A bio-inspired geometric model for sound reconstruction

Master's Thesis

Rand Asswad

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ROUEN NORMANDIE

Department of Applied Mathematics
Natalie FORTIER
Cecilia ZANNI-MERK

L2S Laboratoire Signaux & Systèmes

Dario PRANDI
Ugo BOSCAIN

23 September 2021

Outline

1. Introduction
2. Image reconstruction model
   - Neuro-geometric model of V1
   - Image reconstruction model
3. Sound reconstruction model
   - From V1 to A1
   - Time-Frequency representation
   - The lift to the augmented space
   - Cortical activations in A1
4. Implementation
   - The `WCA1.jl` package
   - Published results
5. Conclusion
   - Reviewing the model
   - Acquired knowledge
   - Future project
   - References

Section 1

Introduction

Laboratory of Signals and Systems (L2S)

- Created in 1974
- Affiliations:
  - CNRS (Centre National de la Recherche Scientifique)
  - CentraleSupélec
  - University of Paris-Saclay
- Research fields:
  - Systems and control
  - Signal processing and statistics
  - Networks and telecommunication

Supervision

- Dario Prandi
  - Affiliations: L2S, CNRS, CentraleSupélec, Université Paris-Saclay
  - Specialties:
    - Geometric control theory
    - Biomimetic image processing
    - Diffusions on singular manifolds
- Ugo Boscain
  - Affiliations: Laboratoire Jacques-Louis Lions, CNRS, Inria, Sorbonne Université
  - Specialties:
    - Sub-riemannian geometry
    - Control of quantum mechanical systems
    - (also) optimal control and switched systems

Internship mission

Work on the proposed neuro-geometric sound reconstruction model.

Subtasks:

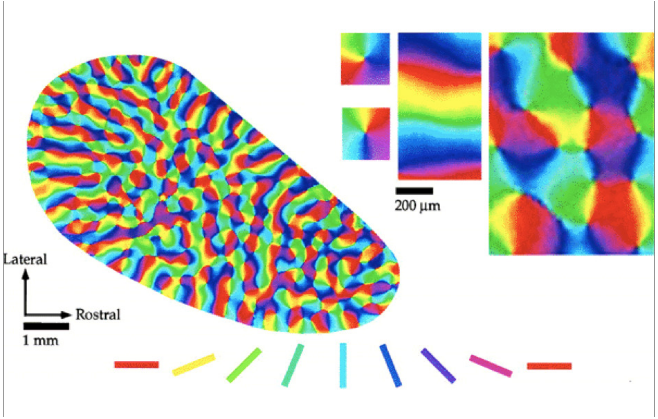
- Study the existing model
- Test on real speech signals
- Publish results
- Reimplement `WCA1.jl` package
- Rethink model & study litterature

Section 2

Image reconstruction model

Basis of the V1 model - starting point

1. Hubel and Weisel (1959) [13] observed that there are groups of neurons sensitive to positions and directions

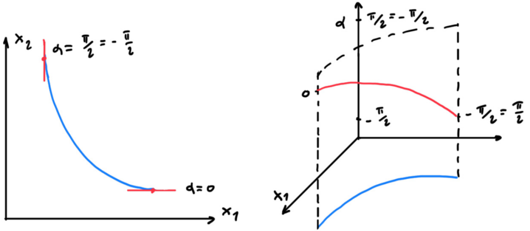


Basis of the V1 model - 3D representation

2. Which inspired Hoffman (1989) [12] to model V1 as a contact space (a 3D manifold endowed with a smooth map)
3. The Citti-Petitot-Sarti (CPS) model (2006) [7,16] extended the model to sub-Riemannian structures

The CPS model:

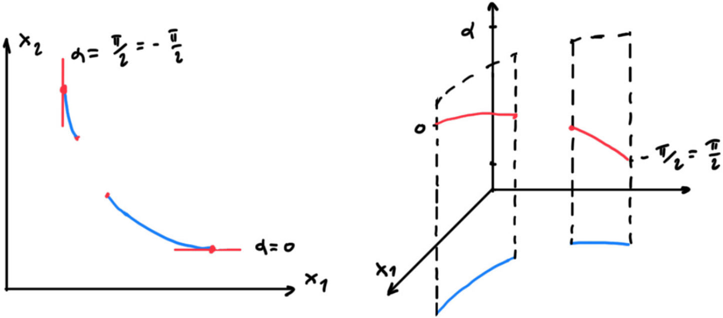
- An image can be seen as a function $f : \mathbb{R}^2 \to \mathbb{R}_+$ representing the grey level at given coordinates
- The primary visual cortex (V1) adds the non-directed angle $\theta \in P^1 = \mathbb{R}/\pi\mathbb{Z}$ of the tangent line to the curve.
  The visual cortex lifts a curve into $\mathbb{R}^2 \times P^1$.



Basis of the V1 model - image reconstruction

4. Ugo Boscain, Dario Prandi, Jean-Paul Gauthier, and their colleagues proposed (in 2017) [2,3] an image reconstruction model based on the CPS model.

If a curve is interrupted in an interval, then the visual cortex tries to reconstruct it by taking the shortest curve in the lifted space.



Wilson-Cowan model [19]

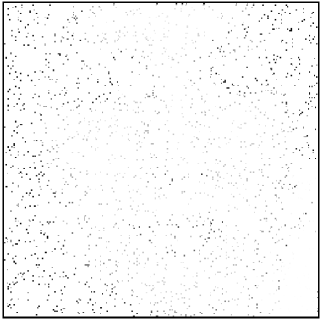
- The Wilson-Cowan (WC) model describes the evolution of neural activations
- WC describes the evolution of excitatory and inhibitory activity in a synaptically coupled neuronal network
- The interaction between the hypercolumns in V1 can be described through the WC equation [5]

Let $a(x, \theta, t)$ be the state of a population of neurons with coordinates $x \in \mathbb{R}^2$ and orientation $\theta \in P^1$ at time $t > 0$, the WC integro-differential equation is given by [2]

$$\frac{\partial}{\partial t} a(x, \theta, t) = -\alpha a(x, \theta, t) + \nu \int_{\mathbb{R}^2 \times P^1} \omega(x, \theta \| x', \theta') \sigma(a(x', \theta', t)) \mathrm{d}x' \mathrm{d}\theta' + h(x, \theta, t)$$

Reconstruction of a 97% corrupted image

original



corrupted

reconstructed

Which begs the question

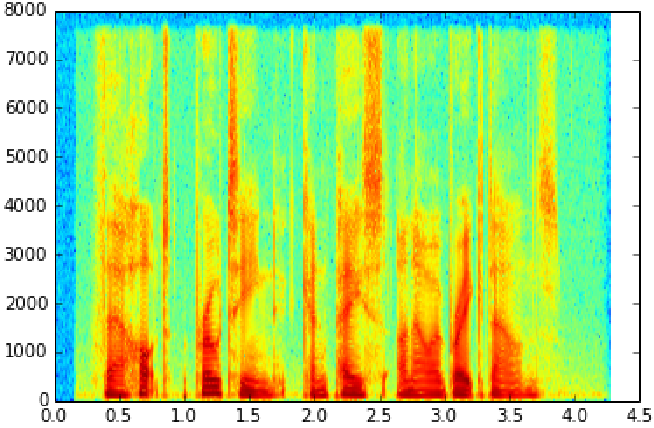
Can we apply these ideas to the problem of sound reconstruction?

