# A bio-inspired geometric model for sound reconstruction

Ugo Boscain, Dario Prandi, Ludovic Sacchelli, Giuseppina Turco

# A bio-inspired geometric model for sound reconstruction

Ugo Boscain* Dario Prandi† Ludovic Sacchelli‡
Giuseppina Turco§

October 19, 2020

The reconstruction mechanisms built by the human auditory system during sound reconstruction are still a matter of debate. The purpose of this study is to propose a mathematical model of sound reconstruction based on the functional architecture of the auditory cortex (A1). The model is inspired by the geometrical modelling of vision, which has undergone a great development in the last ten years. There are however fundamental dissimilarities, due to the different role played by time and the different group of symmetries. The algorithm transforms the degraded sound in an 'image' in the time-frequency domain via a short-time Fourier transform. Such an image is then lifted in the Heisenberg group and is reconstructed via a Wilson-Cowan differo-integral equation. Preliminary numerical experiments are provided, showing the good reconstruction properties of the algorithm on synthetic sounds concentrated around two frequencies.

## 1. Introduction

Listening to speech requires the capacity of the auditory system to map incoming sensory input to lexical representations. When the sound is intelligible, this mapping ("recognition") process is successful. With reduced intelligibility (e.g., due to background noise), the listener has to face the task of recovering the loss of acoustic information. This task is very complex as it requires a higher cognitive load and the ability of repairing

*CNRS, LJLL, Sorbonne Université, Université de Paris, Inria, Paris, France. ugo.boscain@upmc.fr
†Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France. dario.prandi@centralesupelec.fr
‡Univ. Lyon, Université Claude Bernard Lyon 1, CNRS, LAGEPP UMR 5007, 43 bd du 11 novembre 1918, F-69100 Villeurbanne, France. ludovic.sacchelli@univ-lyon1.fr
§CNRS, Laboratoire de Linguistique Formelle, Université de Paris, France. gturco@linguist.univ-paris-diderot.fr

missing input. (See [28] for a review on noise in speech.) Yet, (normal hearing) humans are quite able to recover sounds in several effortful listening situations (see for instance [27]), ranging from sounds degraded at the source (e.g., hypoarticulated and pathological speech), during transmission (e.g., reverberation) or corrupted by the presence of environmental noise.

So far, work on degraded speech has informed us a lot on the acoustic cues that help the listener to reconstruct missing information (e.g.,[18, 31]); the several adverse conditions in which listeners may be able to reconstruct speech sounds (e.g., [2, 28]); and whether (and at which stage of the auditory process) higher-order knowledge (i.e., our information about words and sentences) helps the system to recover lower-level perceptual information (e.g., [22]). However, most of these studies adopt a phenomenological and descriptive approach. More specifically, techniques from previous studies consist in adding synthetic noise to speech sound stimuli, performing spectral and temporal analyses on the stimuli with noise and the same ones without it so to identify acoustic differences, linking the results of these analyses with the outcome from perceptual experiments. In some of these behavioural experiments, for instance, listeners are asked to identify speech units (such as consonants or words) when listening the noisy stimuli. Their accuracy scores provide a measure to the listeners' speech recognition ability.

As it stands, a mathematical model informing us on how the human auditory system is able to reconstruct a degraded speech sound is still missing. The aim of this study is to build a neuro-geometric model for sound reconstruction, stemming from the description of the functional architecture of the auditory cortex.

## 1.1. Modelling the auditory cortex

Knowledge about the functional architecture of the auditory cortex is scarce, and there are difficulties in the application of Gestalt principles for auditory perception. For these reasons, the model we propose is strongly inspired by recent advances in the mathematical modeling of the functional architecture of the primary visual cortex and the processing of visual inputs [24, 32, 13, 9], which recently yield very successful applications to image processing [20, 16, 35, 10]. This idea is not new: neuroscientists take models of V1 as a starting point for understanding the auditory system (see, e.g., [30] for a comparison, and [23] for a related discussion in speech processing). Indeed, biological similarities between the structure of the primary visual cortex (V1) and the primary auditory cortex (A1) are well-known to exist.

An often cited V1-A1 similarity is their "topographic" organization, a general principle determining how visual and auditory inputs are mapped to those neurons responsible for their processing [38]. Substantial evidence for V1-A1 relation is also provided by studies on animals and on humans with deprived hearing or visual functions showing cross-talk interactions between sensory regions [41, 44]. More relevant for our study is the existence of receptive fields of neurons in V1 and A1 that allow for a subdivision of neurons in "simple" and "complex" cells, which supports the idea of a "common canonical processing algorithm within cortical columns" [42, p. 1]. Together with the appearance in A1 of singularities typical of V1 (e.g., pinwheels) [41, 34], these findings

speak in favor of the idea that V1 and A1 share similar mechanisms of sensory input reconstruction. In the next section we present the mathematical model for V1 that will be the basis for our sound reconstruction algorithm.

## 1.2. Neuro-geometric model of V1

The neuro-geometric model of V1 finds its roots in the experimental results of Hubel and Wiesel [25], which inspired Hoffman [24] to model V1 as a *contact space*[1]. This model has then been extended to the so-called sub-Riemannian model in [33, 13, 9, 8]. On the basis of such a model, exceptionally efficient algorithms for image inpainting have been developed (e.g., [10, 15, 16]). These algorithms have now several medical imaging applications (e.g., [45]).

The main idea behind this model is that an image, seen as a function $f : \mathbb{R}^2 \to \mathbb{R}_+$ representing the grey level, is lifted to a distribution on $\mathbb{R}^2 \times P^1$, the bundle of directions of the plane[2]. Here, $P^1$ is the projective line, i.e., $P^1 = \mathbb{R}/\pi\mathbb{Z}$. More precisely, the lift is given by $Lf(x, y, \theta) = \delta_{Sf}(x, y, \theta)f(x, y)$ where $\delta_{S_f}$ is the Dirac mass supported on the set $S_f \subset \mathbb{R}^2 \times P^1$ of points $(x, y, \theta)$ such that $\theta$ is the direction of the tangent line to $f$ at $(x, y)$. Notice that, under suitable regularity assumptions on $f$, $S_f$ is a surface.

When $f$ is corrupted (i.e. when $f$ is not defined in some region of the plane), the lift is corrupted as well and the reconstruction is obtained by applying a deeply anisotropic diffusion adapted to the problem. Such diffusion mimics the flow of information along the horizontal and vertical connections of V1 and uses as an initial condition the surface $S_f$ and the values of the function $f$. Mathematicians call such a diffusion the *sub-Riemannian diffusion* in $\mathbb{R}^2 \times P^1$, cf. [29, 1]. One of the main features of this diffusion is that it is invariant by rototranslation of the plane, a feature that will not be possible to translate to the case of sounds, due to the special role of the time variable.

In what follows, we explain how similar ideas could be translated to the problem of sound reconstruction.

---

[1] A 3 dimensional manifold $M$ becomes a *contact space* once it is endowed with a smooth map $M \ni q \mapsto \mathcal{D}(q)$ where $\mathcal{D}(q)$ is a a plane in the tangent space $T_qM$ passing from $q$. There is an additional requirement on this map. Locally one can always write $\mathcal{D}(q) = \text{span}\{X_0(q), X_1(q)\}$, where $X_0$ and $X_1$ are two smooth vector fields. Then at every point $q$ one should require $\dim(\text{span}_q\{X_0, X_1, [X_0, X_1]\}) = 3$. Here $[\cdot, \cdot]$ is the Lie bracket of the vector fields. The main consequence of this condition is that no surface can be tangent to $\mathcal{D}$ at all points.

By assigning to every $\mathcal{D}(q)$ an inner (Euclidean) product that is smooth as function of $q$, we endow $M$ with a *sub-Riemannian structure*. The simplest way of defining locally such a structure on a 3 dimensional manifold is to assign two vector fields $X_0$ and $X_1$ postulating on one side that $\mathcal{D}(q) = \text{span}\{X_0(q), X_1(q)\}$ (assigning in this way the contact structure) and on the other side that they have norm one and that they are mutually orthogonal (assigning in this way the inner product).

