LINEARIZING DEFORMATIONS OF THE HARMONIC SOURCE-FILTER MODEL WITH THE SPIRAL SCATTERING TRANSFORM

VINCENT LOSTANLEN

ABSTRACT. A classical model for voiced speech production consists in the convolution of a harmonic glottal source $e(t)$ with a vocal tract filter $h(t)$. Introducing indepedent deformations to both components brings realistic variability to pitch and spectral envelope. We study the decomposition of the deformed source-filter model in the spiral scattering transform.

Deformations of the harmonic comb. A 2π-periodic Dirac pulse train is defined as

$$e(t) = 2\pi \sum_{n=-\infty}^{+\infty} \delta(t - 2\pi n).$$

Let $\theta \in \mathcal{C}^3(\mathbb{R})$ a time warp function whose first derivative $\dot{\theta}(t)$ is positive for all $t \in \mathbb{R}$. We defined a warped source as

$$e_\theta(t) = (e \circ \theta)(t) = 2\pi \sum_{n=-\infty}^{+\infty} \delta(\theta(t) - 2\pi n).$$

Observe that $\dot{\theta}(t)$ is the fundamental frequency of $e_\theta(t)$.

Deformations of the broadband filter. Let $h(t)$ be the impulse response of a broadband filter. Its Fourier transform $\hat{h}(\omega)$ is the spectral envelope, and is related to the properties of the vocal tract in speech production. Again, we define a time warp funtion $\nu \in \mathcal{C}^3(\mathbb{R})$ such that $\dot{\nu}(t) > 0$ for all $t \in \mathbb{R}$. Its derivative $\dot{\nu}(t)$ is a local dilation factor for the spectral envelope $\hat{h}(\omega)$.

Deformed source-filter model. We extend the classical source filter model $[e * h](t)$ by warping independently the source and the filter before combining them:

$$x(t) = [e_\theta * h_\nu](t).$$

Main result. We will show that, for $\dot{\theta}(t)$ and $\dot{\nu}(t)$ reasonably regular over the support of first-order wavelets $\psi_{\lambda_1}(t)$, the local maxima of spiral scattering coefficients $x_2(t, \log \lambda_1, \log \lambda_2)$ are clustered on a plane in the (α, β, γ) space of scattering frequencies, where:

- $\alpha > 0$ is the frequency along time, in Hertz.
- β is the "quefrency" along chromas, in cycles per octaves. Typically, $|\beta| \geq 1$.
- γ is the "quefrency" across octaves, in cycles per octaves. Typically, $|\gamma| \leq 1$.

This plane satisfies the Cartesian equation

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

This result means that harmonic sounds overlapping both in time and frequency could be resolved according to their respective source-filter velocities.

Outline of the proof. