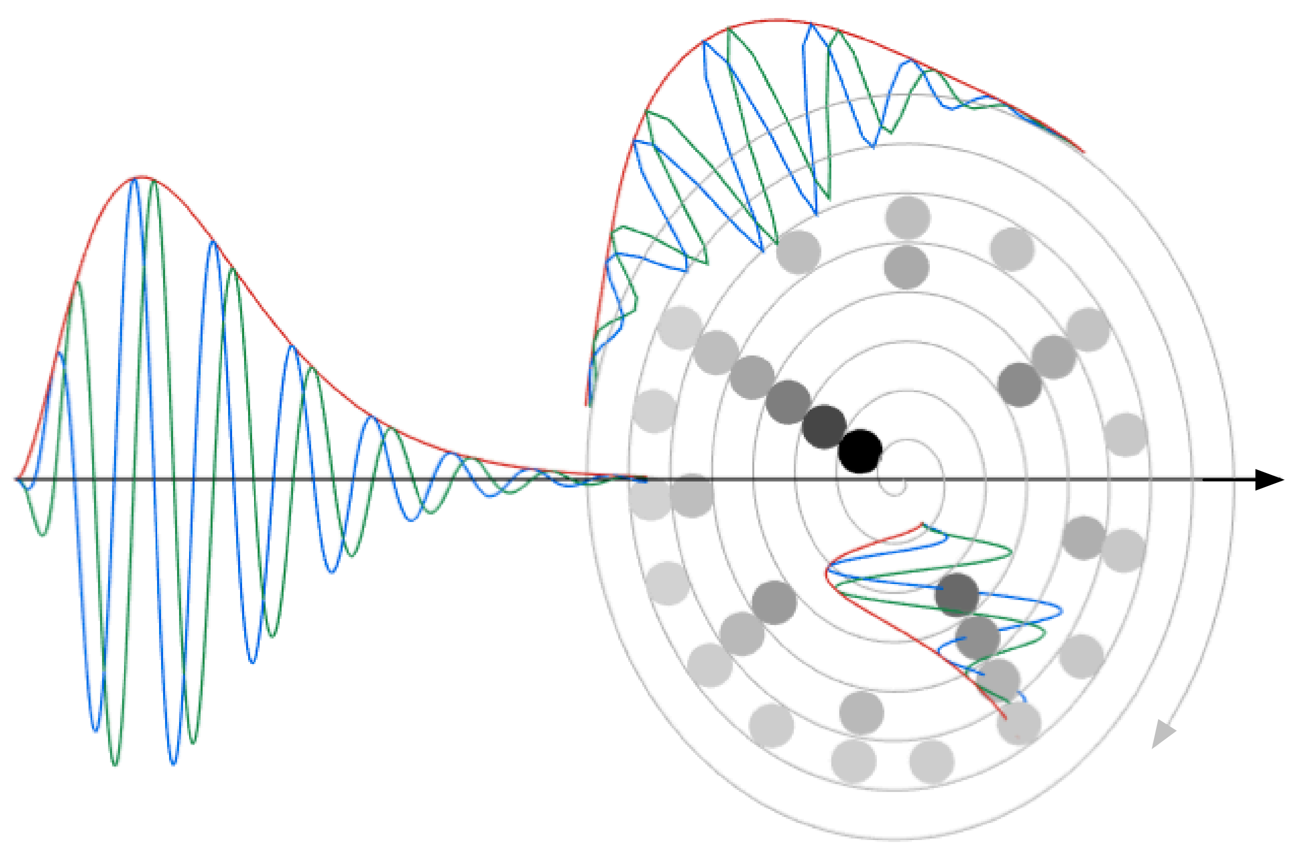
TRANSFORMÉE EN SCATTERING SUR LA SPIRALE TEMPS-CHROMA-OCTAVE



Vincent Lostanlen, Stéphane Mallat
École normale supérieure de Paris

2015-09-07, Lyon, FR. Ce travail est financé par la bourse ERC InvariantClass.

STRUCTURED STATIONARY PROCESSES

Music/speech signals are non-Gaussian, yet notably structured by:

(1) **intermittency**: rhythm, speech rate

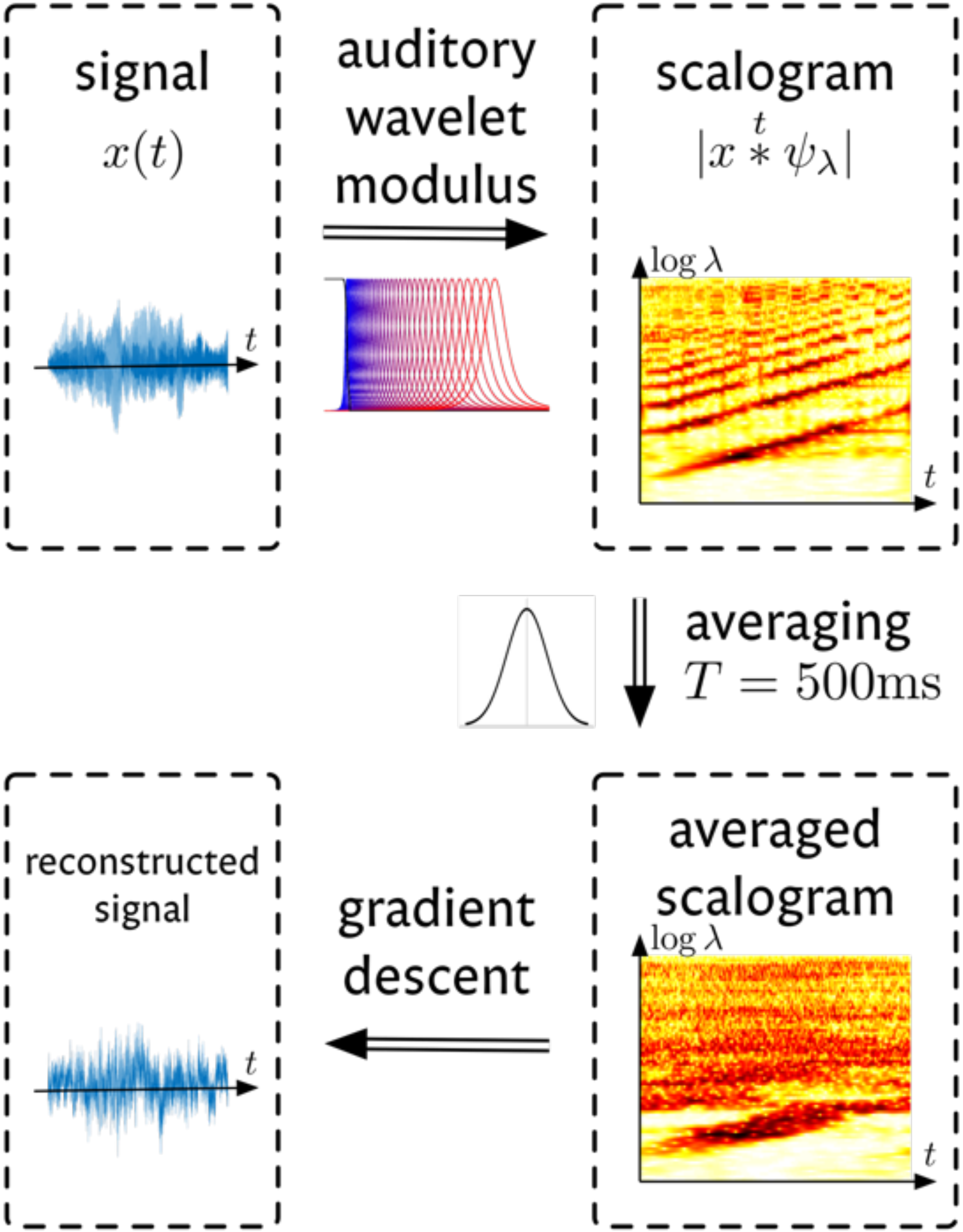
(2) **chirps**: melody, prosody.