The simplest example of sub-Riemannian structure on $\mathbb{R}^3$ is given by the so called Heisenberg group for which the vector fields $X_0 = (1, \nu, 0)^\top$ and $X_1 = (0, 0, 1)^\top$ are orthonormal (here we write coordinates in $\mathbb{R}^3$ as $(\tau, \omega, \nu)$). Such a structure is called Heisenberg group since defining $X_2 = (0, 1, 0)^\top$ one has the Lie brackets $[X_0, X_1] = X_2$, $[X_0, X_2] = [X_1, X_2] = 0$, that are the commutation relations appearing in quantum mechanics.

[2] Note that in mathematics, the term "direction" corresponds to what neurophysiologists call "orientation" and viceversa. In this study, we use the mathematical terminology

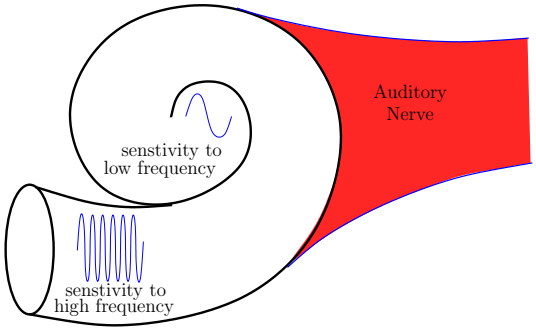



Figure 1: Perceived pitch of a sound depends on the location in the cochlea that the sound wave stimulated. High-frequency sound waves, which correspond to high-pitched noises, stimulate the basal region of the cochlea. Low-frequency sound waves are targeted to the apical region of the cochlear structure and correspond with low-pitched sounds.

## 1.3. From V1 to sound reconstruction

The sensory input reaching A1 comes directly from the cochlea [14]: a spiral-shaped, fluid-filled, cavity that composes the inner ear. Vibrations coming from the ossicles in the middle ear are transmitted to the cochlea, where they propagate and are picked up by sensors (so-called hair cells). These sensors are tonotopically organized along the spiral ganglion of the cochlea in a frequency-specific fashion, with cells close to the base of the ganglion being more sensitive to low-frequency sounds and cells near the apex more sensitive to high-frequency sounds, see Figure 1. This early 'spectrogram' of the signal is then transmitted to higher-order layers of the auditory cortex.

Mathematically speaking, this means that when we hear a sound (that we can think as represented by a function $s : [0, T] \to \mathbb{R}$) our primary auditory cortex A1 is fed by its time-frequency representation[3] $S : [0, T] \times \mathbb{R} \to \mathbb{C}$. If, say, $s \in L^2(\mathbb{R}^2)$ the time-frequency representation $S$ is given by the short-time Fourier transform of $s$, defined as

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

Here, $W : \mathbb{R} \to [0, 1]$ is a compactly supported (smooth) window, so that $S \in L^2(\mathbb{R}^2)$. Since $S$ is complex-valued, it can be thought as the collection of two black-and-white images: $|S|$ and $\arg S$. The function $S$ depends on two variables: the first one is time, that here we indicate with the letter $\tau$, and the second one is frequency, denoted by $\omega$. Roughly speaking, $|S(\tau, \omega)|$ represents the strength of the presence of the frequency $\omega$ at time $\tau$. In the following, we call $S$ the sound image (see Figure 2).

A first attempt to model the process of sound reconstruction into A1 is to apply the algorithm for image reconstruction described in Section 1.2. In a sound image, however, time plays a special role. Indeed:

[3] Actually, its spectrogram $|S| : [0, T] \times \mathbb{R} \to [0, +\infty)$, see Remark 2.

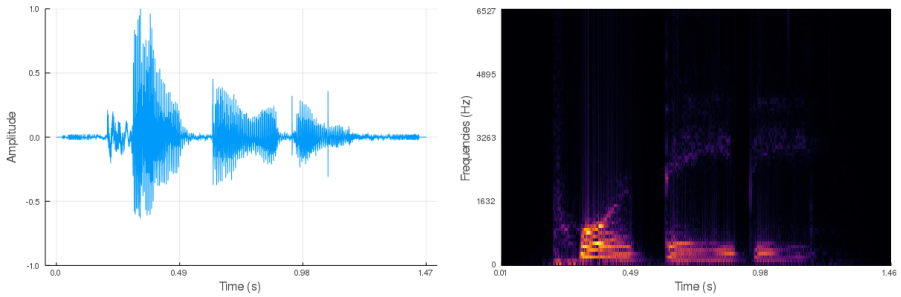



Figure 2: *Left.* Sound signal. *Right.* The corresponding short-time Fourier transform.

1. While for images the reconstruction can be done by evolving the whole image simultaneously, the whole sound image 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 and thus the invariance by rototranslations is lost.

As a consequence, different symmetries have to be taken into account (see Appendix B) and a different model for both the lift and the processing in the lifted space is required.

In order to introduce this model, let us recall that, in V1, neural stimulation stems not only from the input but also from its variations. That is, mathematically speaking, the input image is considered as a real valued function on a 2-dimensional space, and the orientation sensitivity arises from the sensitivity to a first order derivative information on this function, i.e., the tangent directions to level lines. This additional variational information allows to lift the 2-dimensional image space to the aforementioned contact space, and to define the sub-Riemannian diffusion [1, 11].

In our model of A1, we follow the same idea: we consider the variations of the input as additional variables. Input sound signals are time dependent real-valued functions subjected to a short time Fourier transform by the cochlea. As a result the A1 input is considered as a function of time and frequency. The first time derivative $\nu = d\omega/d\tau$ of this object, corresponding to the instantaneous chirpiness of the sound, allows to add a supplementary dimension to the domain of the input. As in the case of V1, this gives rise to a natural lift of the signal to an *augmented space*, which in this case turns out to be $\mathbb{R}^3$ with the Heisenberg group structure. (This structure very often appears in signal processing, see for instance [21] and Appendix B.)

As we already mentioned, the special role played by time in sound signals does not permit to model the flow of information as a pure hypoelliptic diffusion, as was done for static images in V1. We thus turn to a different kind of model: Wilson-Cowan equations [43]. Such a model, based on an integro-differential equation, has been successfully applied to describe the evolution of neural activations. In particular, it allowed to theoretically predict complex perceptual phenomena in V1, such as the emergence of hallucinatory patterns [17, 12], and has been used in various computational models of

the auditory cortex [26, 37, 46]. Recently, these equations have been coupled with the neuro-geometric model of V1 to great benefit. For instance, in [5, 4, 6] they allowed to replicate orientation-dependent brightness illusory phenomena, which had proved to be difficult to implement for non-cortical-inspired models. See also [39], for applications to the detection of perceptual units.

On top of these positive results, Wilson-Cowan equations present many advantages from the point of view of A1 modelling: i) they can be applied independently of the underlying structure, which is only encoded in the kernel of the integral term; ii) they allow for a natural implementation of delay terms in the interactions; iii) they can be easily tuned via few parameters with a clear effect on the results. On the basis of these positive results, we emulate this approach in the A1 context. Namely, we will consider the lifted sound image $I(\tau, \omega, \nu)$ to yield an A1 activation $a(\tau, \omega, \nu)$ via the following Wilson-Cowan equations:

$$\partial_t a(t,\omega,\nu) = -\alpha a(t,\omega,\nu) + \beta I(t,\omega,\nu) + \gamma \int_{\mathbb{R}^2} k_\delta(\omega,\nu\|\omega',\nu')\sigma(a(t-\delta,\omega',\nu'))\,d\omega'\,d\nu'. \quad \text{(WC)}$$

Here, $(t, \omega, \nu)$ are coordinates on the augmented space corresponding to time, frequency, and chirpiness, respectively; $\alpha, \beta, \gamma > 0$ are parameters; $\sigma : \mathbb{C} \to \mathbb{C}$ is a non-linear sigmoid; $k_\delta(\omega, \nu\|\omega', \nu')$ is a weight modelling the interaction between $(\omega, \nu)$ and $(\omega', \nu')$ after a delay of $\delta > 0$. The presence of this delay term models the fact that the time-scale of the input signal and of the neuronal activation are comparable.

