Wavelet scattering on the Shepard pitch spiral

18 mai 2015

May 18, 2015

1 Introduction

On the two dimensions of musical pitch

- Behind the circularity of the musical scale, lies a simple fact: when raising a harmonic note by exactly one octave, even-numbered partials seem unchanged, whereas odd-numbered partials disappear.
- Although frequency is expressed in Hertz as a rectilinear quantity, there are two natural degrees of freedom for pitch.
- Definition of pitch chroma and pitch height.
- Locality relations between pure tones is better expressed in a spiral (Shepard) than on a line.

Problem

- The spiral model is well-known in music theory and experimental psychology.
- However, existing methods in audio classification do not fully take advantage from its richness: they either picture pitch on a line (MFCC) or on a circle (chroma features).
- Consequently, they often fail to capture the joint dynamics of amplitude, pitch and timbre, that arise beyond 25 ms.

Contributions

- We present a new representation of sounds that linearizes the dynamics of pitch chroma and pitch height, while remaining stable to deformations in the time-frequency plane.
- It is an instance of the scattering transform, a generic operator which cascade wavelet convolutions and modulus nonlinearities.

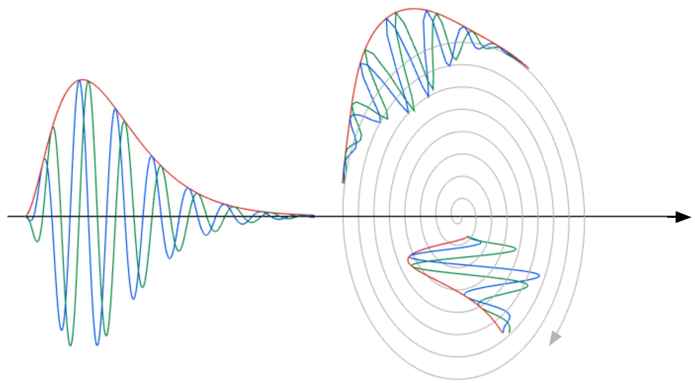

Figure 1: Wavelets in time, log-frequency and octave index.

- It is derived from the Shepard pitch spiral, in that convolutions are performed in time, log-frequency (correlated to pitch chroma) and octave index (correlated with pitch height).
- We give numerical results on transient signals of varying pitch chroma and pitch height, showing that their velocities on the spiral are disentangled and characterized.

2 Scattering transform on the wavelet scalogram

Shepard pitch spiral

$$\log_2 \lambda_1 = \lfloor \log_2 \lambda_1 \rfloor + \{\log_2 \lambda_1\}$$

The integer part $\lfloor \log_2 \lambda_1 \rfloor$ is related to pitch height, whereas the fractional part $\{\log_2 \lambda_1\}$ is related to pitch chroma.

Spiral wavelet

$$\Psi_{\lambda_2}(t, \lambda_1) = \psi_\alpha(t) \times \psi_\beta(\log_2 \lambda_1) \times \psi_\gamma(\lfloor \log_2 \lambda_1 \rfloor)$$

$$x_2(t, \lambda_1, \lambda_2) = |x_1 * \Psi_{\lambda_2}|(t, \lambda_1)$$

Visualization

- We visualize x_2 in 2d by fixing t, λ_1 and α. The axes correspond to β and γ.

3 Capturing pitch chroma and pitch height velocities

3.1 Shepard tones

Definition

- A Shepard tone is a sum of sine waves which are one octave apart:

$$x(t) = \sum_{k=0}^{K-1} \sin(2\pi 2^k f_0 t)$$

- The original definition definition specifies a bell curve in relative amplitudes across octaves, that we choose to drop for simplicity.

Shepard tones on the spiral

- Shepard tones do not have a specific pitch height ; instead, they are entirely determined by their pitch chroma.

Self-similarity

- Shepard tones are self-similar: $x(2t) \approx x(t)$.