Section 3

Sound reconstruction model

Motivation

A sound signal $s(t)$ can be seen as an image in the time-frequency domain $|S|(\tau, \omega)$



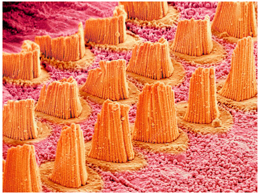
Taking into account:

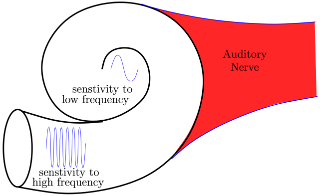
1. In image reconstruction the whole image is evolved simultaneously. However, the sound image (spectrogram) does not reach the auditory cortex simultaneously but *sequentially*. Hence, the reconstruction can be performed only in a sliding window.
2. A rotated sound image corresponds to a completely different input sound, therefore the invariance by rototranslation is lost.

Sound signal processing in the cochlea

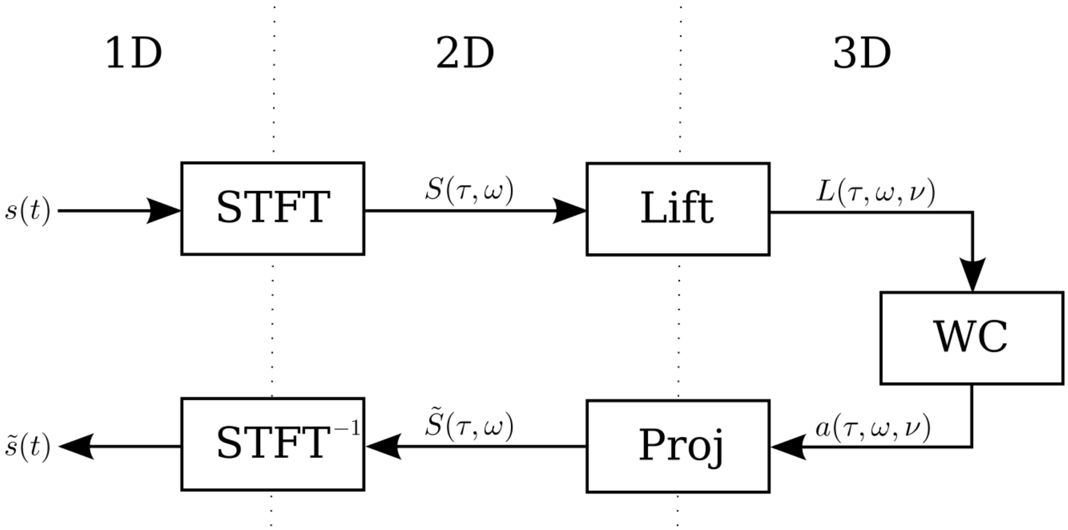
The primary auditory cortex (A1) receives the sensory input directly from the cochlea [8], which is a spiral-shaped fluid-filled cavity that composes the inner ear.

- The mechanical vibrations along the basilar membrane are transduced into electrical activity along a dense, topographically ordered, array of auditory-nerve fibers (hair cells) which convey these electrical potentials to the central auditory system.
- Since the inner hair cells are topographically ordered along the cochlea spiral, different regions of the cochlea are sensitive to frequencies as follows [20]:
  - Hair cells close to the base are more sensitive to low-frequency sounds
  - near the apex are more sensitive to high-frequency sounds





Sound reconstruction pipeline



Time representation & Frequency representation

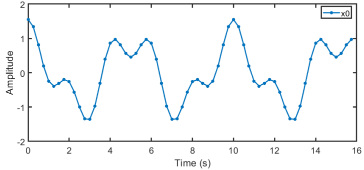
We consider a realizable sound signal $s \in L^2(\mathbb{R})$

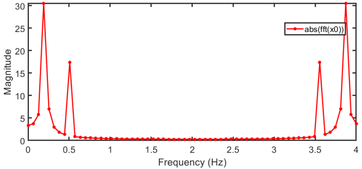
- **Frequency representation:**

$$\hat{s}(\omega) = \mathcal{F}\left\{s(t)\right\}(\omega) = \int_{\mathbb{R}} s(t)e^{-2\pi i\omega t}\mathrm{d}t$$

- **Time representation:**

$$s(t) = \mathcal{F}^{-1}\left\{\hat{s}(\omega)\right\}(t) = \int_{\mathbb{R}} \hat{s}(\omega)e^{2\pi i\omega t}\mathrm{d}\omega$$





Since $s = \mathcal{F}^{-1}\left\{\hat{s}\right\}$, we can say about s and $\hat{s}$ that they

- both contain the exact same information
- both represent the same object $s \in L^2(\mathbb{R})$
- they simply show different features of s

A time-frequency representation would combine the features of both s and $\hat{s}$ into a single function. Such representation provides an *instantaneous frequency spectrum* of the signal at any given time [11].

Short-Time Fourier Transform (STFT)

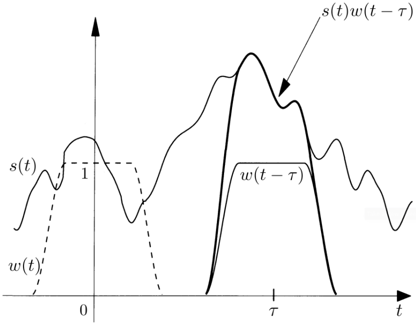
Definition (Short-Time Fourier Transform)

Let $s \in L^2(\mathbb{R})$ be a time signal, let $\omega \in L^2(\mathbb{R})$ be a compactly supported window centered around 0. The STFT of s with respect to the window w is defined as

$$S(\tau, \omega) = \mathrm{STFT}\{s(t)\}(\tau, \omega) = \int_{\mathbb{R}} s(t)w(t-\tau)e^{-2\pi i \omega t}\mathrm{d}t$$

The STFT is

- a very common time-frequency representation of a signal
- the Fourier transform of the $s(t)w(t-\tau)$, the signal taken over a sliding window along the time axis
- usually taken along a smooth window because a sharp cut-off introduces discontinuities and aliasing issues [11]



Time and frequency shifts operators

Definition (Time and frequency shifts operators)

Let $s \in L^2(\mathbb{R})$ be a time signal, we define for all $\tau, \omega \in \mathbb{R}$

- **Time shift operator:** $T_\tau s(t) = s(t - \tau)$
- **Phase shift operator:** $M_\omega s(t) = e^{2\pi i \omega t} s(t)$