The proposed algorithm to process a sound signal $s : [0, T] \to \mathbb{R}$, is the following:

A. **Preprocessing:**
   a) Compute the time-frequency representation $S : [0, T] \times \mathbb{R} \to \mathbb{C}$ of $s$, via standard short time Fourier transform (STFT);
   b) Lift this representation to the Heisenberg group, which encodes redundant information about chirpiness, obtaining $I : [0, T] \times \mathbb{R} \times \mathbb{R} \to \mathbb{C}$ (see Section 2.1 for details);

B. **Processing:** Process the lifted representation $I$ via a Wilson-Cowan equations adapted to the Heisenberg structure, obtaining $a : [0, T] \times \mathbb{R} \times \mathbb{R} \to \mathbb{C}$.

C. **Postprocessing:** Project $a$ to the processed time-frequency representation $\hat{S} : [0, T] \times \mathbb{R} \to \mathbb{C}$ and then apply an inverse STFT to obtain the resulting sound signal $\hat{s} : [0, T] \to \mathbb{R}$.

**Remark 1.** *All the above operations can be performed in real-time, as they only require the knowledge of the sound on a short window* $[t - \delta, t + \delta]$.

**Remark 2.** *Notice that in the presented algorithm we are assuming neural activations to be complex-valued functions, due to the use of the STFT. This is inconsistent with neural modelling, as it is known that the cochlea sends to A1 only the spectrogram of the STFT*

*(that is $|S|$), see [40]. If striving for a biologically plausible description, one can easily modify the above algorithm in this direction (i.e., by computing the lifted representation $I$ starting from $|S|$ instead than $S$). However, during the post-processing phase, in order to invert the STFT and obtain an audible signal, one then needs to reconstruct the missing phase information via heuristic algorithms. See, for instance [19].*

### 1.4. Structure of the paper

In Section 2, we present the reconstruction model. We first present the lift procedure of a sound signal to a function on the augmented space, and then introduce the Wilson-Cowan equations modelling the cortical stimulus. In Section 3, we describe the numerical implementation of the algorithm, together with some of its crucial properties. This implementation is then tested in Section 4, were we show the results of the algorithm on some simple synthetic signals. Such numerical examples can be listened at `www.github.com/dprn/WCA1`, and should be considered as a very preliminary step toward the construction of an efficient cortical-inspired algorithm for sound reconstruction. Finally, in Appendix B, we show how the proposed algorithm preserves the natural symmetries of sound signals.

## 2. The reconstruction model

As discussed in the introduction, the cochlea decomposes the input sound $s : [0, T] \to \mathbb{R}$ in its time-frequency representation $S : [0, T] \times \mathbb{R} \to \mathbb{C}$, obtained via a short-time Fourier transform (STFT). This corresponds to interpret the "instantaneous sound" at time $\tau \in [0, T]$, instead of as a sound level $s(\tau) \in \mathbb{R}$, as a function $\omega \mapsto S(\tau, \omega)$ which encodes the instantaneous presence of each given frequency, with phase information.

### 2.1. The lift to the augmented space

In this section, we present an extension of the time-frequency representation of a sound, which is at the core of the proposed algorithm. Roughly speaking, the instantaneous sound will be represented as a function $(\omega, \nu) \mapsto I(\tau, \omega, \nu)$, encoding the presence of both the frequency and the chirpiness $\nu = d\omega/d\tau$.

Assume for the moment that the sound has a single time-varying frequency, e.g.,

$$s(\tau) = A \sin(\omega(\tau)\tau), \qquad A \in \mathbb{R}. \tag{2}$$

If the frequency is varying slowly enough and the window of the STFT is large enough, its sound image (up to the choice of normalisations constants in the Fourier transform) coincides roughly with

$$S(\tau, \omega) \sim \frac{A}{2i}\Big(\delta_0(\omega - \omega(\tau)) - \delta_0(\omega + \omega(\tau))\Big),$$

where $\delta_0$ is the Dirac delta distribution centered at 0. That is, $S$ is concentrated on the two curves $\tau \mapsto (\tau, \omega(\tau))$ and $\tau \mapsto (\tau, -\omega(\tau))$, see Figure 3. Let us focus only on the first curve.

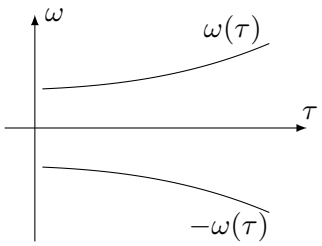



Figure 3: Short-time Fourier transform of the signal in (2), for a positive and increasing $\omega(\cdot)$.

Because of the sensitivity to variations of the input, as discussed in Section 1, the curve $\omega(\tau)$ is lifted in a bigger space by adding a new variable $\nu = d\omega/d\tau$. In mathematical terms the 3-dimensional space $(\tau, \omega, \nu)$ is called the *augmented space*. It will be the basis for the geometric model of A1 that we are going to present.

Up to now the curve $\omega(\tau)$ was parameterized by one of the coordinates of the contact space (the variable $\tau$), but it will be more convenient to consider it as a parametric curve in the space $(\tau, \omega)$. More precisely, the original curve $\omega(\tau)$ is represented in the space $(\tau, \omega)$ as $t \mapsto (t, \omega(t))$ (thus imposing $\tau = t$). Similarly, the lifted curve is parameterized as $t \mapsto (t, \omega(t), \nu(t))$. To every regular enough curve $t \mapsto (t, \omega(t))$, one can associate a lift $t \mapsto (t, \omega(t), \nu(t))$ in the contact space simply by computing $\nu(t) = d\omega/dt$. Vice-versa, a regular enough curve in the contact space $t \mapsto (\tau(t), \omega(t), \nu(t))$ is a lift of planar curve $t \mapsto (t, \omega(t))$ if $\tau(t) = t$ and if $\nu(t) = d\omega/dt$. Now, defining $u(t) = d\nu/dt$ we can say that a curve in the contact space $t \mapsto (\tau(t), \omega(t), \nu(t))$ is a lift of a planar curve if there exists a function $u(t)$ such that:

$$\frac{d}{dt}\begin{pmatrix}\tau\\ \omega\\ \nu\end{pmatrix} = \begin{pmatrix}1\\ \nu\\ 0\end{pmatrix} + u(t)\begin{pmatrix}0\\ 0\\ 1\end{pmatrix}. \tag{3}$$

Letting $q = (\tau, \omega, \nu)$, equation (3) can be equivalently written as the control system

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

where the $X_0$ and $X_1$ are the two vector fields in $\mathbb{R}^3$

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

Notice that the two vector fields appearing in this formula generate the Heisenberg group. However, we are not dealing here with a sub-Riemannian structure, since the space $\{X_0 + uX_1 \mid u \in \mathbb{R}\}$ is a line and not a plane. (One would get a plane by considering two controls, namely $\{u_0X_0 + u_1X_1 \mid (u_0, u_1) \in \mathbb{R}^2\}$.)

Following [9], when $s$ is a general sound signal, we lift each level line of $|S|$. By the implicit function theorem, this yields the following subset of the contact space:

$$\Sigma = \left\{(\tau, \omega, \nu) \in \mathbb{R}^3 \mid \nu\partial_\omega|S|(\tau,\omega) + \partial_\tau|S|(\tau,\omega) = 0\right\}. \tag{4}$$

If $|S| \in C^2$ and $\operatorname{Hess}|S|$ is non-degenerate, the set $\Sigma$ is indeed a surface. Finally, the external input from the cochlea is given by

$$I(\tau, \omega, \nu) = S(\tau, \omega)\delta_\Sigma(\tau, \omega, \nu). \tag{5}$$

Here, $\delta_\Sigma$ denotes the Dirac delta distribution concentrated at $\Sigma$. The presence of this distributional term is necessary for a well-defined solution to the evolution equation (WC). Such en equation is introduced in the next section.