(3) **harmonic combs**: timbre, formants.

Goal: capture (1), (2), and (3) without detection nor training.

Tool: nonlinear transformations of the wavelet scalogram.

Evaluation: signal reconstruction from averaged representation.

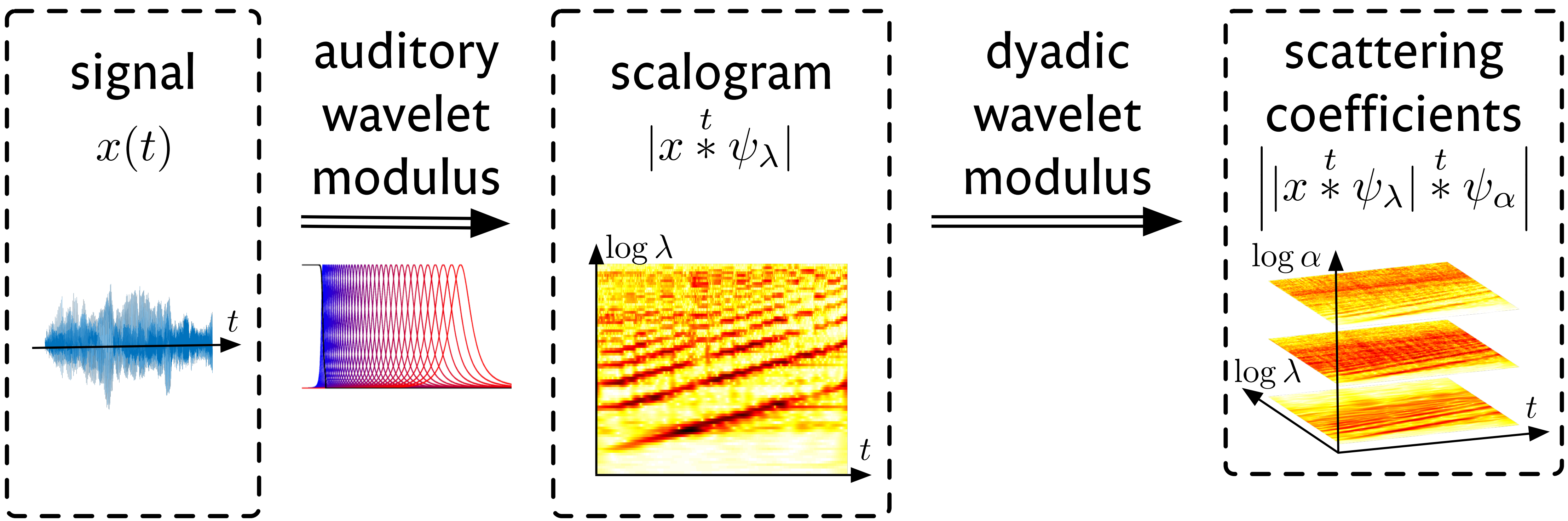


signal
$x(t)$
auditory
wavelet
modulus
scalogram
$|x \overset{t}{*} \psi_\lambda|$
$\log \lambda$
t
averaging
$T = 500\text{ms}$
averaged
scalogram
reconstructed
signal
gradient
descent

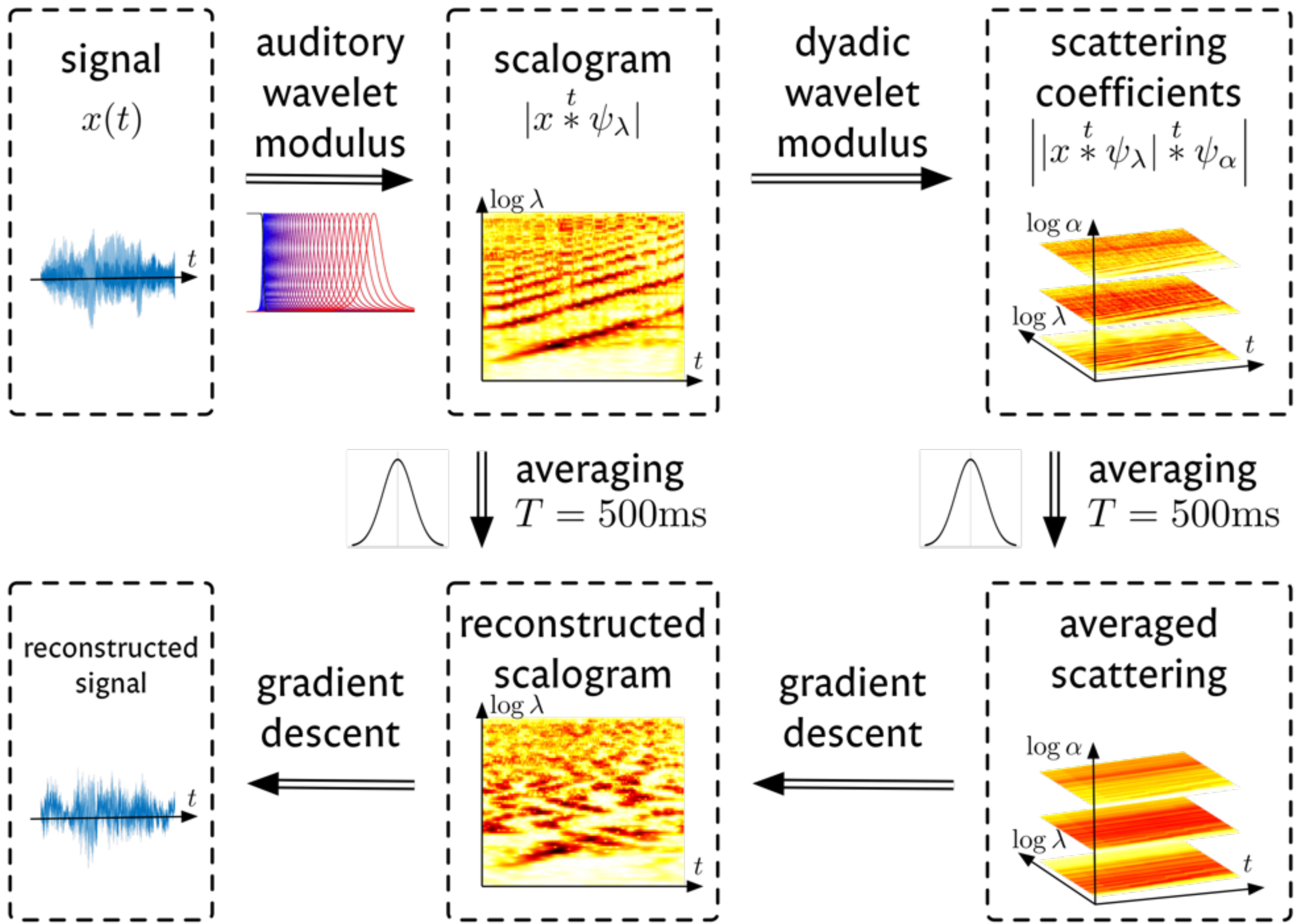
(1) Capturing intermittency

Non-regular information is lost after averaging.