We call T_τ and M_ω unitary operators in $\mathcal{U}(L^2(\mathbb{R}))$

The STFT can be formulated using these unitary operators

$$\begin{aligned} S(\tau, \omega) &= \int_{\mathbb{R}} s(t) w(t - \tau) e^{-2\pi i \omega t} \mathrm{d}t \\ &= \int_{\mathbb{R}} s(t) \overline{M_\omega T_\tau w(t)} \mathrm{d}t \\ &= \langle s, M_\omega T_\tau w \rangle_{L^2(\mathbb{R})} \end{aligned}$$

We can redefine the STFT as an operator V_w on $s \in L^2(\mathbb{R})$ defined in function of $T_\tau, M_\omega \in \mathcal{U}(L^2(\mathbb{R}))$ [4,11].

$$V_w s(\tau, \omega) = \langle s, M_\omega T_\tau w \rangle_{L^2(\mathbb{R})}$$

Discrete STFT

Similarly to the continuous STFT, the discrete STFT is the Discrete Fourier Transform (DFT) of the signal over a sliding window. Nevertheless, the window cannot slide continuously along the time axis, instead the signal is windowed at different frames with an overlap. The window therefore hops along the time axis.

Discrete STFT parameters:

- Window size (DFT size): N
- Overlap size: R
- Hop size: $H = N - R$
- Overlap ratio: $r = R/N \in [0, 1[$

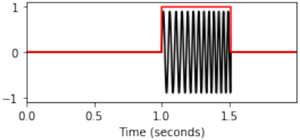
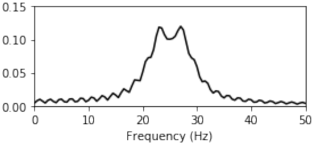
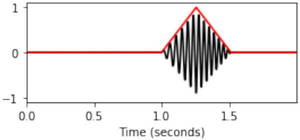
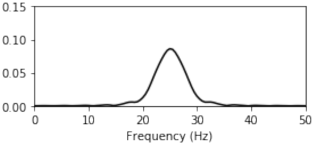
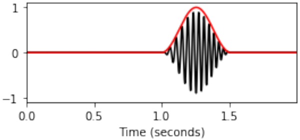
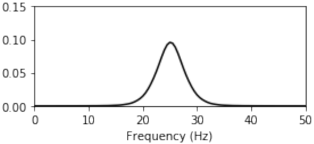
Definition (Discrete Short-Time Fourier Transform)

The discrete STFT of a signal $s \in L^2([0, T])$ over a window w is defined as

$$S[m, \omega] = \sum_{t=0}^{T} s[t]w[t - mH]e^{-2\pi i\omega t}$$

STFT windowing

The choice of the window affects the quality of the Fourier transform.

Window function	$s(t)w(t-t_0)$	$\mathcal{F}\left\{s(t)w(t-t_0)\right\}(\omega)$
Rectangular		
Triangular		
Hann		

STFT windowing - invertibility constraints

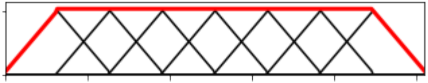
The STFT is invertible if its parameters satisfy the two following constraints [10,15]:

- **Nonzero OverLap Add (NOLA):**

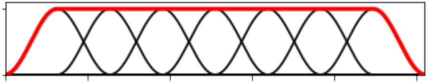
$$\sum_{m\in\mathbb{Z}} w^2[t - mH] \neq 0$$

- **Constant OverLap Add (COLA):**

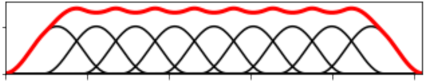
$$\sum_{m\in\mathbb{Z}} w[t - mH] = 1$$



Triangular window, overlap ratio $r = \frac{1}{2}$



Hann window, overlap ratio $r = \frac{1}{2}$



Hann window, overlap ratio $r = \frac{3}{8}$

The NOLA condition is met for any window given an overlap ratio $r \in [0, 1[$. It is worth noting that this condition can be found without the square depending on the inverse STFT algorithm.

The COLA constraint defines the partition of unity over the discrete time axis, imposing a stronger condition.

STFT windowing - Hann window

Remark

In typical applications, the window functions used are non-negative, smooth, bell-shaped curves.

In our model we use the Hann window, which satisfies the COLA condition for any overlap ratio of $r = \frac{n}{n+1}, n \in \mathbb{N}^*$.

The Hann window of length L is defined as

$$w(x) = \begin{cases} \frac{1+\cos\left(\frac{2\pi x}{L}\right)}{2} & \text{if } |x| \leq \frac{L}{2} \\ 0 & \text{if } |x| > \frac{L}{2} \end{cases}$$

Uncertainty principles

In mathematics, uncertainty principles are

- limits to the accuracy with which the values for certain physical pairs can be obeserved
- inequilities that involve pairs of complementary/disjoint variables

Common examples are

- **Heisenberg's Uncertainty Principle:** a particle's momentum and its position
- **The Heisenberg-Gabor limit:** a signal's time and frequency

Theorem (Heisenberg-Pauli-Weyl inequality)

Let $f \in L^2(\mathbb{R})$, *then* $\forall a, b \in \mathbb{R}$

$$\left(\int_{\mathbb{R}} (t-a)^2 \left|f(t)\right|^2 \mathrm{d}t\right)^{1/2} \left(\int_{\mathbb{R}} (\omega-b)^2 \left|\hat{f}(\omega)\right|^2 \mathrm{d}\omega\right)^{1/2} \geq \frac{\|f\|_2^2}{4\pi}$$

Uncertainty principle - the Heisenberg-Gabor limit

From the Heisenberg-Pauli-Weyl Inequality, we obtain the following theorem

Theorem (Heisenberg-Gabor limit)

Let $f \in L^2(\mathbb{R})$, if $\|f\|_2 = 1$ then

$$\sigma_t \cdot \sigma_\omega \geq \frac{1}{4\pi}$$

where σ_t and σ_ω are the standard deviations of the time and frequency respectively.

Interpretation of the standard deviations:

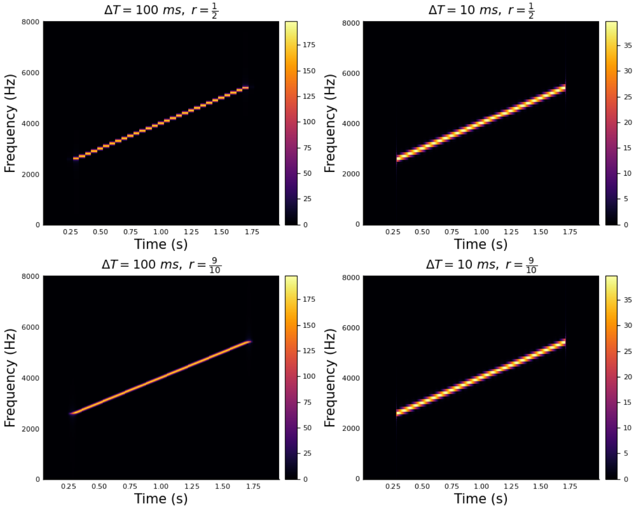
- σ_t is the size of the *essential support* of f
- σ_ω is the size of the *essential bandwidth* of the signal centered around the average frequency $\bar{\omega}$