## 2.2. Cortical activations in A1

On the basis of what described in the previous section and the well-known tonotopical organization of A1 (cf. Section 1), we propose to consider A1 to be the space of $(\omega, \nu) \in \mathbb{R}^2$. When hearing a sound $s(\cdot)$, the external input fed to A1 at time $t > 0$ is then given as the slice at $\tau = t$ of the lift $I$ of $s$ to the contact space. That is, hearing an "instantaneous sound level" $s(t)$ reflects in the external input $I(t, \omega, \nu)$ to the "neuron" $(\omega, \nu)$ in A1 as follows: The "neuron" receives an external charge $S(t, \omega)$ if $(t, \omega, \nu) \in \Sigma$, and no charge otherwise, where $\Sigma$ is defined in (4).

We model the neuronal activation induced by the external stimulus $I$ by adapting to this setting the well-known Wilson-Cowan equations. These equations are widely used and proved to be very effective in the study of V1 [43, 12]. According to this framework, the resulting activation $a : [0, T] \times \mathbb{R} \times \mathbb{R} \to \mathbb{C}$ is the solution of the following equation with delay $\delta > 0$:

$$\partial_t a(t, \omega, \nu) = -\alpha a(t, \omega, \nu) + \beta I(t, \omega, \nu) + \gamma \int_{\mathbb{R}^2} k_\delta(\omega, \nu \| \omega', \nu')\sigma(a(t - \delta, \omega', \nu'))\, d\omega'\, d\nu', \tag{6}$$

with initial condition $a(t, \cdot, \cdot) \equiv 0$ for $t \leq 0$. Here, $\alpha, \beta, \gamma > 0$ are parameters, $k_\delta$ is an interaction kernel, and $\sigma : \mathbb{C} \to \mathbb{C}$ is a (non-linear) saturation function, or sigmoid. In the following, we let $\sigma(\rho e^{i\theta}) = \tilde{\sigma}(\rho)e^{i\theta}$ where $\tilde{\sigma}(x) = \min\{1, \max\{0, \kappa x\}\}$, $x \in \mathbb{R}$, for some fixed $\kappa > 0$. The fact that the non-linearity $\sigma$ does not act on the phase is one of the key ingredients in proving that this processing preserves the natural symmetries of sound signals, see Proposition 4 in Appendix B.

When $\gamma = 0$, equation (6) becomes the standard low-pass filter $\partial_t a = -\alpha a + I$, whose solution is the convolution of the input signal $I$ with the function

$$\varphi(t) = \begin{cases} e^{-t\alpha} & \text{if } t > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Setting $\gamma \neq 0$ adds a non-linear delayed interaction term on top of this exponential smoothing, encoding the inhibitory and excitatory interconnections between neurons. Next section is devoted to the choice of the integral kernel $k_\delta$.

**Remark 3.** *In* (6) *we chose to consider a simple form for the interaction term. A more precise choice would indeed need to take into account the whole history of the process, for example by considering*

$$\int_{\tau}^{+\infty} e^{-\varrho(s-\tau)} \int_{\mathbb{R}^2} k_s(\omega,\nu\|\omega',\nu')\sigma(a(t-s,\omega',\nu'))\,d\omega'\,d\nu'\,ds, \qquad \varrho>0.$$

## 2.3. The neuronal interaction kernel

Considering A1 as a slice of the augmented space allows to deduce a natural structure for neuron connections as follows. Going back to a sound composed by a single time-varying frequency $t \mapsto \omega(t)$, we have that its lift is concentrated on the curve $t \mapsto (\omega(t), \nu(t))$, such that

$$\frac{d}{dt}\begin{pmatrix}\omega\\ \nu\end{pmatrix} = Y_0(\omega,\nu) + u(t)Y_1(\omega,\nu), \tag{7}$$

where $Y_0(\omega,\nu) = (\nu,0)^\top$, $Y_1(\omega,\nu) = (0,1)^\top$, and $u : [0,T] \to \mathbb{R}$.

As in the case of V1 [8], we model neuronal connections via these dynamics. In practice, this amounts to assume that the excitation starting at a neuron $X_0 = (\omega',\nu')$ evolves as the stochastic process $\{A_t\}_{t\geq 0}$ naturally associated with (7). This is given by the following stochastic differential equation,

$$dA_t = Y_0(A_t)dt + Y_1(A_t)dW_T, \qquad A_0 = (\omega',\nu'), \tag{8}$$

where $\{W_t\}_{t\geq 0}$ is a Wiener process. The generator of $\{A_t\}_{t\geq 0}$ is the second order differential operator

$$\mathcal{L} = Y_0 + (Y_1)^2 = \nu\partial_\omega + b\partial_\nu^2.$$

In this formula, the vector fields $Y_0$ and $Y_1$ are interpreted as first-order differential operators. Moreover, we added a scaling parameter $b>0$, modelling the relative strength of the two terms.

It is natural to model the influence $k_\delta(\omega,\nu\|\omega',\nu')$ of neuron $(\omega',\nu')$ on neuron $(\omega,\nu)$ at time $\delta>0$ as the transition density of the process $\{A_t\}_{t\geq 0}$. It is well-known that such transition density is obtained by computing the integral kernel at time $\delta$ of the Fokker-Planck equation corresponding to (8), that reads

$$\partial_t I = \mathcal{L}^* I, \quad \text{where} \quad \mathcal{L}^* = -Y_0 + (Y_1)^2 = -\nu\partial_\omega + b\partial_\nu^2. \tag{9}$$

The existence of an integral kernel for (9) is a consequence of the hypoellipticity[4] of $(\partial_t - \mathcal{L}^*)$. The explicit expression of $k_\delta$ is well-known and we recall it in the following result, proved in Appendix A.

**Proposition 1.** *The integral kernel of equation* (9) *is*

$$k_\delta(\omega,\nu\|\omega',\nu') = \frac{\sqrt{3}}{2\pi b\delta^2}\exp\left(-\frac{g_\delta(\omega,\nu\|\omega',\nu')}{b\delta^3}\right), \tag{10}$$

[4]That is, if $f$ is a distribution defined on an open set $\Omega$ and such that $(\partial_t - \mathcal{L}^*)f \in C^\infty(\Omega)$, then $f \in C^\infty(\Omega)$.

*where,*

$$g_\delta(\omega,\nu\|\omega',\nu') = 3(\omega-\omega')^2 - 3\delta(\omega-\omega')(\nu+\nu') + \delta^2\left(\nu^2+\nu\nu'+\nu'^2\right).$$

# 3. Numerical implementation

For the numerical experiments, we chose to implement the proposed algorithm in Julia [7]. As already presented, this process consists in a preprocessing phase, in which we build an input function $I$ on the 3D contact space, a main part, where $I$ is used as the input of the Wilson-Cowan equation (WC), and a post-processing phase, where the reconstructed sound is recovered from the result of the first part.

In the following, we present these phases separately.

## 3.1. Pre-processing

The input sound $s$ is lifted to a time-frequency representation $S$ via a classical implementation of STFT, i.e., by performing FFTs of a windowed discretised input. In the proposed implementation we chose to use a standard Hann window (see, e.g., [36]):

$$W(x) = \begin{cases} \dfrac{1+\cos(2\pi x/L)}{2} & \text{if } |x| < L/2, \\ 0 & \text{otherwise.} \end{cases}$$

The resulting time-frequency signal is then lifted to the contact space through an approximate computation of the gradient $\nabla|S|$ and the following discretisation of (5):

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

**Discretisation issues** While the discretisation of the time and frequency domains is a well understood problem, dealing with the additional chirpiness variable requires some care. Indeed, even if we assume that the significant frequencies of the input sound $s$ belong to a bounded interval $\Lambda \subset \mathbb{R}$, in general the set $\{\nu \in \mathbb{R} \mid I(\tau,\omega,\nu) \neq 0\}$ is unbounded. Indeed, one can check that as $(\tau,\omega)$ moves to a point where the countour lines of $|S|$ become vertical, the set of chirpinesses $\nu$'s such that $\nu\partial_\omega|S|(\tau,\omega) = -\partial_\tau|S|(\tau,\omega)$ will converge to $\pm\infty$.