3.2 Varying pitch chroma: the Shepard-Risset glissando

Definition

- A Shepard-Risset glissando is built by playing a Shepard tone at a tape speed that increases exponentially.
- This corresponds to a uniform motion of the pitch chroma.
- Shepard-Risset glissandi are periodic because of the circularity of chromas.
- A Shepard-Risset glissando is defined as:

$$x_{\mathrm{c}}(t) = x\left(\frac{2^{V_{\mathrm{c}} t}}{V_{\mathrm{c}} \log 2}\right)$$

- V_c is the pitch chroma velocity, measured in octaves per second.

First-order coefficients

- The scalogram is proportional to

$$x_{1,\mathrm{c}}(t, \lambda_1) \propto \widehat{\psi_{\lambda_1}}\left(2^{k(t,\lambda_1)+V_{\mathrm{c}} t} f_0\right)$$

where the integer-valued function $k(t, \lambda_1)$ is the index of the nearest partial from the point (t, λ_1) in the time-frequency plane.

Second-order coefficients

- The scattering coefficients of the Shepard-Risset glissando can be expressed in closed form, by applying a result proven by Joakim Andén on joint time-frequency characterization of exponential chirps:

$$x_{2,c}(t, \lambda_1, \lambda_2) \propto \widehat{\psi}\left(-\frac{\beta}{\alpha} V_{\mathrm{c}}\right)$$

  Consequently, x_2 is maximal when the slope α/β is equal to $-V_c$, that is, when the direction of oscillation is orthogonal to the glissando trajectories.

- The joint scattering transform behaves like an oriented edge detector in the time-scale plane.

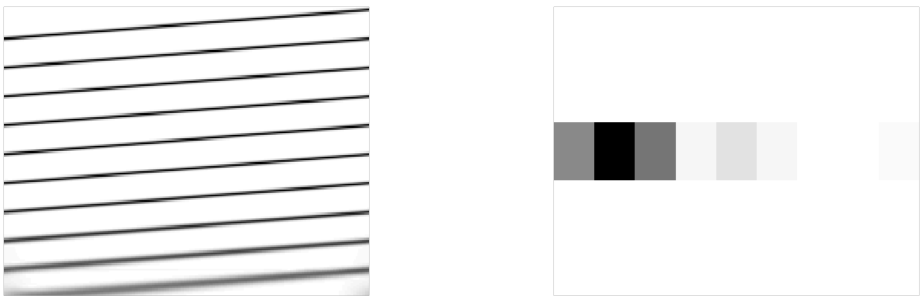

Figure 2: The Shepard-Risset glissando: first-order coefficients (left) and second-order spiral coefficients (right).

3.3 Varying pitch height: the "Shepard arpeggio"

Definition

- Less popular than the Shepard-Risset glissando of varying chroma and fixed height, is the "Shepard arpeggio" of varying height and fixed chroma.

$$x_g(t) = \sum_{k=0}^{K-1} 2^k g(V_{\mathrm{h}} t - k) \sin(2\pi 2^k f_0 t)$$

- V_{h} is the pitch height velocity, measured in octaves per second.
- g is a zero-centered bump function ; we have chosen a standard Gaussian in our experiments.

First-order coefficients

$$x_{1,\mathrm{g}}(t, \lambda_1) \propto g\left(V_{\mathrm{h}} t - k(\lambda_1)\right) \times \widehat{\psi_{\lambda_1}}\left(2^{k(\lambda_1)} f_0\right)$$

where the integer-valued function $k(\lambda_1)$ is the index of the nearest partial from the frequency λ_1.

Second-order coefficients We have an approximate closed form if g is a Gaussian. For (t, λ_1) roughly within the arpeggio, $\beta = 0$ and (α, γ) such that α/γ is close to $-V_{\mathrm{h}}$.

$$x_{2,g}(t, \lambda_1, \lambda_2) \propto g\left(\frac{\frac{\alpha}{\gamma} + V_{\mathrm{h}}}{\alpha}\right).$$

Consequently, x_2 is maximal when the slope α/γ is equal to $-V_h$, that is, when the direction of oscillation is orthogonal to the arpeggio trajectory.

Figure 3: The Shepard arpeggio: first-order coefficients (left) and second-order spiral coefficients (right).