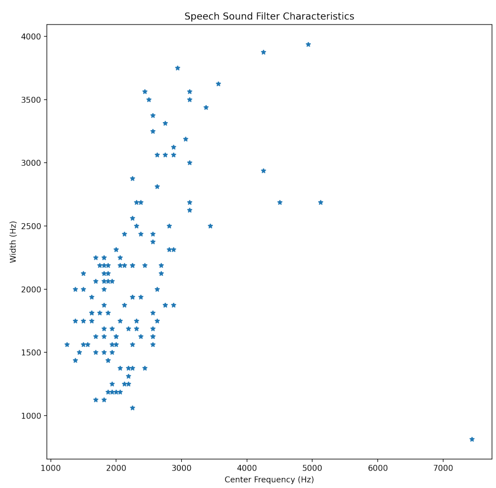
Finally, the bird sounds and synthesizer sounds appear to be drastically more periodic. For the bird sounds, do remember that our audio was not pre-processed to isolate the bird cries. It also contains a lot more noise, due to weather and environmental sounds. This is a notable difference from the original Lewicki [16] paper. That is also why our components appear to be both a mix of periodic sounds, that correspond to harmonic bird sounds, and aperiodic, environmental sounds. The results obtained from our synthesizer data outlines the idea of periodic components even better. The components extracted correspond to periodic frequency modulations that appear in our data.



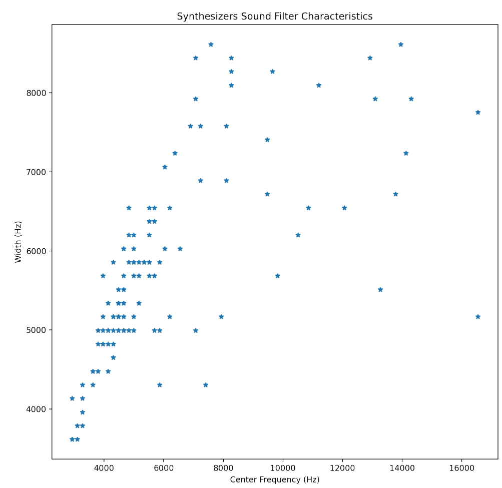
(a) Bird Sounds



(b) Environmental Sounds



(c) Speech Sounds



(d) Synthesizer Sounds

Figure 5: Auditory Filter Center Frequencies and Widths

Figure 5 displays the center frequencies of each filter plotted against their width. Note that we define width here as the difference between the upper and lower frequencies that concentrate the middle 50% of the component's power. We can observe that for aperiodic signals, the width is increased as the center frequency increases. This is most notable in the environmental and speech sounds. The average width is also quite low. These results further show that the components obtained from aperiodic sounds are akin to wavelets. For the bird song signals, on the other hand, we do not observe any correlations between center frequency and width, suggesting components closer to sinusoids. Finally, our synthesizer results are interesting because we observe a little bit of both ideas. This is probably due to the fact that our components capture frequency-modulated periods. These can vary greatly in frequency, and can also be quite localized in time, hence the structure of our plot.