In the numerical implementation we chose to restrict the admissible chirpinesses to a bounded interval $N \subset \mathbb{R}$. This set is chosen in a case by case fashion in order to contain the relevant slopes for the examples under consideration. Work is ongoing to automate this procedure.

## 3.2. Processing

Equation (WC) can be solved via a standard forward Euler method. Hence, the delicate part of the numerical implementation is the computation of the interaction term.

As is clear from the explicit expression given in Proposition 1, $k_\delta$ is not a convolution kernel. That is, $k_\delta(\omega,\nu\|\omega',\nu')$ cannot be expressed as a function of $(\omega-\omega',\nu-\nu')$. As a consequence, a priori we need to explicitly compute all values $k_\delta(\omega,\nu\|\omega',\nu')$ for $(\omega,\nu)$ and $(\omega',\nu')$ in the considered domain. As is customary, in order to reduce computation times, we fix a threshold $\varepsilon>0$ and for any given $(\omega,\nu)$ we compute only values for $(\omega',\nu')$ in the compact set

$$\mathrm{K}_\delta^\varepsilon(\omega,\nu)=\{(\omega',\nu')\mid k_\delta(\omega,\nu\|\omega',\nu')\geq\varepsilon\}.$$

The structure of $\mathrm{K}_\delta^\varepsilon(\omega,\nu)$ is given in the following, whose proof we defer to Appendix A.

**Proposition 2.** *For any $\varepsilon>0$ and $(\omega,\nu)\in\mathbb{R}^2$, we have that $\mathrm{K}_\delta^\varepsilon(\omega,\nu)$ is the set of those $(\omega',\nu')\in\mathbb{R}^2$ that satisfy*

$$|\nu-\nu'|^2\leq C_\varepsilon:=-4b\delta\log\left(\frac{2\pi b\tau^2}{\sqrt{3}}\varepsilon\right),$$
$$\left|\omega'-\omega+\frac{\delta(\nu+\nu')}{2}\right|\leq\frac{\delta}{2\sqrt{3}}\sqrt{C_\varepsilon-|\nu-\nu'|^2}.$$

**Remark 4.** *It holds $C_\varepsilon\geq 0$ if and only if*

$$\varepsilon\leq\frac{\sqrt{3}}{2\pi b\delta^2}.$$

*Indeed, for any $(\omega,\nu)\in\mathbb{R}^2$, the r.h.s. above corresponds to $\max k_\delta(\omega,\nu\|\cdot,\cdot)$, and thus $\mathrm{K}^\varepsilon(\omega,\nu)=\varnothing$ for larger values of $\varepsilon$.*

The above allows to numerically implement $k_\delta$ as a family of sparse arrays. That is, let $G\subset\Lambda\times N$ be the chosen discretisation of the significant set of frequencies and chirpinesses. Then, to $\xi=(\omega,\nu)\in G$ we associate the array $M_\xi:G\to\mathbb{R}$ defined by

$$M_\xi(\xi')=\begin{cases}k_\delta(\xi\|\xi') & \text{if } \xi'\in\mathrm{K}^\varepsilon(\xi),\\ 0 & \text{otherwise.}\end{cases}$$

Therefore, up to choosing the tolerance $\varepsilon\ll 1$ sufficiently small, the interaction term in (WC), evaluated at $\xi=(\omega,\nu)\in G$, can be efficiently estimated by

$$\int_{\mathbb{R}^2} w(\xi\|\xi')\sigma(a(t-\delta,\xi'))\,d\xi'\approx\sum_{\xi'\in K^\varepsilon(\xi)}M_\xi(\xi')a(t-\delta,\xi').$$

### 3.3. Post-processing

Both operations in the pre-processing phase are inversible: the STFT by inverse STFT, and the lift by integration along the $\nu$ variable (that is, summation of the discretized solution). The final output signal is thus obtained by applying the inverse of the pre-processing (integration then inverse STFT) to the solution $a$ of (WC). That is, the resulting signal is given by

$$\hat{s}(t)=\mathrm{STFT}^{-1}\left(\int_{-\infty}^{+\infty}a(t,\omega,\nu)\,\mathrm{d}\,\nu\right).$$

The following guarantees that $\hat{s}$ is real-valued and thus correctly represents a sound signal. From the numerical point of view, this implies that we can focus on solutions of (WC) in the half-space $\{\omega \geq 0\}$, which can then be extended to the whole space by mirror symmetry.

**Proposition 3.** *It holds that $\hat{s}(t) \in \mathbb{R}$ for all $t > 0$.*

*Proof.* Let us denote

$$\hat{S}(t, \omega) = \int_{-\infty}^{+\infty} a(t, \omega, \nu) \, \mathrm{d}\, \nu,$$

so that $\hat{s} = \mathrm{STFT}^{-1}(\hat{S})$. Moreover, for any function $f(t, \omega, \nu)$, we let $f^\star(t, \omega, \nu) := \bar{f}(t, -\omega, -\nu)$.

To prove the statement it is enough to show that

$$a(t, \cdot, \cdot) \equiv a^\star(t, \cdot, \cdot) \qquad \forall t \geq 0. \tag{11}$$

This is trivially satisfied for $t \leq 0$, since in this case $a(t, \cdot, \cdot) \equiv 0$.

We now claim that if (11) holds on $[0, T]$ it holds on $[0, T + \delta]$, which will prove it for all $t \geq 0$. By definition of $I$ and the fact that $S(t, -\omega) = \overline{S(t, \omega)}$, we immediately have $I \equiv I^\star$. On the other hand, the explicit expression of $k_\delta$ in (10) yields that

$$k_\delta(-\omega, -\nu \| \omega', \nu') = k_\tau(\omega, \nu \| - \omega', -\nu').$$

Then, for all $t \leq T + \delta$, we have

$$\begin{aligned}
&\int_{\mathbb{R}^2} k_\delta(-\omega, -\nu \| \omega', \nu') \sigma(a(t - \tau, \omega', \nu')) \, d\omega' \, d\nu' \\
&\quad = \int_{\mathbb{R}^2} k_\delta(\omega, \nu \| \omega'', \nu'') \sigma(a^\star(t - \tau, \omega'', \nu'')) \, d\omega'' \, d\nu'' \\
&\quad = \int_{\mathbb{R}^2} k_\delta(\omega, \nu \| \omega'', \nu'') \sigma(a(t - \tau, \omega'', \nu'')) \, d\omega'' \, d\nu''.
\end{aligned}$$

A simple argument, e.g., using the variation of constants method, shows that these two facts imply the claim, and thus the statement. □

## 4. Experiments

In Figure 4 we present a series of experiments on simple synthetic sounds in order to exhibit some key features of our algorithm. These experiments can be reproduced via the code available at `https://www.github.com/dprn/WCA1`.

The first example, Figure 4a, is a simple linear chirp, such that the dominating frequency depends linearly on time (i.e., corresponding to $\omega(t) = \mu t$ for some $\mu \in \mathbb{R}$). One observes that the processed sound presents the same feature but for a longer duration. The parameters in the experiment have been chosen to emphasize the effect of the modelling equation: the reconstruction should not present a tail that is as pronounced, however this allows to highlight the diffusive effect along the lifted slope.

The second example, Figure 4b, corresponds to the same linear chirp as Figure 4a, that has been interrupted in its middle section, creating two disjoint linear chirps. Thanks to the transport effect of the algorithm, the gap between the two chirps is bridged in the processed signal. For this illustration, the interruption lasts about twice as long as the delay.

The third example, Figure 4c, consists of the sum of two linear chirps with different slopes. The slopes have been picked to suggest that linear continuations of the chirps should intersect. This is indeed what happens in the processed signal. However, notice that the resulting crossing happens almost as a sum of the two chirps processed independently, with close to no interaction at the crossing. This is purely an effect of the lift procedure. The increasing chirp is (predominantly) lifted to a stratum corresponding to a positive slope, while the decreasing chirp is lifted to a negative slope stratum. De facto, their evolution under the Wilson–Cowan equation is decoupled in the 3D augmented space.