The Gabor limit means that

- "a realizable signal occupies a region of area at least one in the time-frequency plane."
- we cannot sharply localize a signal in both the time domain and frequency domain
- the concept of an instantaneous frequency is impossible [11]

Uncertainty principle - resolution issues

STFT resolution with respect to different window sizes ΔT and overlap ratios r



Influence of the window size and the overlap ratio:

- **Window size:**
  - Larger windows $\Longrightarrow$ higher frequency resolution & lower time resolution
  - Smaller windows $\Longrightarrow$ lower frequency resolution & higher time resolution
- **Overlap:**
  - Small overlaps $\Longrightarrow$ time discontinuities & computationally cheaper
  - Big overlaps $\Longrightarrow$ more time precision & computationally costly

Inverse STFT

Theorem (Parseval's Formula for the STFT)

Consider two signals $s_1, s_2 \in L^2(\mathbb{R})$, and two windows $w_1, w_2 \in L^2(\mathbb{R})$, then

$$\langle V_{w_1} s_1, V_{w_2} s_2 \rangle_{L^2(\mathbb{R}^2)} = \langle s_1, s_2 \rangle_{L^2(\mathbb{R})} \overline{\langle w_1, w_2 \rangle}_{L^2(\mathbb{R})}$$

Proposition

If $\|w\|_2 = 1$ then the STFT operator W_w is an isometry from $L^2(\mathbb{R})$ to $L^2(\mathbb{R}^2)$.
This can be easily shown from Parseval's Formula

$$\forall s, w \in L^2(\mathbb{R}), \|V_w s\|_2 = \|s\|_2 \|w\|_2 \implies \|V_w s\|_2 = \|s\|_2, \forall s \in L^2(\mathbb{R}) \text{ if } \|w\|_2 = 1$$

Theorem (Inverse Short-Time Fourier Transform)

Let $w, h \in L^2(\mathbb{R})$ with $\langle w, h \rangle \neq 0$. Then for all $s \in L^2(\mathbb{R})$

$$s(t) = \frac{1}{\langle w, h \rangle} \iint_{\mathbb{R}^2} V_w s(\tau, \omega) M_\omega T_\tau h(t) \mathrm{d}\omega \mathrm{d}\tau = \frac{1}{\langle w, h \rangle} \iint_{\mathbb{R}^2} S(\tau, \omega) h(t - \tau) e^{2\pi i \omega t} \mathrm{d}\omega \mathrm{d}\tau$$

Inverse STFT - Griffin-Lim Algorithm [10]

Advantages:

- efficient and easy to implement
- works on modified STFT

General idea:

- Let $Y \in L^2(\mathbb{R}^2)$ be a modified STFT
- There might not be $y \in L^2(\mathbb{R})$ such that $Y = V_w y$
- The GLA finds a signal $x \in L^2(\mathbb{R})$ with $X = V_w x$ that minimizes $d(X, Y) = \|X - Y\|_2^2$
- We consider x the inverse STFT of the modified STFT Y.

Algorithm:

- Calculate $y_\tau \in L^2(R^2)$ the inverse Fourier transform of Y with respect to the frequency ω at a fixed time τ.

$$y_\tau(t) = \int_{\mathbb{R}} Y(\tau, \omega) e^{2\pi i \omega t} \mathrm{d}\omega$$

- Find iteratively the signal x that minimizes $d(X, Y)$

$$x[t] = \frac{\sum_\tau y_\tau[t] w[t - \tau]}{\sum_\tau w^2[t - \tau]}$$

The sound chirpiness

3D representation in our models

- **V1 model:** sensitivity to directions

$$\theta \in P^1 = \mathbb{R}/\pi\mathbb{Z}$$

- **A1 model:** sensitivity to sound chirpiness

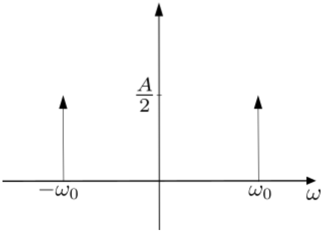
$$\nu = \frac{\mathrm{d}\omega}{\mathrm{d}\tau} \in \mathbb{R}$$

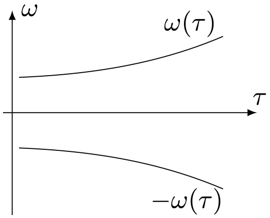
Interpretation of the *instantaneous chirpiness*:

- the time derivative of the frequency
- the slope of the frequency $w(t)$
- the tangent of the sound image directions $\tan\theta$

The sound chirpiness - single frequency spectrum

Single constant frequency	Single time-varying frequency
$s(t) = A \cdot sin(\omega_0 t)$	$s(t) = A \cdot sin(\omega(t)t)$
$\hat{s}(\omega) = \frac{A}{2i}(\delta_0(\omega - \omega_0) - \delta_0(\omega + \omega_0))$	$S(\tau, \omega) = \frac{A}{2i}(\delta_0(\omega - \omega(\tau)) - \delta_0(\omega + \omega(\tau)))$





The sound chirpiness - single time-varying frequency

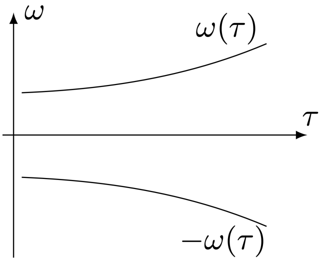
Parametric representation of the sound

$$s(t) = A \cdot sin(\omega(t)t)$$

- In the time-frequency domain: $t \mapsto (t, \omega(t))$
- In the *augmented space*: $t \mapsto (t, \omega(t), \nu(t))$

with

$$\nu(t) = \frac{\mathrm{d}\omega}{\mathrm{d}t}(t)$$



Representation in contact space - control system

What's the nature of the curve $t \mapsto (t, \omega(t), \nu(t))$?