5 Discussion

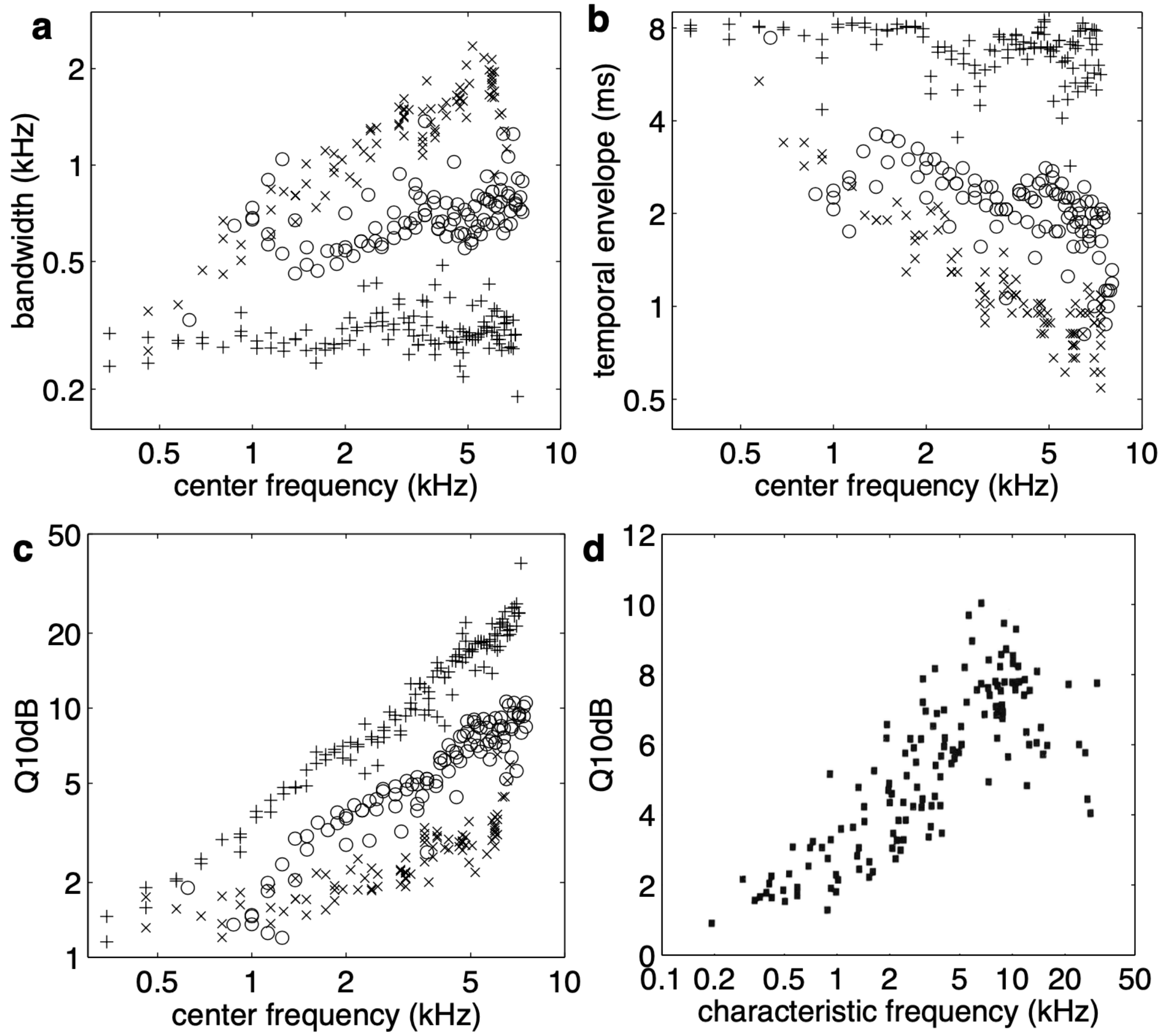


Figure 6: Filter characteristics as a function of center frequency for environmental sounds (×), speech (○), and vocalizations (+). Filter bandwidth (a). Filter temporal envelope width (b). Filter sharpness or center frequency divided by bandwidth (Q10dB) (c). Q10dB measured from cat auditory nerve fibers (d). From [8].

When comparing our results to *Efficient Coding of Natural Sounds* [16], one can notice similarities. This is notably the case in the sources obtained by independent component analysis. Although our results are obtained on longer and noisier sections of audio, the conclusions drawn by the paper are noticeable in our results. These conclusions are that the auditory filters obtained for environmental, non-harmonic sounds resemble wavelets. On the other hand, those obtained from harmonic or periodic sounds resemble sinusoids. Our results regarding center and bandwidth frequencies are also quite similar. Please consult the appendix to see the auditory filters obtained by Lewicki. Figure 6, parts a) and b), outline the temporal envelopes, center frequencies, and bandwidths observed during Lewicki's experiments. Another reason behind our differing results could be the function G used to measure negentropy ($G(a) = -exp(-|a|^q)$ for Lewicki), and the differences between the ICA algorithms used.

Lewicki's paper [16], however, really shines when comparing the derived filter's center frequency divided by bandwidth (Q_{10dB}) to the Q_{10dB} measured from cat auditory nerve fibers (Figure 7, parts c) and d)). We notice a similar pattern, suggesting that the derived auditory filters are similar to those found in the nervous system.

6 Future Work

Although we believe our results are quite strong, some elements need to be improved. For one, the data sets need to be cleaned. Estimating efficient codes is very hard on noisy data, and ours was clearly much noisier than the data used by Lewicki in 2002. This is especially important for the animal noises. We believe that the sound decompositions obtained using ICA would have been much more similar to regular sinusoids if we had taken the time to isolate every bird sound from its surrounding noise. More importantly, since the sample sections for each sound are chosen randomly in our data sets, a number of our data points used for ICA were probably just noise, leading to inaccurate filters for harmonic sounds. We also believe a more thorough study of the center and bandwidth frequencies of each one of our sources is needed, so as to link our results to the nervous system and its responses to natural sounds.

After the publication of *Efficient Coding of Natural Sounds*, Lewicki published at least two works exploring the link between the auditory filters derived by ICA and their links to those found in the cochlear system (especially in terms of time-frequency) [26] [25]. This work further reinforces the links between information theory and the brain.

On a separate note, perhaps independent component analysis could be used more often in the realm of machine learning. ICA is already quite popular in the area of brain imaging [15]. Perhaps using the auditory sources derived by the algorithm could serve as a way to augment one's data for problem such as audio scene recognition.

