Wavelet Scattering on the Pitch Spiral

Vincent Lostanlen and Stéphane Mallat
École normale supérieure de Paris, France

How to capture the spectro-temporal evolution of harmonic spectra ?

Purpose: classification, blind source separation, music transcription.

Auditory wavelets

Constant-Q band-pass filters :

$$\underbrace{\lambda}_{\text{center frequency}} = 2^{\overbrace{j}^{\text{octave}} + \frac{\overbrace{\chi}^{\text{chroma}}}{\underbrace{Q}_{\text{quality factor}}}}$$

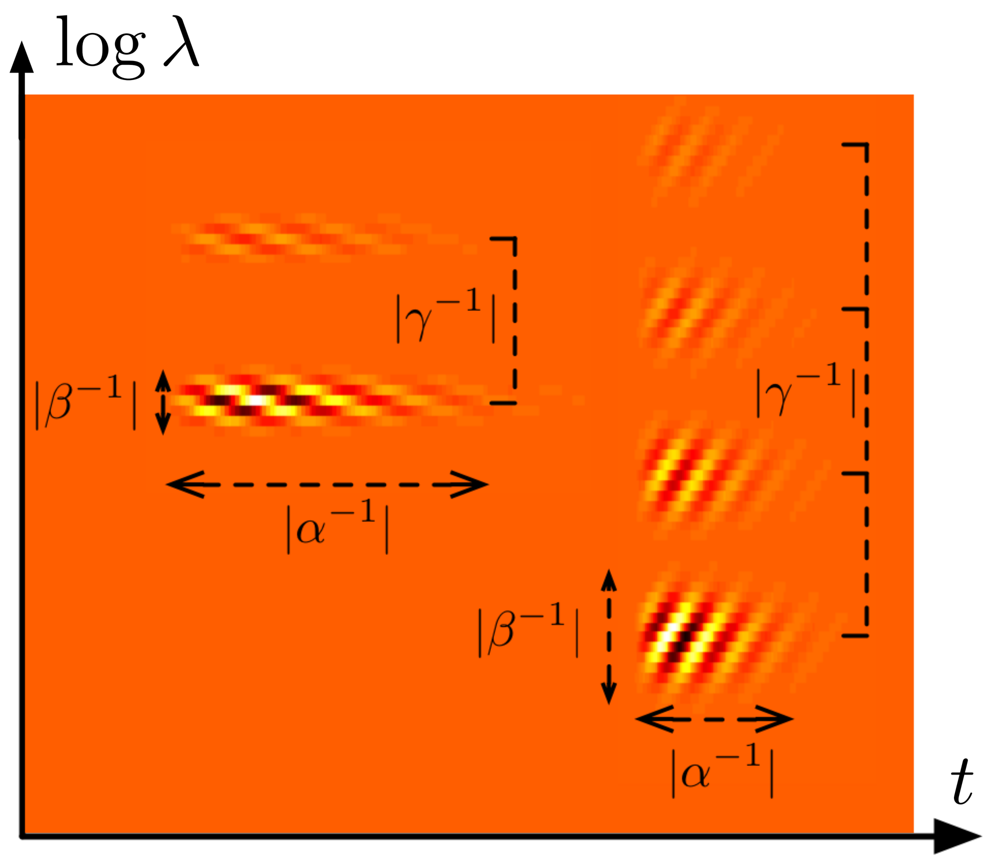
Convolutions with the wavelet filter bank and complex modulus yield the wavelet scalogram:

$$x_1(t, \log\lambda) = |x \overset{t}{*} \psi_\lambda|$$

Scattering transforms

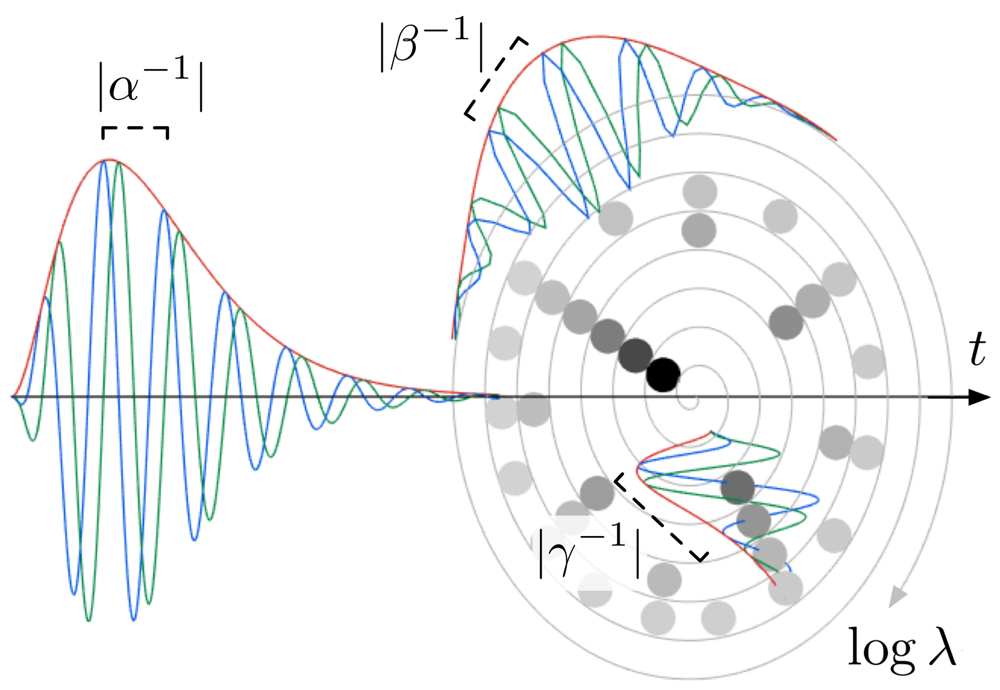
« Scatter » the scalogram with modulation wavelets to improve regularity.

We compose wavelet filter banks:
along time t with frequencies α ;
along chromas χ with frequencies β ;
along octaves j with frequencies γ.



See [Andén 2012] on α-scattering; [Andén 2015] on (α,β)-scattering.

Pitch spiral



Corroborated by music theory, psychology [Shepard, 1964], and neuroimaging [Warren, 2003].

Second-order spiral coefficients:

$$x_2 = \left| x_1 \overset{t}{*} \psi_\alpha \overset{\log\lambda_1}{*} \psi_\beta \overset{\lfloor \log\lambda_1 \rfloor}{*} \psi_\gamma \right|$$

Deformations of the source-filter model

Let θ and η be diffeomorphisms.

$$x_{\theta,\eta}(t) = \Big((e \circ \theta) \overset{t}{*} (h \circ \eta) \Big)(t)$$

If the following conditions are met:

(a) The scalogram separates partials

$$Q > 2\lambda / \dot{\theta}(t)$$

(b) Slowly varying source

$$1/Q \gg \lambda \| \ddot{\theta} / \dot{\theta} \|_\infty$$

(c) Slowly varying filter

$$1/Q \gg \lambda \| \ddot{\eta} / \dot{\eta} \|_\infty$$

(d) Spectral smoothness

$$Q \gg \lambda \times \| \mathrm{d}(\log|\hat{h}|)/\mathrm{d}\omega \|_\infty \times \| 1/\dot{\eta} \|_\infty$$

Therefore, the ridge coefficients of $x_2(t, \log\lambda, \alpha, \beta, \gamma)$ lie on a plane whose Cartesian equation is

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

Unsupervised learning of musical timbre

We computed the scattering coefficients of isolated notes from 16 instruments with varying pitches (28), nuances (3), and manufacturers (3). [Goto 2003]

We performed max-pooling along time and across B neighboring log-frequency bands, and applied logarithmic compression.