Let's define $u(t) = \mathrm{d}\nu/\mathrm{d}t$, the curve in the contact space $t \mapsto (t, \omega(t), \nu(t))$ is *a lift of a planar curve if there exists a function* $u(t)$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\tau\\ \omega\\ \nu\end{pmatrix} = \begin{pmatrix}1\\ \nu\\ 0\end{pmatrix} + u(t)\begin{pmatrix}0\\ 0\\ 1\end{pmatrix}$$

Let $q = (\tau, \omega, \nu)$, the previous equations is the state equation of a control system written as

$$\frac{\mathrm{d}}{\mathrm{d}t}q(t) = X_0(q(t)) + u(t)X_1(q(t))$$

where $X_0(q(t))$ and $X_1(q(t))$ are two vector fields in $\mathbb{R}^3$

$$X_0\begin{pmatrix}\tau\\ \omega\\ \nu\end{pmatrix} = \begin{pmatrix}1\\ \nu\\ 0\end{pmatrix}, \quad X_1\begin{pmatrix}\tau\\ \omega\\ \nu\end{pmatrix} = \begin{pmatrix}0\\ 0\\ 1\end{pmatrix}$$

Representation in contact space - Heisenberg group

The two vector fields in $\mathbb{R}^3$

$$X_0\begin{pmatrix}\tau\\ \omega\\ \nu\end{pmatrix}=\begin{pmatrix}1\\ \nu\\ 0\end{pmatrix},\quad X_1\begin{pmatrix}\tau\\ \omega\\ \nu\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 1\end{pmatrix}$$

Generate the **Heisenberg group** because [4,11]

- $Z=[X_0,X_1]\neq 0$
- $[Z,X_0]=[Z,X_1]=0$

Lift to the contact space

We lift the each level line of the spectrum $|S|(\tau,\omega)$ to the contact space. Yeilding the following subset of the contact space, which is a well-defined surface if $|S| \in \mathcal{C}^2$ and $\mathrm{Hess}\,|S|$ is non-degenerate [4].

$$\Sigma = \left\{(\tau,\omega,\nu) \in \mathbb{R}^3 | \nu\partial_\omega |S|(\tau,\omega) + \partial_\tau |S|(\tau,\omega) = 0\right\}$$

Which allows to finally define the sound lift in the contact space as

$$L(\tau,\omega,\nu) = S(\tau,\omega)\cdot\delta_\Sigma(\tau,\omega,\nu) = \begin{cases} S(\tau,\omega) & \text{if } (\tau,\omega,\nu) \in \Sigma \\ 0 & \text{otherwise} \end{cases}$$

The time-frequency representation is obtained from the lifted sound by applying the projection operator defined as

$$\mathrm{Proj}\left\{L(\tau,\omega,\nu)\right\}(\tau,\omega) = \int_{\mathbb{R}} L(\tau,\omega,\nu)\mathrm{d}\nu$$

Cortical activations in A1 - Wilson-Cowan model

We model the cortical activations in A1 as follows

- The primary auditory cortex (A1) is a space of $(\omega, \nu) \in \mathbb{R}^2$.
- A1 receives the sound lift to the contact space $L(t, \omega, \nu)$ at every instant t.
- The *neuron* receives an external charge $S(t, \omega)$ if $(t, \omega, \nu) \in \Sigma$ and no charge otherwise.

We need to model these neural activations $\rightsquigarrow$ Wilson-Cowan model

- Successfully applied to describe neural activations in V1 and A1 [2,3,6,9,14,17,21]
- Flexible model, applies independently to the underlying geometric structure
- Geometric structure is encoded in the kernel of the integral term
- Implementation of delay terms

Wilson-Cowan equation

$$\frac{\partial}{\partial t}a(t,\omega,\nu) = -\alpha a(t,\omega,\nu) + \beta L(t,\omega,\nu) + \gamma \int_{\mathbb{R}^2} k_\delta(\omega,\nu\|\omega',\nu')\sigma(a(t-\delta,\omega',\nu'))\mathrm{d}\omega'\mathrm{d}\nu'$$

where

- $\alpha, \beta, \gamma > 0$ are (tuning) parameters
- $\sigma : \mathbb{C} \to \mathbb{C}$ is a non-linear sigmoid
  - $\sigma(\rho e^{i\theta}) = \tilde{\sigma}(\rho)e^{i\theta}$
  - $\tilde{\sigma}(x) = \min\{\max\{0, \kappa x\}, 1\}, \forall x \in \mathbb{R}$ given a fixed $\kappa > 0$
- $k_\delta(\omega,\nu\|\omega',\nu')$ is a weight modeling the interaction between (ω,ν) and (ω',ν') after a delay $\delta > 0$ via the kernel of the transport-diffusion operator associated to the contact structure of A1

Wilson-Cowan equation with no delay

When $\gamma = 0$, the WC equation becomes a standard low-pass filter

$$\partial_t a(t,\omega,\nu) = -\alpha a(t,\omega,\nu) + L(t,\omega,\nu)$$

whose solution is simply

$$a(t,\omega,\nu) = \int_0^t e^{-\alpha(s-t)} L(t,\omega,\nu) \mathrm{d}s$$

Here, ω and ν are parameters $\rightsquigarrow$ there is no interaction between regions sensitive to different ω and ν.

Wilson-Cowan equation with delayed interaction

$$\frac{\partial}{\partial t}a(t,\omega,\nu) = -\alpha a(t,\omega,\nu) + \beta L(t,\omega,\nu) + \gamma \int_{\mathbb{R}^2} k_\delta(\omega,\nu\|\omega',\nu')\sigma(a(t-\delta,\omega',\nu'))\mathrm{d}\omega'\mathrm{d}\nu'$$

With $\gamma \neq 0$, a non-linear term is added on top of the low-pass filter:

- The added term describes the diffusion of the activation in side A1
- The added term encodes the inhibitory and excitatory interconnections between neurons
- The sigmoid is a non-linear function that saturates the signal a

Section 4

Implementation

The Julia language

The Julia language is

- New
    - First appeared in 2012
    - Version 1.0 was released in 2018
- Fast: comparable to Fortran and C
- Easy to use: similar to Python, Matlab, and R
- General-purpose
- Great for scientific computing

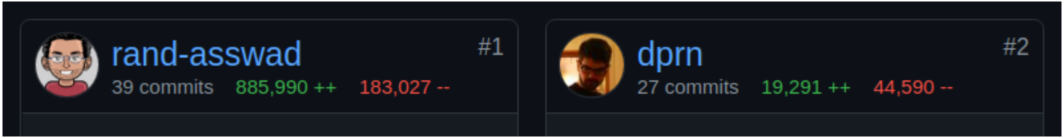
Julia community is small: in 2021 Stack Overflow Developer Survey [22] "Which language developers wanted to work in over the next year?"

- Julia: 1.29%
- Python: 48.24%
- Matlab: 4.66%

Result: less stable scientific libraries in Julia than other languages

The `WCA1.jl` package

- Original code: `https://github.com/dprn/WCA1`
- Forked repository: `https://github.com/rand-asswad/WCA1`



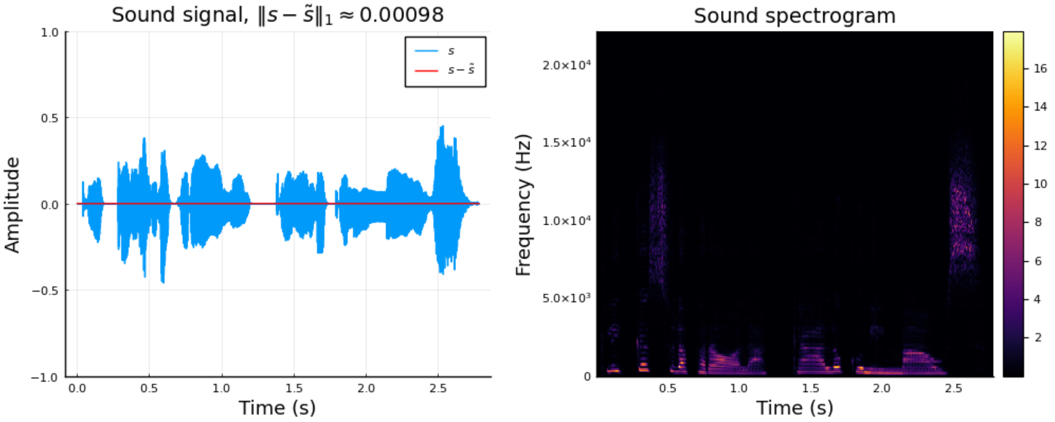
Issues with original code:

- Unstable $\rightsquigarrow$ failed to run on speech signals
- Far from optimal $\rightsquigarrow$ took long time to run on speech signals
- Low readability
- Non-conforming to Julia's code norms and performance recommendations

The STFT module

Issue: no implementation of the inverse STFT in Julia's standard libraries (FFTW.jl andDSP.jl).