Scatter each scalogram channel with wavelet modulus.



Andén and Mallat, IEEE TSP 2014. *Deep Scattering Spectrum*.

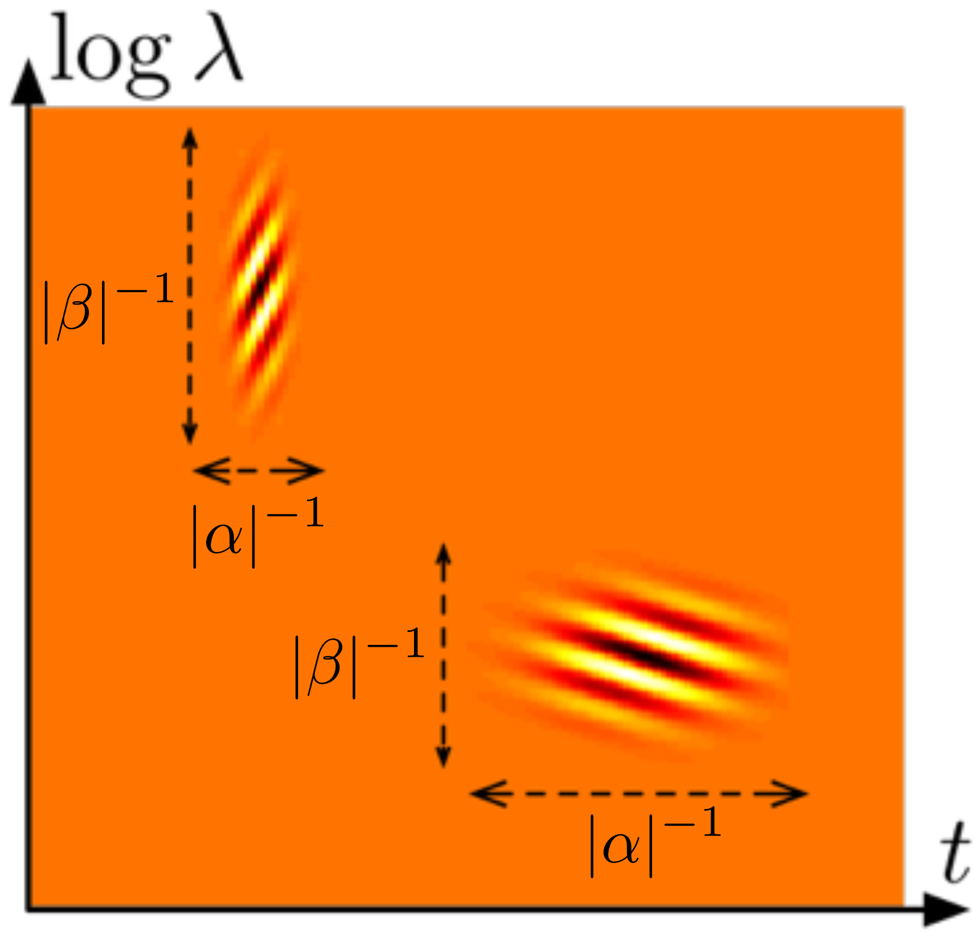


signal
$x(t)$
t
auditory wavelet modulus
scalogram
$|x \overset{t}{*} \psi_\lambda|$
$\log \lambda$
t
dyadic wavelet modulus
scattering coefficients
$\left| |x \overset{t}{*} \psi_\lambda| \overset{t}{*} \psi_\alpha \right|$
$\log \alpha$
$\log \lambda$
t
averaging
$T = 500\text{ms}$
averaging
$T = 500\text{ms}$
reconstructed signal
t
gradient descent
reconstructed scalogram
$\log \lambda$
t
gradient descent
averaged scattering
$\log \alpha$
$\log \lambda$
t

(2) Capturing chirps

« Plain » time scattering loses time-frequency coherence.

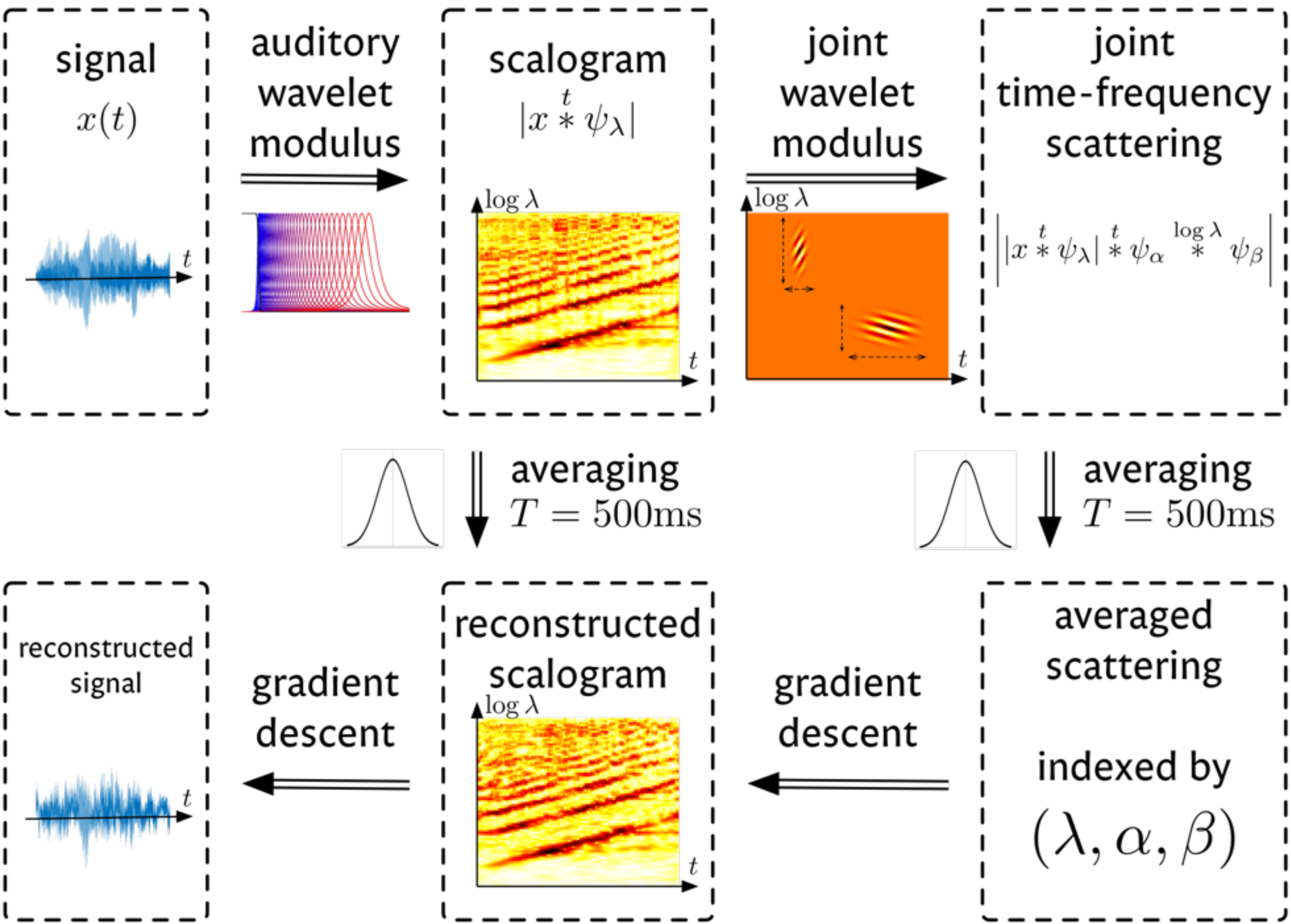
Scatter **jointly** in time and log-frequency.