The fourth and last example, Figure 4d, corresponds to a nonlinear chirp, roughly corresponding to choosing $\omega(\tau) = \sin(m\tau)$ in (2). The construction of the model favors linearity in the evolution of perceived frequencies. We can observe how the more linear elements of the input result in more diffusion.

## 5. Conclusion

In this work we presented a sound reconstruction framework inspired by the analogies between visual and auditory cortices. Building upon the successful cortical inspired image reconstruction algorithms, the proposed framework lifts time-frequencies representations of signals to the 3D contact space, by adding instantaneous chirpiness information. These redundant representations are then processed via adapted diffeo-integral Wilson-Cowan equations.

The promising results obtained on simple synthetic sounds, although preliminary, suggest possible applications of this framework to the problem of degraded speech. The next step will be to test the reconstruction ability of normal-hearing humans on originally degraded speech material compared to the same speech material after algorithm reconstruction. Such an endeavour will contribute to the understanding of the auditory mechanisms emerging in adverse listening conditions. It will furthermore help to deepen our knowledge on general organization principles underlying the functioning of the human auditory cortex.

## Acknowledgments

The first three authors have been supported by the ANR project SRGI ANR-15-CE40-0018 and by the ANR project Quaco ANR-17-CE40-0007-01. This study was also supported by the IdEx Universite de Paris, ANR-18-IDEX-0001, awarded to the last author.

## A. Integral kernel of the Kolmogorov operator

The result in Proposition 1 is well-known. E.g., by applying [3, Proposition 9] and letting $x = (\omega, \nu)$ and $x' = (\omega', \nu')$ one gets that the kernel is

$$k_\delta(\omega, \nu \| \omega', \nu') = \frac{1}{2\pi\sqrt{\det D_\delta}} \exp\left[-\frac{1}{2}\left(x' - \mathrm{e}^{\delta A}\, x\right)^\top D_\delta^{-1}\left(x' - \mathrm{e}^{\delta A}\, x\right)\right],$$

where

$$A = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} 0 \\ \sqrt{2b} \end{pmatrix},$$

and

$$D_\delta = \mathrm{e}^{\delta A}\left[\int_0^\tau \mathrm{e}^{-\sigma A}\, BB^*\, \mathrm{e}^{-\sigma A^*}\, \mathrm{d}\,\sigma\right] \mathrm{e}^{\tau A^*}.$$

Direct computations yield

$$D_\delta = 2b\begin{pmatrix} \delta^3/3 & -\tau^2/2 \\ -\delta^2/2 & \tau \end{pmatrix}, \quad \text{and} \quad \det D_\delta = \frac{b^2\tau^4}{3}.$$

Therefore,

$$\frac{1}{2}D_\delta^{-1} = \frac{1}{b\tau^3}M, \quad \text{where} \quad M = \begin{pmatrix} 3 & 3\delta/2 \\ 3\delta/2 & \delta^2 \end{pmatrix}.$$

Finally, the statement follows by letting

$$g_\delta(x \| x') = \left(x' - \mathrm{e}^{\delta A}\, x\right)^\top M \left(x' - \mathrm{e}^{\delta A}\, x\right).$$

We now turn to an argument for Proposition 2. Observe that $k_\delta(x\|x') \geq \varepsilon$ if and only if

$$g_\delta(x\|x') \leq \eta := -b\delta^3 \log\left(\frac{2\pi b\delta^2}{\sqrt{3}}\varepsilon\right).$$

Then, we start by solving $z^\top M z \leq \eta$, for $z \in \mathbb{R}^2$. One can check that this is verified if and only if

$$|z_2| \leq \sqrt{\frac{4\eta}{\delta^2}}, \quad \text{and} \quad \left|z_1 + \frac{\delta}{2}z_2\right| \leq \frac{\delta}{2\sqrt{3}}\sqrt{\frac{4\eta}{\delta^2} - z_2^2}.$$

Since $C_\varepsilon = 4\eta/\delta^2$ the statement follows by computing the above at $z = x' - e^{\tau A}x$.

## B. Heisenberg group action on the contact space

Recall that the short-time Fourier transform of a signal $s \in L^2(\mathbb{R})$ is given by

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

Here, $W : \mathbb{R} \to [0,1]$ is a compactly supported (smooth) window, so that $S \in L^2(\mathbb{R}^2)$. Fundamental operators in time frequency analysis [21] are time and phase shifts, acting on signals $s \in L^2(\mathbb{R})$ by

$$T_\theta s(t) := s(t-\theta) \quad \text{and} \quad M_\lambda s(t) := e^{2\pi i \lambda t} s(t),$$

for $\theta, \lambda \in \mathbb{R}$. One easily checks that $T_\theta$ and $M_\lambda$ are unitary operators on $L^2(\mathbb{R})$. By conjugation with the short-time Fourier transform, they naturally define the unitary operators on $L^2(\mathbb{R}^2)$ given by

$$\begin{aligned} T_\theta S(\tau,\omega) &= e^{-2\pi i \omega \theta} S(\tau-\theta,\omega), \\ M_\lambda S(\tau,\omega) &= S(\tau,\omega-\lambda). \end{aligned} \tag{12}$$

The relevance of the Heisenberg group in time-frequency analysis is a direct consequence of the commutation relation

$$T_\theta M_\lambda T_\theta^{-1} M_\lambda^{-1} = e^{-2\pi i \lambda \theta} \,\mathrm{Id}\,.$$

Indeed, this shows that the operator algebra generated by $(T_\theta)_{\theta\in\mathbb{R}}$ and $(M_\lambda)_{\lambda\in\mathbb{R}}$ coincides with the Heisenberg group $\mathbb{H}^1$ via the representation $U : \mathbb{H}^1 \to \mathcal{U}(L^2(\mathbb{R}^2))$ defined by

$$U(\theta,\lambda,\zeta) = e^{-2\pi i \zeta} T_\theta M_\lambda. \tag{13}$$

The above discussion shows that the Heisenberg group can be regarded as the natural space of symmetries of sound signals. In particular, any meaningful treatment of these signals should respect such symmetry. In the case of our model, this is the content of the following result.

**Proposition 4.** *The sound processing algorithm presented in this paper commutes with the Heisenberg group action* (13) *on sound signals. That is, if the input sound signal $s \in L^2(\mathbb{R}^2)$ yields $\hat{s}$ as a result, then, for any $(\theta,\lambda,\zeta) \in \mathbb{H}^1$, the input $U(\theta,\lambda,\zeta)s$ yields $U(\theta,\lambda,\zeta)\hat{s}$ as a result.*

*Proof.* We can schematically write the algorithm as:

$$\hat{s} = \mathrm{STFT}^{-1} \circ \mathrm{Proj} \circ \mathrm{WC} \circ \mathrm{Lift} \circ \mathrm{STFT}(s).$$

Here, Lift is the lift operator defined in Section 2.1, WC denotes the Wilson-Cowan evolution (WC), and Proj denotes the projection from the augmented space to the time-frequency representation, defined by

$$\mathrm{Proj}\, a(t,\omega) = \int_{-\infty}^{+\infty} a(t,\omega,\nu)\, d\nu.$$

Observe that (12) shows that $U$ induces a representation of $\mathbb{H}^1$ on $L^2(\mathbb{R}^2)$, the codomain of the STFT, which we will denote by $\tilde{U}$. Thus, to prove the statement it suffices to show that

$$[\mathrm{Proj} \circ \mathrm{WC} \circ \mathrm{Lift}, \tilde{U}(\theta,\lambda,\zeta)] = 0, \qquad \forall (\theta,\lambda,\zeta) \in \mathbb{H}^1.$$

Recall now that Lift associates with $S \in L^2(\mathbb{R}^2)$ a distribution of the form $\mathrm{Lift}[S](\tau,\omega,\nu) = S(\omega,\nu)\delta_\Sigma(\tau,\omega,\nu)$ for some $\Sigma \subset \mathbb{R}^3$. Due to the fact that $\Sigma$ is defined via the modulus of $S$, it is unaffected by the phase factors appearing in the representation $\tilde{U}$. That is, the lift of $\tilde{U}(\theta,\lambda,\zeta)S$ is given by

$$\begin{aligned}\mathrm{Lift}\left[\tilde{U}(\theta,\lambda,\zeta)S\right](\tau,\omega,\nu) &= e^{-2\pi i(\zeta+\omega\theta)}\delta_\Sigma(\tau-\theta,\omega-\lambda,\nu)S(\tau-\theta,\omega-\lambda)\\ &= e^{-2\pi i(\zeta+\omega\theta)}\,\mathrm{Lift}[S](\tau-\theta,\omega-\lambda,\nu).\end{aligned}$$

It is then immediate to check that $[\mathrm{Proj}\circ\mathrm{Lift},\tilde{U}(\theta,\lambda,\zeta)] = 0$.