Solution: implemented the Griffin-Lim algorithm [10] from scratch



The Lift module - calculating chirpiness values

The sound chirpiness is defined as

$$\nu\partial_\omega |S| (\tau, \omega) + \partial_\tau |S| (\tau, \omega) = 0$$

We compute the chirpiness with respect to each time-frequency pair by calculating the gradient of the spectrum $\nabla |S|$.

$$\nu(\tau, \omega) = \begin{cases} -\frac{\partial_\tau |S|(\tau,\omega)}{\partial_\omega |S|(\tau,\omega)} & \text{if } |\partial_\omega |S| (\tau, \omega)| > \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

where ε is a small threshold.

The Lift module - chirpiness sampling issue

Issue: the chipriness values ν are unbounded since

$$\nu\partial_\omega |S| (\tau, \omega) + \partial_\tau |S| (\tau, \omega) = 0$$

and there exists points (τ_0, ω_0) such that $\partial_\omega |S| (\tau_0, \omega_0) = 0$

therefore chirpiness values stretch over the entire real line (coverge to $\pm\infty$)

Original solution: manually restrict chirpiness values to $\nu \in [\nu_{\text{min}}, \nu_{\text{max}}]$ for synthetic signals (the limits are determined after visualizing the histogram of the chirpiness values).

Needed solution: a reliable method to automatically determine the interval $[\nu_{\text{min}}, \nu_{\text{max}}]$ without losing (a lot of) values.

The Lift module - chirpiness values distribution

We noticed that the chirpiness values of speech signals follow a Cauchy distribution [1]

Let X be a random variable following $\mathrm{Cauchy}(x_0, \gamma)$

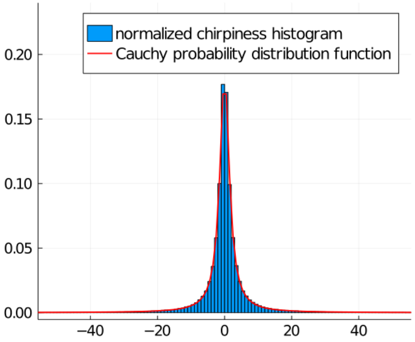
- **Location parameter** x_0**:** location of the peak
- **Scale parameter** γ**:** half the interquartile range

Probability density function (PDF):

$$f_X(x) = \frac{1}{\pi\gamma\left(1 + \left(\frac{x - x_0}{\gamma}\right)^2\right)}$$

Cumulative distribution function (CDF):

$$F_X(x) = \frac{1}{\pi}\arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}$$



The Lift module - chirpiness values distribution

Estimating Cauchy parameters $\mathrm{Cauchy}(x_0, \gamma)$:

- x_0: the chirpiness samples median
- γ: half the interquartile range (difference between the 75th and the 25th percentile)

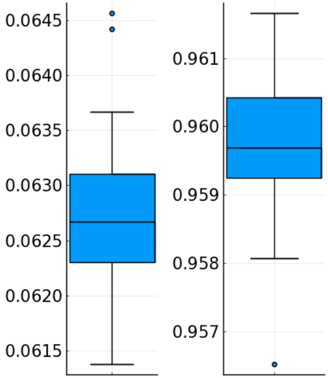
Assumption:

$$\nu \sim \mathrm{Cauchy}\left(\mathrm{median}(\nu), \frac{Q(75\%) - Q(25\%)}{2}\right)$$

Stastical tests on a library of real speech signals **rejected** the assumption.
Nevertheless, the fit is quite good according to the Kolomogorov-Smirnov statistic

$$D_n = \sup_x |F_n(x) - F_X(x)|$$

where F_n is the empirical distribution function



Box plots for estimated Cauchy distributions of speech signals chirpiness values

- *left:* Kolmogorov-Smirnov statistic values.
- *right:* percentage of values falling in $I_{0.95}$

The Lift module - chirpiness sampling

1. Calculate chirpiness values for each (τ, ω)
2. Compute values to Cauchy distribution to find confidence interval $I_p = [\nu_{\min}, \nu_{\max}]$
3. Discretize chirpiness values $\nu \in I_p$ as follows

Let $(\nu_n)_{1 \leq n \leq N}$ such that $\nu_{\min} = \nu_1 < \cdots < \nu_N = \nu_{\max}$.

Each value ν is rounded to the nearest ν_n.

$$n(\nu) = \left\lfloor \frac{\nu - \nu_{\min}}{\nu_{\max} - \nu_{\min}}(N-1) + 1 \right\rceil, \quad \forall \nu \in I_p$$

where $\lfloor \cdot \rceil : \mathbb{R} \to \mathbb{Z}$ is the rounding function to the nearest integer.

The Lift module - chirpiness sampling optimization

The function $n(\nu)$ can be optimized by rewriting it as an affine function

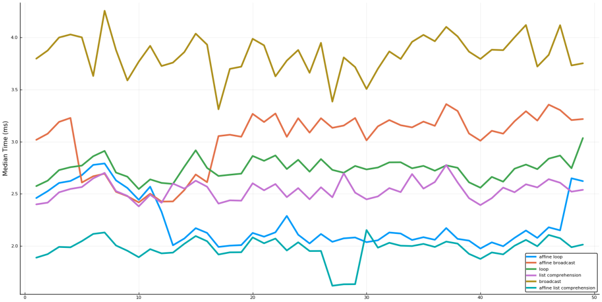
$$n(\nu) = \left\lfloor \frac{\nu - \nu_{\text{min}}}{\nu_{\text{max}} - \nu_{\text{min}}}(N-1) + 1 \right\rceil = \left\lfloor \underbrace{\left(\frac{N-1}{\nu_{\text{max}} - \nu_{\text{min}}} \right)}_{a} \cdot \nu + \underbrace{\left(1 - \frac{(N-1)\nu_{\text{min}}}{\nu_{\text{max}} - \nu_{\text{min}}} \right)}_{b} \right\rceil = \lfloor a \cdot \nu + b \rceil$$

This reduces the number of arithmetic operations inside the loop in $O(n)$ complexity.