$$\left| \, |x \overset{t}{*} \psi_\lambda| \overset{t}{*} \psi_\alpha \overset{\log \lambda}{*} \psi_\beta \right|$$

related to Shamma's STRF
(spectro-temporal response fields)
in auditory neuroscience.

Andén, Lostanlen, and Mallat, to be presented at IEEE MLSP 2015.
Joint time-frequency scattering for audio classification.



signal
$x(t)$
auditory wavelet modulus
scalogram
$|x \overset{t}{*} \psi_\lambda|$
joint wavelet modulus
joint time-frequency scattering
$\left||x \overset{t}{*} \psi_\lambda| \overset{t}{*} \psi_\alpha \overset{\log\lambda}{*} \psi_\beta\right|$
averaging
$T = 500\text{ms}$
averaging
$T = 500\text{ms}$
reconstructed signal
gradient descent
reconstructed scalogram
gradient descent
averaged scattering
indexed by
(λ, α, β)

(3) Capturing harmonicity

Harmonic combs are irregular along the log-frequency axis...

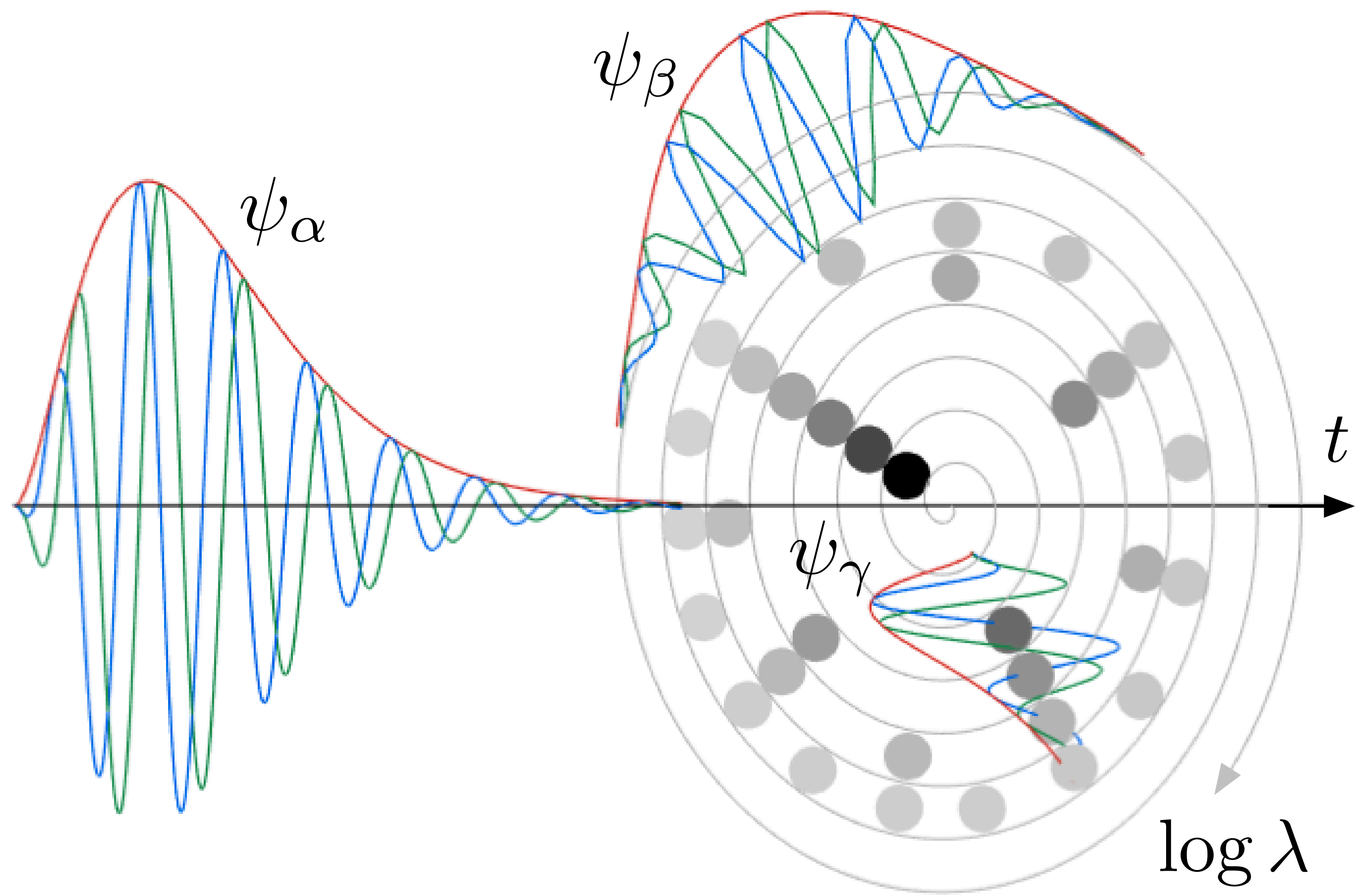
... but they create radial patterns in the **pitch spiral.**

See Shepard (1964), Risset (1969), Warren (2003), Deutsch (2008).

This **Snail Analyzer Tuner** beta demo is brought to you thanks to:
Ch. Picasso, Th. Hélie, H. Vinet, F. Rousseau ; Ircam and CNRS.
The software will soon be released commercially.