7 Conclusion

In this project, we explored efficient coding through the lense of audition. We derived the sources associated with four sets of sounds with different properties using independent component analysis. The sources obtained allowed us to further demonstrate that the more a sound is periodic or harmonic, the more its filters are too. The methodology used in this paper was heavily inspired by *Efficient Coding of Natural Sounds* by Michael S. Lewicki [16]. We did however propose a few alterations, such as the use of the Fast ICA algorithm and a different set of testing sounds. This project allowed us to learn a lot about the vast field of efficient coding. We hope to expand upon the knowledge gained in the future, and thank Professor Jean-Pierre Nadal for the opportunity to explore this field.

Acknowledgements

All the code and results used for this report can be found on the following Github:

- username: **d-dawg78**
- repository: **MVA_Neuro**

The slides from [22] and [20] were also used extensively when writing the background information section.

References

[1] Joseph J Atick and A Norman Redlich. "Towards a theory of early visual processing". In: *Neural computation* 2.3 (1990), pp. 308–320.

[2] Horace B Barlow et al. "Possible principles underlying the transformation of sensory messages". In: *Sensory communication* 1.01 (1961).

[3] Anthony J Bell and Terrence J Sejnowski. "An information-maximization approach to blind separation and blind deconvolution". In: *Neural computation* 7.6 (1995), pp. 1129–1159.

[4] Anthony J Bell and Terrence J Sejnowski. "The "independent components" of natural scenes are edge filters". In: *Vision research* 37.23 (1997), pp. 3327–3338.

[5] Franz Berger et al. "Bird audio detection-dcase 2018". In: *DCASE2018 Workshop, Surrey (UK)*. 2018.

[6] Laurel H Carney and TC Yin. "Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model". In: *Journal of neurophysiology* 60.5 (1988), pp. 1653–1677.

[7] E De Boer and HR De Jongh. "On cochlear encoding: Potentialities and limitations of the reverse-correlation technique". In: *The Journal of the Acoustical Society of America* 63.1 (1978), pp. 115–135.

[8] EF Evans. "Cochlear nerve and cochlear nucleus". In: *Auditory system*. Springer, 1975, pp. 1–108.

[9] J Hans van Hateren and Dan L Ruderman. "Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex". In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265.1412 (1998), pp. 2315–2320.

[10] Donald Olding Hebb. *The organization of behavior: A neuropsychological theory*. Psychology Press, 2005.

[11] David H Hubel and Torsten N Wiesel. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". In: *The Journal of physiology* 160.1 (1962), pp. 106–154.

[12] Aapo Hyvärinen. "Survey on independent component analysis". In: (1999).

[13] Aapo Hyvärinen and Erkki Oja. "Independent component analysis: algorithms and applications". In: *Neural networks* 13.4-5 (2000), pp. 411–430.

[14] Simon Laughlin. "A simple coding procedure enhances a neuron's information capacity". In: *Zeitschrift für Naturforschung c* 36.9-10 (1981), pp. 910–912.

[15] Steven Lemm et al. "Introduction to machine learning for brain imaging". In: *Neuroimage* 56.2 (2011), pp. 387–399.

[16] Michael S Lewicki. "Efficient coding of natural sounds". In: *Nature neuroscience* 5.4 (2002), pp. 356–363.

[17] Michael S Lewicki and Bruno A Olshausen. "Probabilistic framework for the adaptation and comparison of image codes". In: *JOSA A* 16.7 (1999), pp. 1587–1601.

[18] Stéphane Mallat. *A wavelet tour of signal processing*. Elsevier, 1999.

[19] Annamaria Mesaros, Toni Heittola, and Tuomas Virtanen. "TUT database for acoustic scene classification and sound event detection". In: *2016 24th European Signal Processing Conference (EUSIPCO)*. IEEE. 2016, pp. 1128–1132.

[20] Jean-Pierre Nadal. *Lecture notes on Modelling in Neuroscience*. Spring 2021.

[21] Bruno A Olshausen and David J Field. "Emergence of simple-cell receptive field properties by learning a sparse code for natural images". In: *Nature* 381.6583 (1996), pp. 607–609.

[22] Jonathan Pillow. *Lecture notes on Information Theory and the Efficient Coding Hypothesis (NEU 314)*. Spring 2016.

[23] William S Rhode and Philip H Smith. "Characteristics of tone-pip response patterns in relationship to spontaneous rate in cat auditory nerve fibers". In: *Hearing research* 18.2 (1985), pp. 159–168.

[24] Claude E Shannon. "A mathematical theory of communication". In: *The Bell system technical journal* 27.3 (1948), pp. 379–423.

[25] Evan Smith and Michael S Lewicki. “Efficient coding of time-relative structure using spikes”. In: *Neural computation* 17.1 (2005), pp. 19–45.