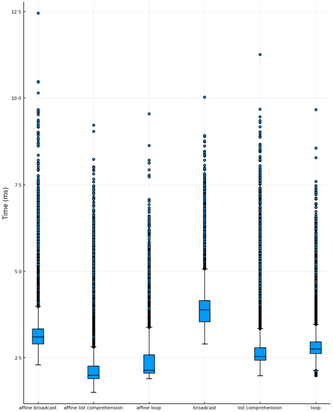
```
function discretize_chirpiness(v::AbstractMatrix, vMin::Real, vMax::Real, N::Integer=100)
    a = (N-1) / (vMax - vMin)
    b = 1.0 - ((N-1) * vMin / (vMax - vMin))
    index = [vMin ≤ x ≤ vMax ? round(Int, a*x + b) : nothing for x in v]
    return index, range(vMin, vMax, length = N)
end
```

The Lift module - chirpiness sampling benchmark

Using Julia's standard benchmark tools, we ran a benchmark on the speech library samples with different chirpiness implementations.



The benchmarked median time for each method ploted against the speech samples

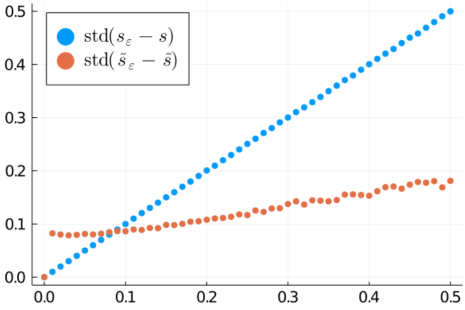


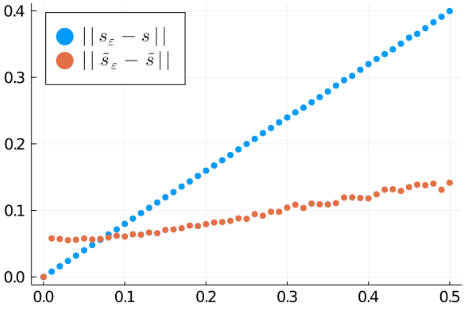
Box plots of the benchmarked time for each method on the speech samples

Denoising experiment [1]

We apply a gaussian random noise $g_\varepsilon \sim \mathcal{N}(0, \varepsilon)$ to a an input sound s, we process the noisy sound input through the algorithm pipeline

- **Input:** $s_\varepsilon = s + g_\varepsilon$
- **Output:** $\tilde{s}_\varepsilon = \mathrm{STFT}^{-1} \circ \mathrm{Proj} \circ \mathrm{WC} \circ \mathrm{Lift} \circ \mathrm{STFT}(s_\varepsilon)$





Distance of noisy sound to original one before (blue) and after (red) the processing, plotted against the standard deviation of the noise ε (where $\|s\| = \|s\|_1 \,/\, \dim(s)$)

Section 5

Conclusion

Model analysis

The sound reconstruction model:

- improves noisy speech signals
- is mathematically stable
- has great potential

Conclusion:

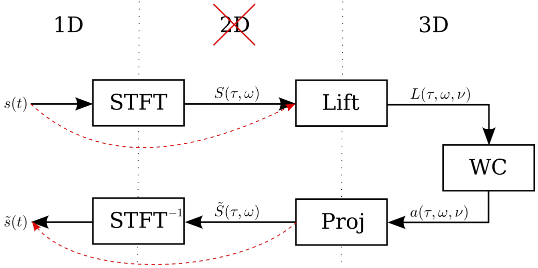
- the model should be improved and adapted to more corrupted sounds
- the model deserves to be the basis of a PhD project

We will see the paths we explored to improve the model

Model analysis - Lift drawbacks

- The lifted representation $L(\tau, \omega, \nu) = S(\tau, \omega)\delta_\sigma(\tau, \omega, \nu)$ depends on the phase of $S(\tau, \omega) \in \mathbb{C}$. This is unrealistic, since the cochlea only transmits the spectrogram $|S(\tau, \omega)|$ because A1 is insensitive to phase.
- At a fixed time $t > 0$, the resulting representation $L(t, \omega, \nu)$ is a distribution, concentrated on a one dimensional curve in the frequency-chirpiness space which is also unrealistic.
- The current procedure to obtain $L(\tau, \omega, \nu)$ requires to first compute $S(\tau, \omega)$ and then to "lift" it. We would like to obtain L directly from the original signal s.

To improve the model, it is crucial to devise a novel lift procedure allowing to bypass these problems.



Alternative sound reconstruction pipeline

Model analysis - Wavelet Transform

By reading state-of-the-art literature on the neurophysiology of the inner ear, we realized that a Wavelet transform represents the signal processing in the cochlea than the STFT transform [18,20].

Definition (Wavelet Transform)

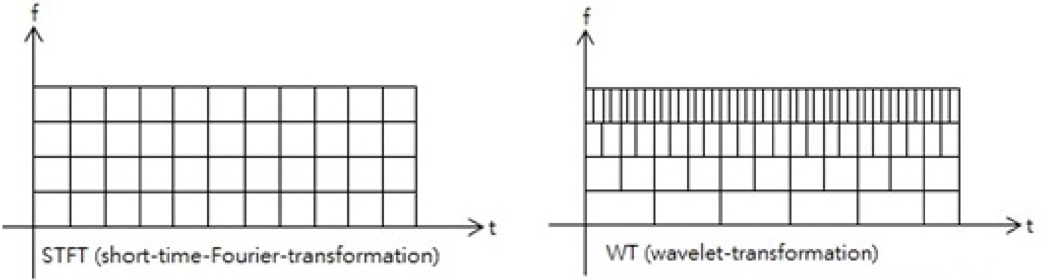
The Wavelet Transform (WT) of a realizable signal $s \in L^2(\mathbb{R})$ along a wavelet $\psi \in L^2(\mathbb{R})$ is defined by

$$W_\psi s(a, t) = \frac{1}{\sqrt{a}} \int_{\mathbb{R}} s(\tau) \overline{\psi\left(\frac{\tau - t}{a}\right)} \mathrm{d}\tau$$

where a is the dilation variable.

Model analysis - Wavelet Transform

Advantage: time resolution increases for higher frequencies in the WT.



Disadvantage: the dilation variable *a* implicitly represents the frequency ω.