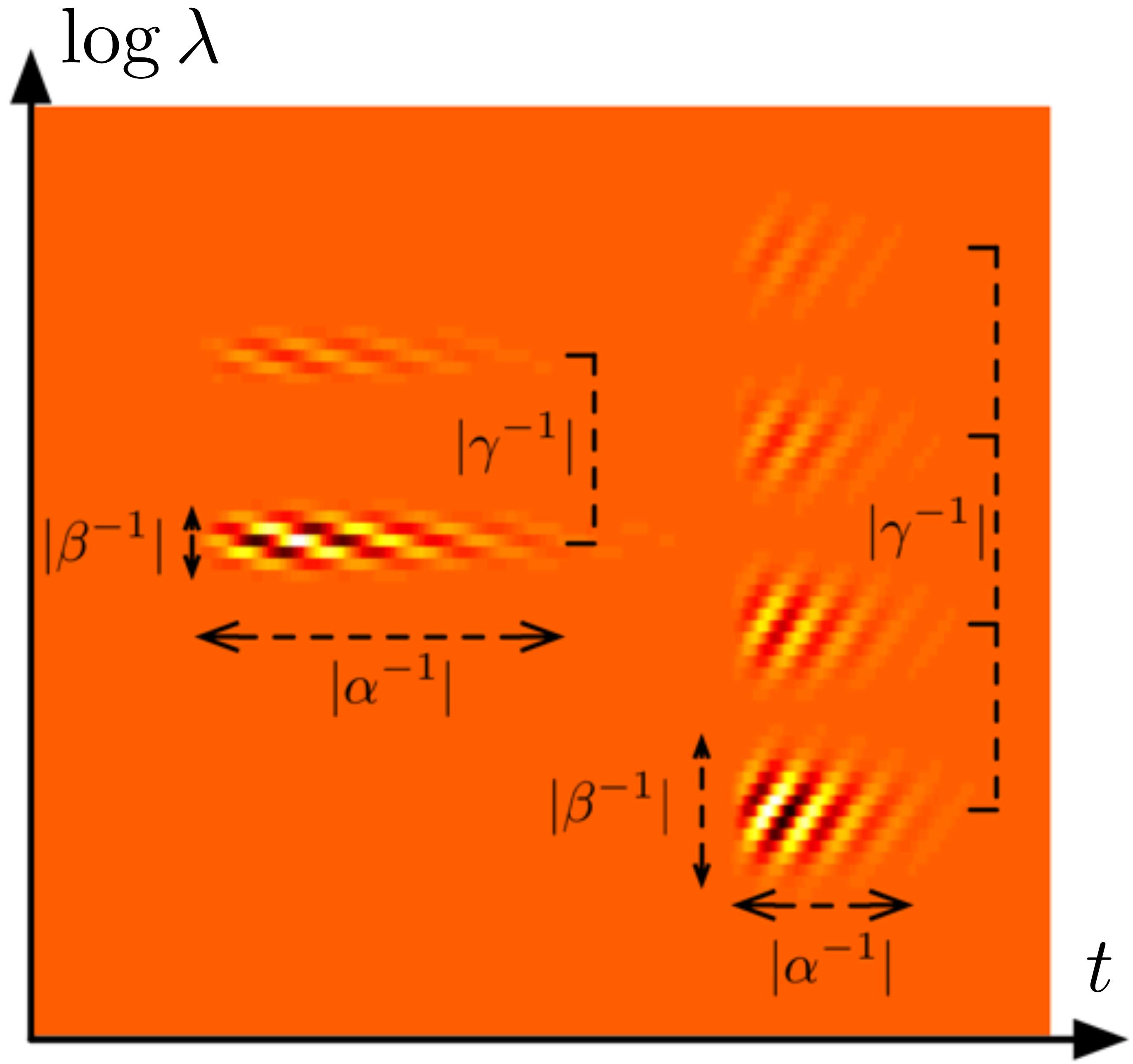
Spiral wavelets: 3D view

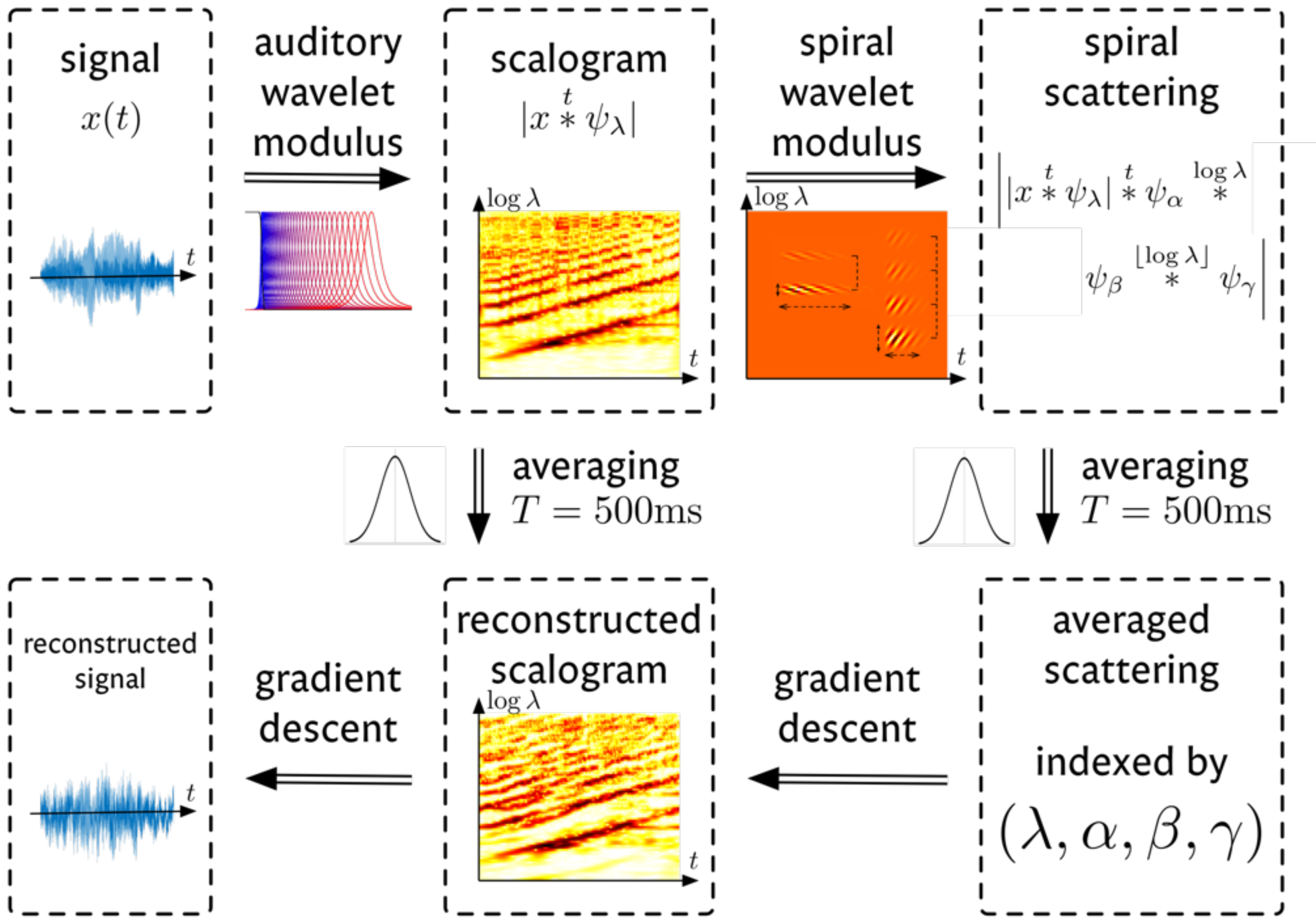
$$\left| |x \overset{t}{*} \psi_\lambda| \overset{t}{*} \psi_\alpha \overset{\log \lambda}{*} \psi_\beta \overset{\lfloor \log \lambda \rfloor}{*} \psi_\gamma \right|$$