$$\mathbf{S_B} x_1 = \log \max_{\substack{t \in \mathbb{R} \\ |b| \leq B}} x_1(t, \log\lambda_1 \pm b)$$

$$\mathbf{S_B} x_2 = \log \max_{\substack{t \in \mathbb{R} \\ |b| \leq B}} x_2(t, \log\lambda_1 \pm b, \alpha, \beta, \gamma)$$

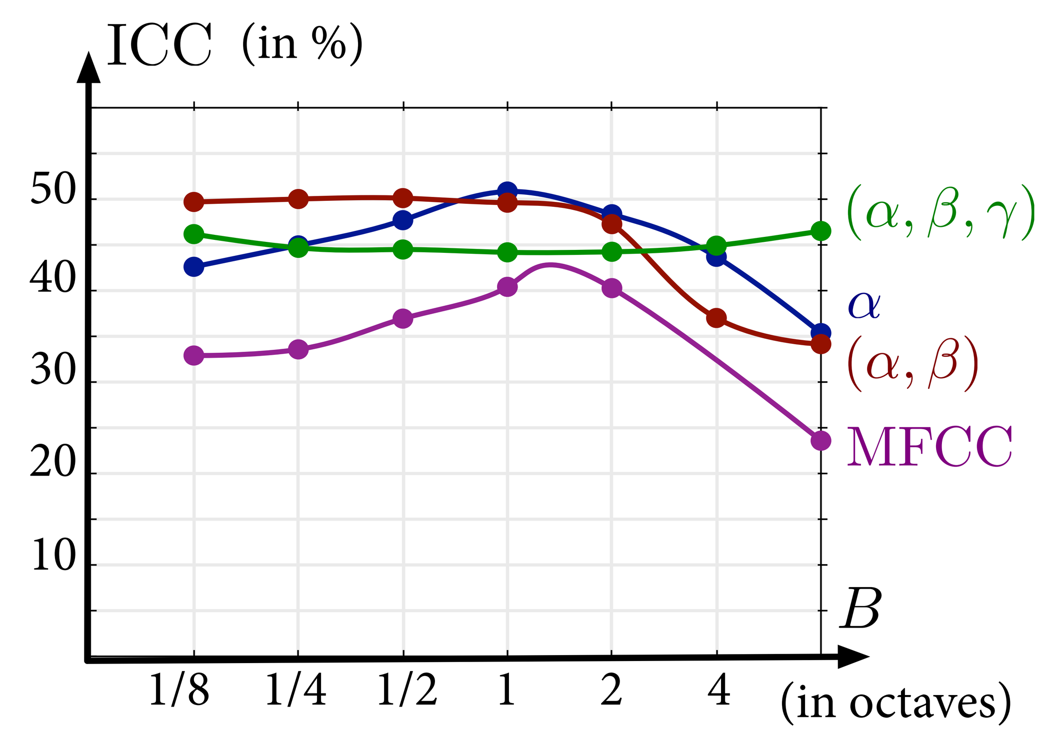
The feature vector is the concatenation of first-order and second-order coefficients.

$$\mathbf{S_B} x = (\mathbf{S_B} x_1, \mathbf{S_B} x_2)$$

We computed the centroid of each instrument in the feature space and measured Fisher's Intra-class Correlation Coefficient (ICC):

$$\mathrm{ICC} = \frac{\text{variance between centroids}}{\text{total variance}}$$

Results are charted below.



References

Andén and Mallat. *Scattering Representation of Modulated Sounds*, DAFx 2012.

Andén, Lostanlen, and Mallat. *Joint Time-frequency Scattering for Audio Classification*, MLSP 2015.

Goto, Hashigushi, Nikimura, and Oka. *RWC Music Database: Music Genre Database and Musical Instrument Sound Database*, ISMIR 2003.

Shepard. *Circularity in Judgments of Relative Pitch*, JASA 1964.

Warren, Uppenkamp, Patterson, and Griffiths. *Separating Pitch Chroma and Pitch Height in the Human Brain*, PNAS 2003.

The source code to reproduce experiments is available at

`www.github.com/lostanlen/scattering.m`

This work is supported by the ERC InvariantClass 320959 grant.