- Obtaining the sound chirpiness ν is not straightforward as in the case of the STFT
- We haven't been able to define an appropriate lift from the WT

Model analysis - the lift operator

We defined the STFT as operator on $L^2(\mathbb{R})$ in function of the unitary shift operators

$$V_w s(\tau, \omega) = \langle s, M_\omega T_\tau w \rangle_{L^2(\mathbb{R})}$$

We would like to have

$$L_\gamma s(\tau, \omega, \nu) = \langle s, C_\nu M_\omega T_\tau \gamma \rangle_{L^2(\mathbb{R})}$$

where $C_\nu \in \mathcal{U}(L^2(\mathbb{R}))$

Such operator would be

- Mathematically stable and elegant
- Computationally cheap

Acquired knowledge

- Fundamental mathematics
  - Geometry
  - Group representations
  - Operator algebra
  - Time-Frequency analysis
- Applied mathematics & programming
  - Signal processing & DSP
  - Julia language
  - Neural activations models
- The neuro-physiology of the inner ear
- Research experience
  - Studying state-of-the-art litterature
  - Co-writing a conference paper
  - Attending the GSI 2021 conference

My future project

After my internship, I have decided to pursue

- a Master's degree in fundamental mathematics at Université de Lorraine, focusing on PDEs and Control Theory
- a PhD thesis in the domains of PDEs and Control Theory
- a career in academic research

Thank you for your attention!

References I

[1] Rand Asswad, Ugo Boscain, Giuseppina Turco, Dario Prandi, and Ludovic Sacchelli. 2021. An Auditory Cortex Model for Sound Processing.. 56–64. DOI:https://doi.org/10.1007/978-3-030-80209-7_7

[2] Marcelo Bertalmío, Luca Calatroni, Valentina Franceschi, Benedetta Franceschiello, and Dario Prandi. 2018. A cortical-inspired model for orientation-dependent contrast perception: A link with Wilson-Cowan equations. *arXiv:1812.07425 [cs]* (December 2018). Retrieved November 12, 2020 from http://arxiv.org/abs/1812.07425

[3] Ugo Boscain, Roman Chertovskih, Jean-Paul Gauthier, Dario Prandi, and Alexey Remizov. 2017. Cortical-inspired image reconstruction via sub-Riemannian geometry and hypoelliptic diffusion. In *SMAI 2017 - 8e Biennale Française des Mathématiques Appliquées et Industrielles*, La Tremblade, France, 37–53. DOI:https://doi.org/10.1051/proc/201864037

[4] Ugo Boscain, Dario Prandi, Ludovic Sacchelli, and Giuseppina Turco. 2021. A bio-inspired geometric model for sound reconstruction. *The Journal of Mathematical Neuroscience* 11, 1 (January 2021), 2. DOI:https://doi.org/10.1186/s13408-020-00099-4

References II

[5] Paul C. Bressloff and Jack D. Cowan. 2002. An Amplitude Equation Approach to Contextual Effects in Visual Cortex. *Neural Computation* 14, 3 (March 2002), 493–525. DOI:https://doi.org/10.1162/089976602317250870

[6] Paul C. Bressloff, Jack D. Cowan, Martin Golubitsky, Peter J. Thomas, and Matthew C. Wiener. 2002. What Geometric Visual Hallucinations Tell Us about the Visual Cortex. *Neural Computation* 14, 3 (March 2002), 473–491. DOI:https://doi.org/10.1162/089976602317250861

[7] G. Citti and A. Sarti. 2006. A Cortical Based Model of Perceptual Completion in the Roto-Translation Space. *Journal of Mathematical Imaging and Vision* 24, 3 (May 2006), 307–326. DOI:https://doi.org/10.1007/s10851-005-3630-2

[8] P. Dallos. 1996. Overview: Cochlear Neurobiology: Springer Handbook of Auditory Research. *The Cochlea: Springer Handbook of Auditory Research* (1996), 1–43. Retrieved August 13, 2021 from https://www.scholars.northwestern.edu/en/publications/overview-cochlear-neurobiology-springer-handbook-of-auditory-rese

References III

[9] G. B. Ermentrout and J. D. Cowan. 1979. A mathematical theory of visual hallucination patterns. *Biological Cybernetics* 34, 3 (October 1979), 137–150. DOI:https://doi.org/10.1007/BF00336965

[10] D. Griffin and Jae S. Lim. 1983. Signal estimation from modified short-time Fourier transform. *undefined* (1983). Retrieved September 3, 2021 from https://www.semanticscholar.org/paper/Signal-estimation-from-modified-short-time-Fourier-Griffin-Lim/14bc876fae55faf5669beb01667a4f3bd324a4f1

[11] Karlheinz Gröchenig. 2001. *Foundations of Time-Frequency Analysis*. Birkhäuser Basel. DOI:https://doi.org/10.1007/978-1-4612-0003-1

[12] William C. Hoffman. 1989. The visual cortex is a contact bundle. *Applied Mathematics and Computation* 32, 2 (August 1989), 137–167. DOI:https://doi.org/10.1016/0096-3003(89)90091-X

[13] D. H. Hubel and T. N. Wiesel. 1959. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology* 148, 3 (1959), 574–591. DOI:https://doi.org/10.1113/jphysiol.1959.sp006308

References IV

[14] Alex Loebel, Israel Nelken, and Misha Tsodyks. 2007. Processing of sounds by population spikes in a model of primary auditory cortex. *Frontiers in Neuroscience* 1, 1 (November 2007), 197–209. DOI:https://doi.org/10.3389/neuro.01.1.1.015.2007

[15] Meinard Müller. 2015. *Fundamentals of Music Processing - Audio, Analysis, Algorithms, Applications*. Springer. Retrieved from https://www.audiolabs-erlangen.de/fau/professor/mueller/bookFMP

[16] Jean Petitot and Yannick Tondut. 1999. Vers une neurogéométrie. Fibrations corticales, structures de contact et contours subjectifs modaux. *Mathématiques et Sciences Humaines* 145, (1999), 5–101. Retrieved August 13, 2021 from https://eudml.org/doc/94522

[17] James Rankin, Elyse Sussman, and John Rinzel. 2015. Neuromechanistic Model of Auditory Bistability. *PLoS computational biology* 11, 11 (November 2015), e1004555. DOI:https://doi.org/10.1371/journal.pcbi.1004555

[18] Hans Martin Reimann. 2011. Signal processing in the cochlea: The structure equations. *The Journal of Mathematical Neuroscience* 1, 1 (June 2011), 5. DOI:https://doi.org/10.1186/2190-8567-1-5

References V

[19] Hugh R. Wilson and Jack D. Cowan. 1972. Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons. *Biophysical Journal* 12, 1 (January 1972), 1–24. DOI:https://doi.org/10.1016/S0006-3495(72)86068-5

[20] Xiaowei Yang, Kuansan Wang, and Shihab Shamma. 1992. Auditory representations of acoustic signals. *Information Theory, IEEE Transactions on* 38, (April 1992), 824–839. DOI:https://doi.org/10.1109/18.119739

[21] Isma Zulfiqar, Michelle Moerel, and Elia Formisano. 2019. Spectro-Temporal Processing in a Two-Stream Computational Model of Auditory Cortex. *Frontiers in Computational Neuroscience* 13, (2019), 95. DOI:https://doi.org/10.3389/fncom.2019.00095

[22] 2021. Stack Overflow Developer Survey 2021. Retrieved from https://insights.stackoverflow.com/survey/2021#section-most-popular-technologies-programming-scripting-and-markup-languages