Spiral wavelets: 2D view

$$\left| \left| x \overset{t}{*} \psi_\lambda \right| \overset{t}{*} \psi_\alpha \overset{\log \lambda}{*} \psi_\beta \overset{\lfloor \log \lambda \rfloor}{*} \psi_\gamma \right|$$



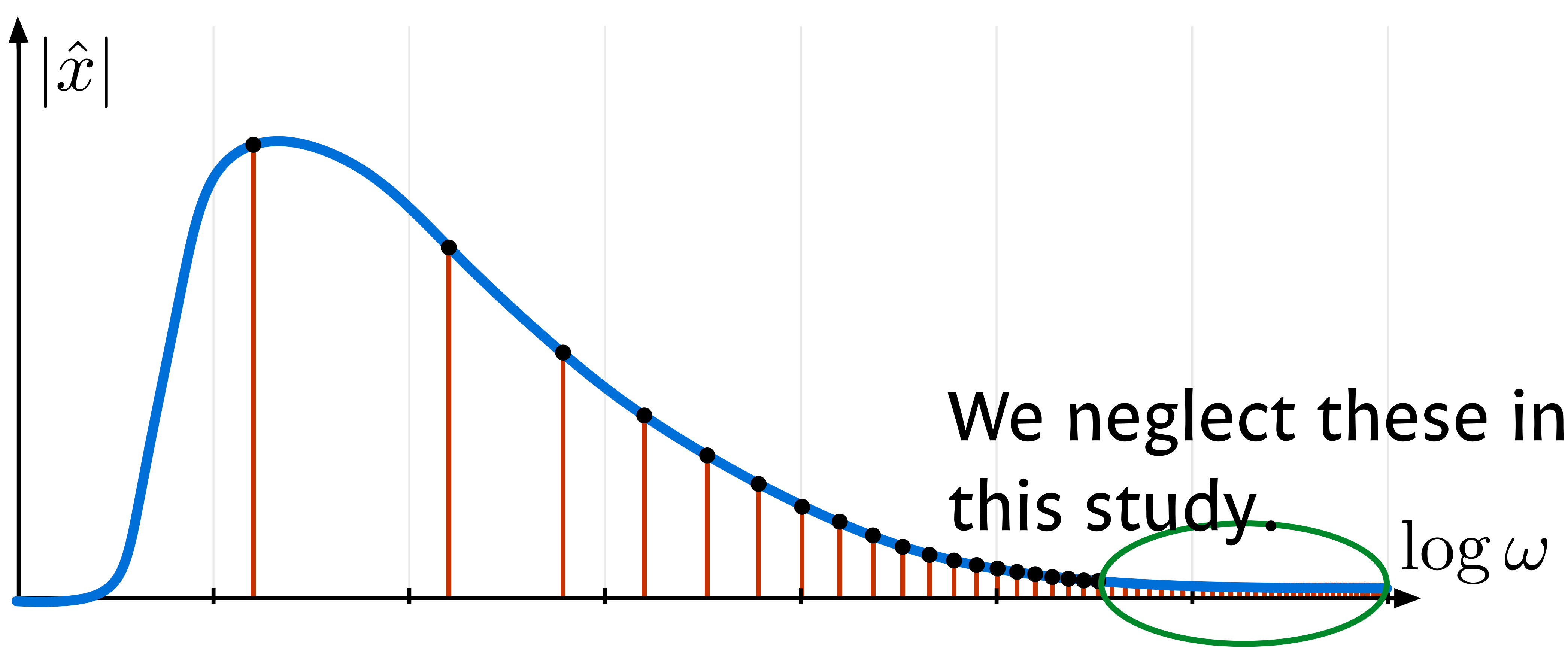


signal
$x(t)$
auditory
wavelet
modulus
scalogram
$|x \overset{t}{*} \psi_\lambda|$
spiral
wavelet
modulus
spiral
scattering
$\Big| \big| x \overset{t}{*} \psi_\lambda \big| \overset{t}{*} \psi_\alpha \overset{\log \lambda}{*} \psi_\beta \overset{\lfloor \log \lambda \rfloor}{*} \psi_\gamma \Big|$
log λ
t
averaging
$T = 500\mathrm{ms}$
averaging
$T = 500\mathrm{ms}$
reconstructed
signal
gradient
descent
reconstructed
scalogram
gradient
descent
averaged
scattering
indexed by
$(\lambda, \alpha, \beta, \gamma)$