We are left to verify that the operator WC commutes with $\tilde{U}(\theta,\lambda,\zeta)$. The commutation is trivial for the linear terms. On the other hand, the non-linearity introduced in the integral term commutes with $\tilde{U}(\theta,\lambda,\zeta)$ thanks to the fact that $\sigma(\rho e^{i\phi}) = e^{i\phi}\sigma(\rho)$ for all $\rho > 0, \phi \in \mathbb{R}$. □

## References


[1] A. Agrachev, D. Barilari, and U. Boscain. *A Comprehensive Introduction to Sub-Riemannian Geometry.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2019.

[2] P. Assmann and Q. Summerfield. *The Perception of Speech Under Adverse Conditions*, pages 231–308. Springer New York, New York, NY, 2004.

[3] D. Barilari and F. Boarotto. Kolmogorov-fokker-planck operators in dimension two: heat kernel and curvature, 2017.

[4] M. Bertalmío, L. Calatroni, V. Franceschi, B. Franceschiello, A. Gomez Villa, and D. Prandi. Visual illusions via neural dynamics: Wilson-cowan-type models and the efficient representation principle. *Journal of Neurophysiology*, 2020. PMID: 32159409.

[5] M. Bertalmío, L. Calatroni, V. Franceschi, B. Franceschiello, and D. Prandi. A cortical-inspired model for orientation-dependent contrast perception: A link with wilson-Cowan equations. In *Scale Space and Variational Methods in Computer Vision*, Cham, 2019. Springer International Publishing.

[6] M. Bertalmìo, L. Calatroni, V. Franceschi, B. Franceschiello, and D. Prand i. Cortical-inspired Wilson-Cowan-type equations for orientation-dependent contrast perception modelling. *J. Math. Imaging Vision*, 2020.

[7] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98, 2017.

[8] U. Boscain, R. Chertovskih, J.-P. Gauthier, and A. Remizov. Hypoelliptic diffusion and human vision: a semi-discrete new twist on the petitot theory. *SIAM J. Imaging Sci.*, 7(2):669–695, 2014.

[9] U. Boscain, J. Duplaix, J.-P. Gauthier, and F. Rossi. Anthropomorphic image reconstruction via hypoelliptic diffusion, 2010.

[10] U. V. Boscain, R. Chertovskih, J.-P. Gauthier, D. Prandi, and A. Remizov. Highly corrupted image inpainting through hypoelliptic diffusion. *J. Math. Imaging Vision*, 60(8):1231–1245, 2018.

[11] M. Bramanti. *An invitation to hypoelliptic operators and Hörmander's vector fields*. SpringerBriefs in Mathematics. Springer, Cham, 2014.

[12] P. C. Bressloff, J. D. Cowan, M. Golubitsky, P. J. Thomas, and M. C. Wiener. Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 356(1407):299–330, 2001.

[13] G. Citti and a. Sarti. A Cortical Based Model of Perceptual Completion in the Roto-Translation Space. *J. Math. Imaging Vis.*, 24(3):307–326, feb 2006.

[14] P. Dallos. *Overview: Cochlear Neurobiology*, pages 1–43. Springer New York, New York, NY, 1996.

[15] R. Duits and E. Franken. Left-invariant parabolic Evolutions on SE(2) and Contour Enhancement via Invertible Orientation Scores. Part I: Linear Left-invariant Diffusion Equations on SE. *Quarterly of Appl. Math., AMS*, 2010.

[16] R. Duits and E. Franken. Left-invariant parabolic evolutions on SE(2) and contour enhancement via invertible orientation scores. Part II: nonlinear left-invariant diffusions on invertible orientation scores. *Q. Appl. Math.*, (0):1–38, 2010.

[17] G. B. Ermentrout and J. D. Cowan. A mathematical theory of visual hallucination patterns. *Biological cybernetics*, 34:137–150, 1979.

[18] T. Fernandes, P. Ventura, and R. Kolinsky. Statistical information and coarticulation as cues to word boundaries: A matter of signal quality. *Perception & Psychophysics*, 69(6):856–864, Aug 2007.

[19] J. Fienup. Phase retrieval algorithms: a comparison. *Applied optics*, 21:2758–69, 08 1982.

[20] E. Franken and R. Duits. Crossing-Preserving Coherence-Enhancing Diffusion on Invertible Orientation Scores. *International Journal of Computer Vision*, 85(3):253–278, 2009.

[21] K. Gröchenig. *Foundations of time-frequency analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 2001.

[22] R. Hannemann, J. Obleser, and C. Eulitz. Top-down knowledge supports the retrieval of lexical information from degraded speech. *Brain Research*, 1153:134 – 143, 2007.

[23] G. Hickok and D. Poeppel. The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5):393–402, 2007.

[24] W. C. Hoffman. The visual cortex is a contact bundle. *Appl. Math. Comput.*, 32(2-3):137–167, Aug. 1989.

[25] D. H. Hubel and T. N. Wiesel. Receptive fields of single neurons in the cat's striate cortex. *The Journal of Physiology*, 148(3):574–591, 1959.

[26] A. Loebel, I. Nelken, and M. Tsodyks. Processing of Sounds by Population Spikes in a Model of Primary Auditory Cortex. *Frontiers in Neuroscience*, 1(1):197–209, 2007.

[27] P. A. Luce and C. T. McLennan. *Spoken Word Recognition: The Challenge of Variation*, chapter 24, pages 590–609. "John Wiley & Sons, Ltd", 2008.

[28] S. Mattys, M. Davis, A. Bradlow, and S. Scott. Speech recognition in adverse conditions: A review. *Language, Cognition and Neuroscience*, 27(7-8):953–978, 9 2012.

[29] R. Montgomery. *A tour of subriemannian geometries, their geodesics and applications*, volume 91 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2002.

[30] I. Nelken and M. B. Calford. *Processing Strategies in Auditory Cortex: Comparison with Other Sensory Modalities*, pages 643–656. Springer US, Boston, MA, 2011.

[31] G. Parikh and P. C. Loizou. The influence of noise on vowel and consonant cues. *The Journal of the Acoustical Society of America*, 118(6):3874–3888, 2005.

[32] J. Petitot and Y. Tondut. Vers une neurogéométrie. fibrations corticales, structures de contact et contours subjectifs modaux. *Mathématiques et Sciences humaines*, 145:5–101, 1999.

[33] J. Petitot and Y. Tondut. Vers une Neurogéométrie. Fibrations corticales, structures de contact et contours subjectifs modaux. pages 1–96, 1999.

[34] T. W. Polger, L. A. Shapiro, and O. U. Press. *The multiple realization book*. Oxford University Press, Oxford, 2016.

[35] D. Prandi and J.-P. Gauthier. *A semidiscrete version of the Citti-Petitot-Sarti model as a plausible model for anthropomorphic image reconstruction and pattern recognition*. SpringerBriefs in Mathematics. Springer International Publishing, Cham, 2017.

[36] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes*. Cambridge University Press, Cambridge, 3rd editio edition, 2007.