[26] Evan C Smith and Michael S Lewicki. “Efficient auditory coding”. In: *Nature* 439.7079 (2006), pp. 978–982.

[27] Gilbert Strang. “Wavelet transforms versus Fourier transforms”. In: *Bulletin of the American Mathematical Society* 28.2 (1993), pp. 288–305.

[28] Victor Zue, Stephanie Seneff, and James Glass. “Speech database development at MIT: TIMIT and beyond”. In: *Speech communication* 9.4 (1990), pp. 351–356.

Appendix

Efficient Coding of Natural Sounds [16]

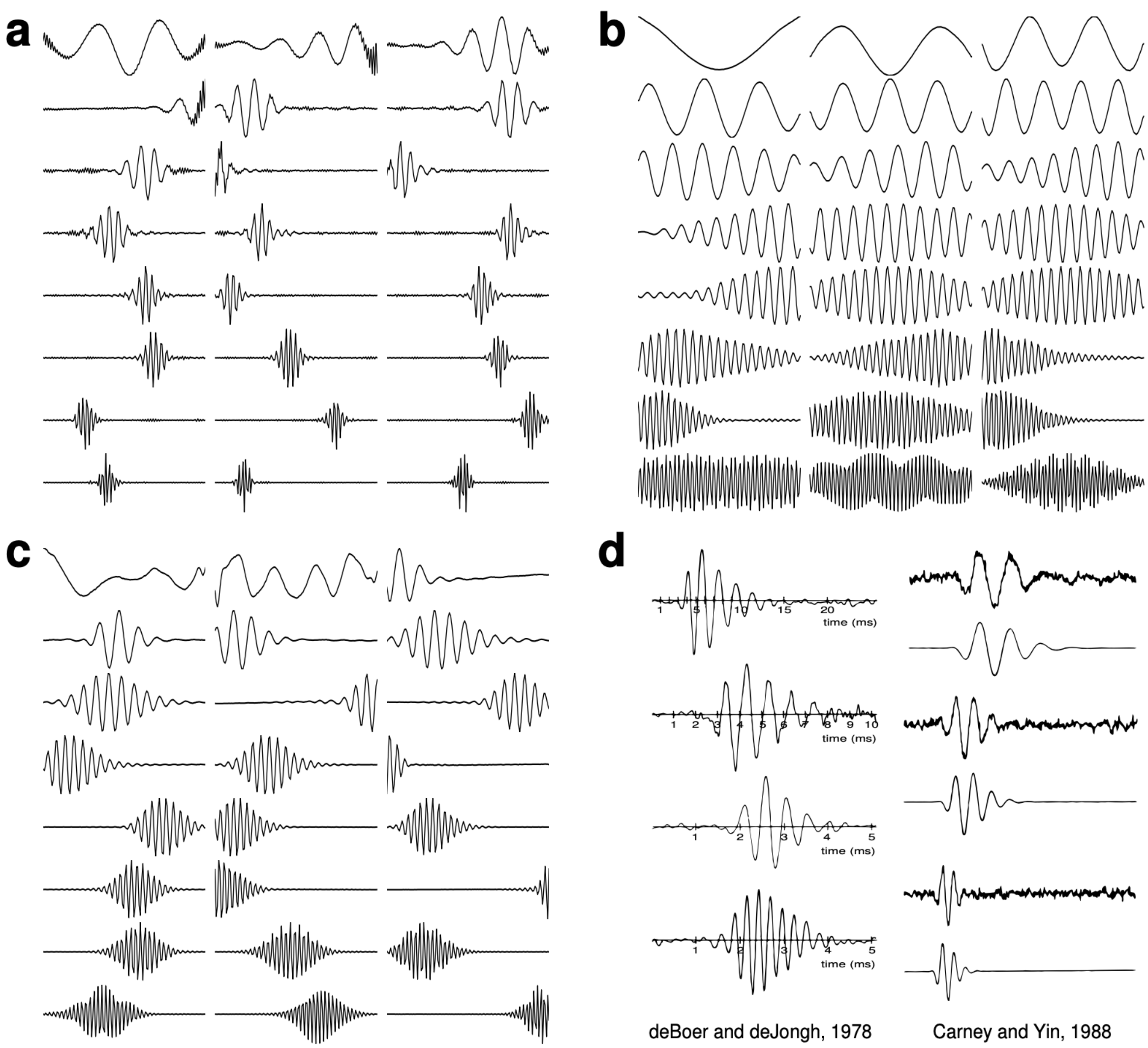


Figure 7: Auditory Filters Obtained from (a) Environmental Sounds. (b) Animal Vocalizations. (c) Speech. (d) Cochlear filter shapes measured experimentally at the auditory nerve. From [7] and [6].