Stationary source-filter model

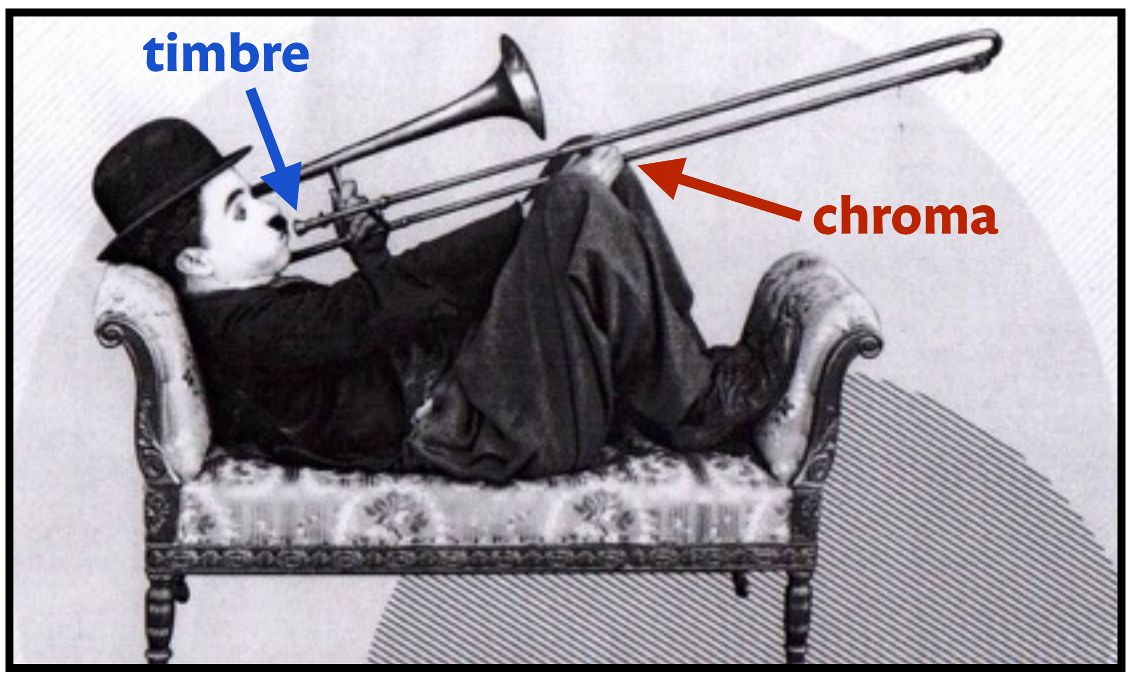
The stationary source-filter model is

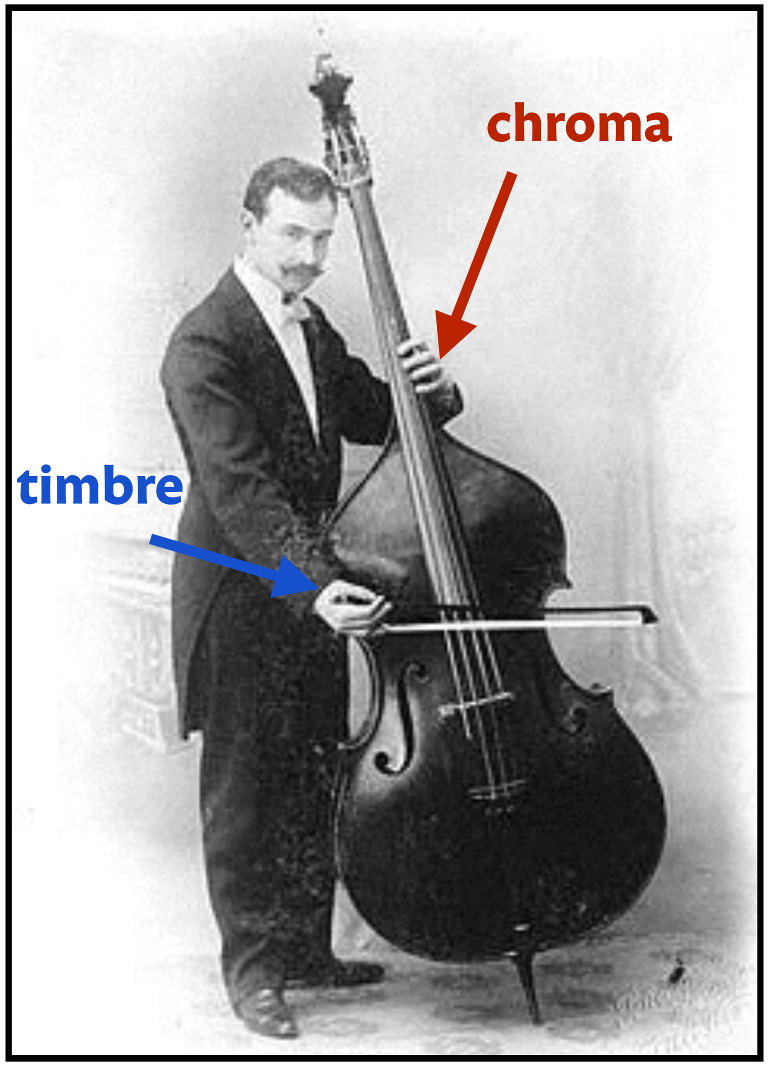
$$x(t) = [e * h](t) \text{ i.e. } \hat{x}(\omega) = [\hat{e} \times \hat{h}](\omega)$$



Two degrees of freedom

Musical transients are not regular in time-frequency...
... but in **time**-**chroma**-**octave**.





SLOW DEFORMATIONS

Let $\theta(t) \in \mathcal{C}^3$ be a time warp function.

$\dot{\theta}(t) > 0$ is the fundamental frequency of

$$e_\theta(t) = (e \circ \theta)(t).$$

$\dot{\nu}(t) > 0$ is the position of the formant (spectral peak)

$$h_\nu(t) = (h \circ \nu)(t).$$

The nonstationary source-filter model is defined as

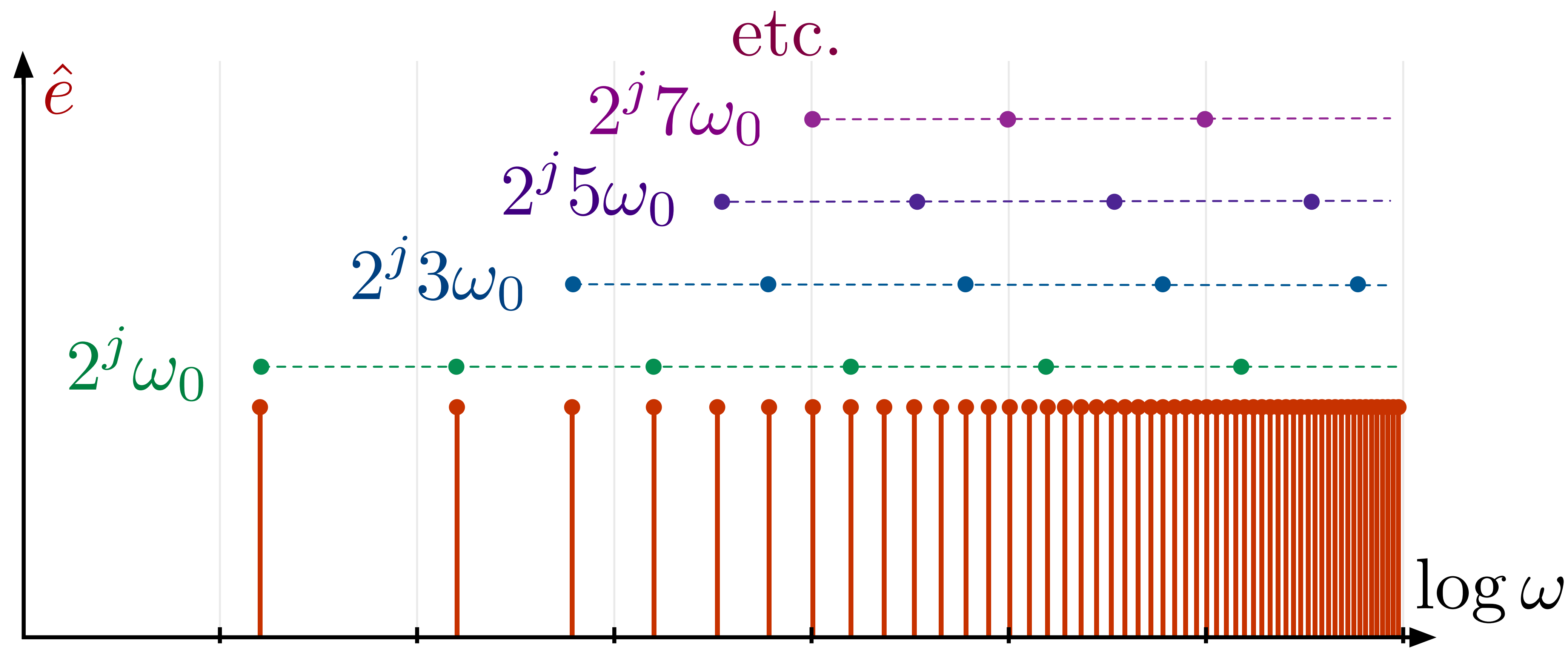
$$x_{\theta,\nu}(t) = [e_\theta * h_\nu](t).$$

Harmonicity property

The harmonic comb is self-similar:

$$\hat{e}(\omega) = \hat{e}(2^j \omega) \text{ for all } \omega > 1 \text{ and } j \in \mathbb{N}.$$

Regularity across octaves for a given chroma:

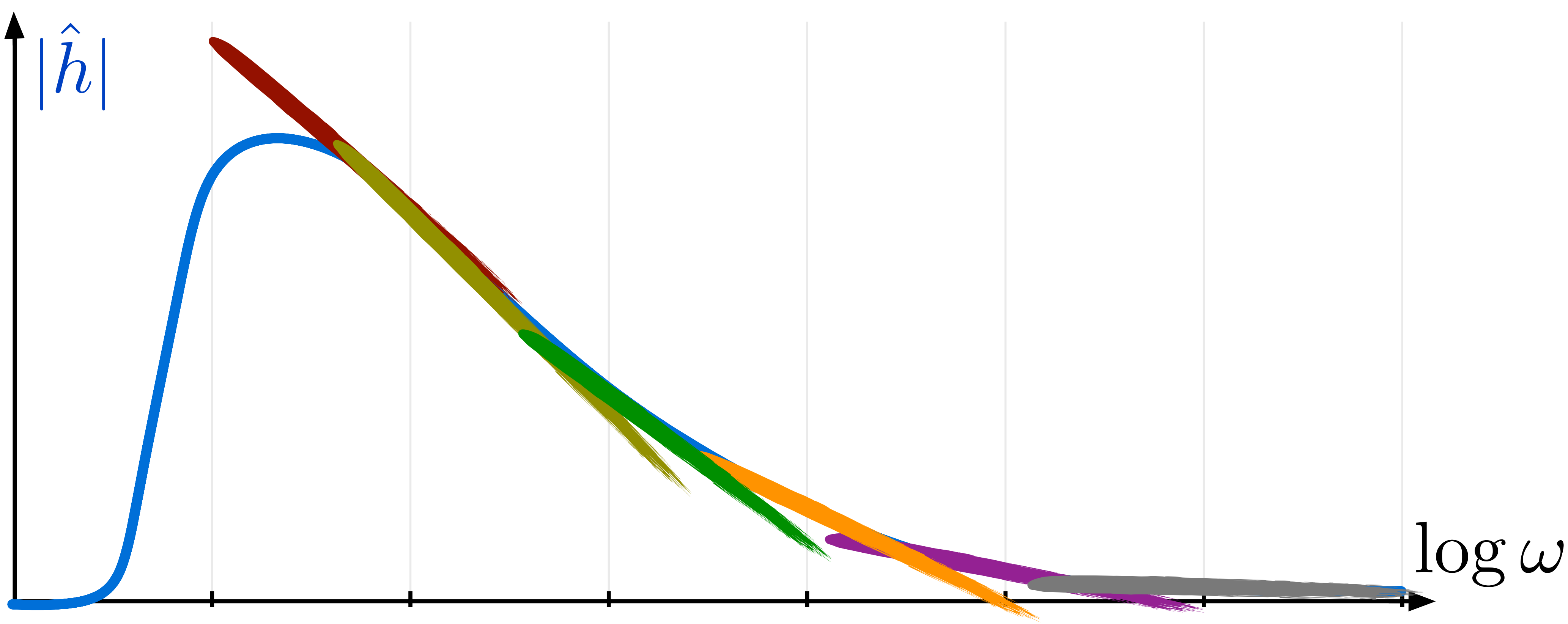


Spectral smoothness property

The spectral envelope is regular across semitones:

$$\frac{\partial \log |\hat{h}|}{\partial \omega} \ll \frac{\partial \log |\hat{e}|}{\partial \omega}.$$

Regularity along chromas within an octave:



Spiral wavelet ridges

- **Vanishing moment** property:
  Convolving a wavelet with a linear function yields almost zero.
- Harmonicity and spectral smoothness rewrite as

$$\left| \mathbf{U_1} e_\theta \overset{j_1}{*} \psi_\gamma \right| = 0 \quad \text{and} \quad \left| \mathbf{U_1} h_\nu \overset{\chi_1}{*} \psi_\beta \right| \approx 0 .$$

- Ridges are on a plane whose Cartesian equation is

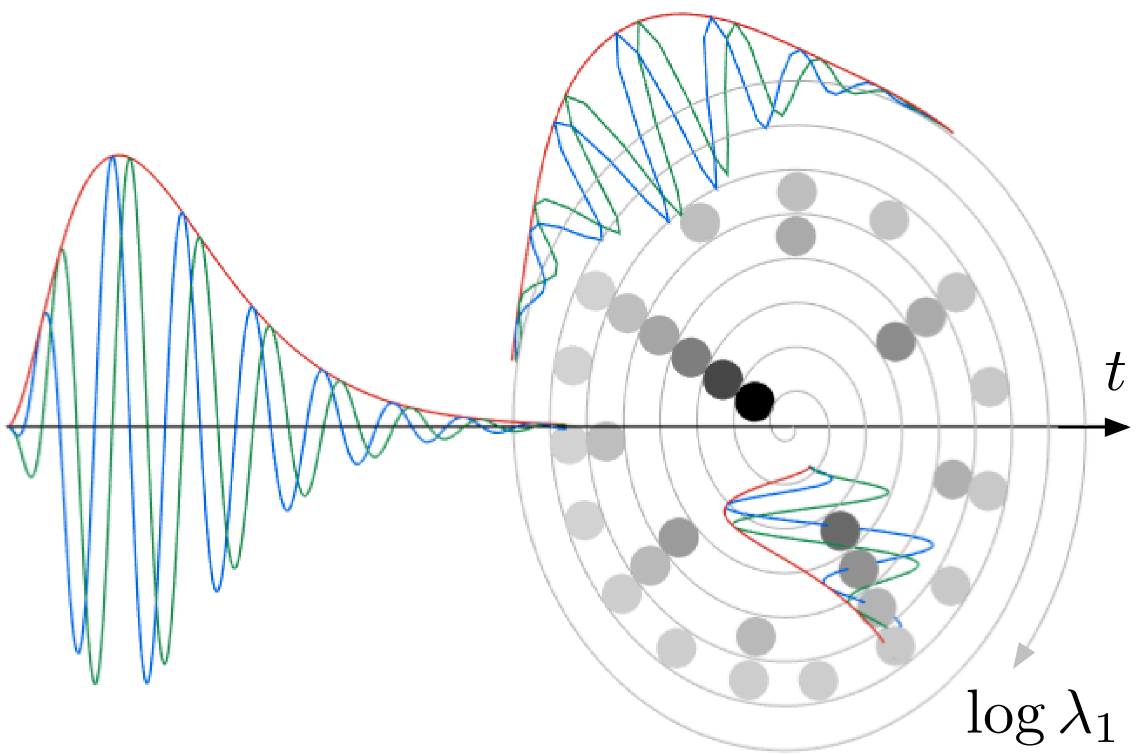
$$\alpha + \frac{\ddot{\theta}(t)}{\dot{\theta}(t)} \beta + \frac{\ddot{\nu}(t)}{\dot{\nu}(t)} \gamma = 0 .$$

- The same holds for averaged coefficients $\mathbf{S_2} x_{\theta,\nu}$ over T if

$$\left| \frac{\dddot{\theta}(t)}{\ddot{\theta}(t)} - \frac{\ddot{\theta}(t)}{\dot{\theta}(t)} \right| \ll T^{-1} \quad \text{and} \quad \left| \frac{\dddot{\nu}(t)}{\ddot{\nu}(t)} - \frac{\ddot{\nu}(t)}{\dot{\nu}(t)} \right| \ll T^{-1} .$$

Conclusions

- Natural sounds are nonstationary, but physically regular.
- In the pitch spiral, source-filter transients become translations.
- Spiral scattering yields source-filter velocities **without detection**.
- Encouraging results in invariant reconstruction.



Experiments can be reproduced at:
www.github.com/lostanlen/