[37] J. Rankin, E. Sussman, and J. Rinzel. Neuromechanistic Model of Auditory Bistability. *PLoS Computational Biology*, 11(11):e1004555, nov 2015.

[38] J. P. Rauschecker. Auditory and visual cortex of primates: a comparison of two sensory systems. *The European journal of neuroscience*, 41(5):579–585, 03 2015.

[39] A. Sarti and G. Citti. The constitution of visual perceptual units in the functional architecture of v1. *Journal of Computational Neuroscience*, 38(2), Apr 2015.

[40] W. Sethares. *Tuning, Timbre, Spectrum, Scale*. Springer London, 2005.

[41] J. Sharma, A. Angelucci, and M. Sur. Induction of visual orientation modules in auditory cortex. *Nature*, 404(6780):841–847, 2000.

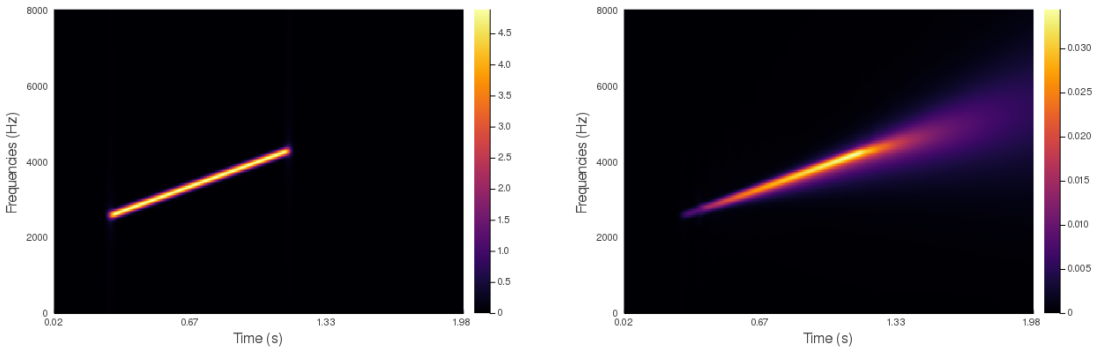
[42] B. Tian, P. Kuśmierek, and J. P. Rauschecker. Analogues of simple and complex cells in rhesus monkey auditory cortex. *Proceedings of the National Academy of Sciences*, 110(19):7892–7897, 2013.

[43] H. R. Wilson and J. D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*, 12(1):1–24, jan 1972.

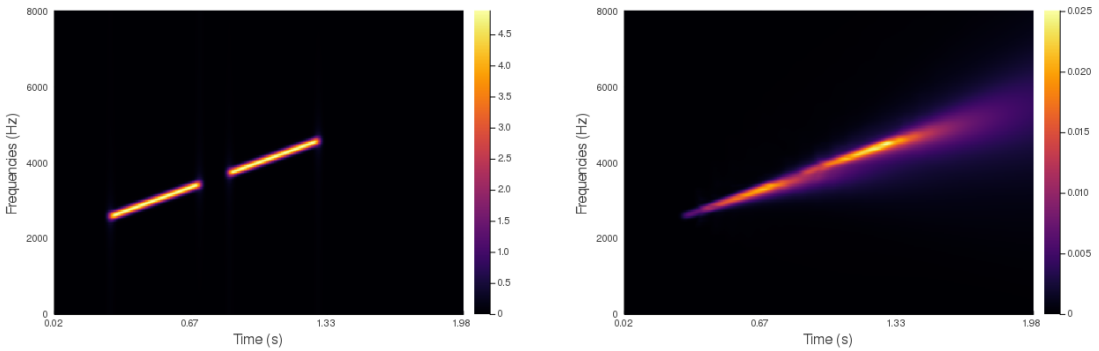
[44] R. J. Zatorre. Do you see what i'm saying? interactions between auditory and visual cortices in cochlear implant users. *Neuron*, 31(1):13 – 14, 2001.

[45] J. Zhang, B. Dashtbozorg, E. Bekkers, J. P. W. Pluim, R. Duits, and B. M. ter Haar Romeny. Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. *IEEE Transactions on Medical Imaging*, 35(12):2631–2644, Dec 2016.

[46] I. Zulfiqar, M. Moerel, and E. Formisano. Spectro-Temporal Processing in a Two-Stream Computational Model of Auditory Cortex. *Frontiers in Computational Neuroscience*, 13:95, jan 2020.

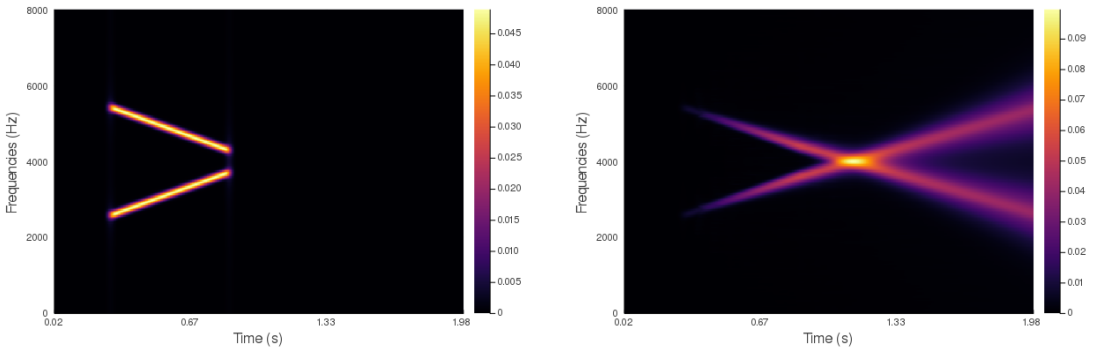



(a) Linear chirp. (Parameters: $\alpha = 55$, $\beta = 1$, $\gamma = 55$, $b = 0.05$.)

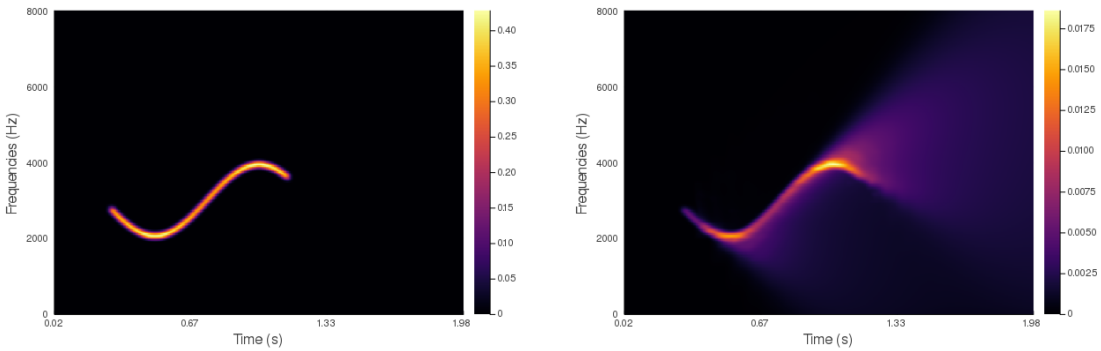



(b) Interrupted chirp. (Parameters: $\alpha = 55$, $\beta = 1$, $\gamma = 55$, $b = 0.05$.)

Figure 4: Experiments on synthetic sounds with varying frequencies. *Left:* The STFT of the original sound. *Right:* The STFT of the processed sound with delay $\delta = 0.0625s$. Parameters vary for each experiment depending on the desired effect we wish to highlight. Each time, only the positive frequencies are shown: negative frequencies are recovered via the Hermitian symmetry of the Fourier transform on real signals.




(c) Intersecting chirps. (Parameters: $\alpha = 53$, $\beta = 1$, $\gamma = 55$, $b = 0.01$.)




(d) Nonlinear chirp. (Parameters: $\alpha = 53$, $\beta = 1$, $\gamma = 55$, $b = 0.2$.)

Figure 4: Experiments on synthetic sounds with varying frequencies. (Continued.)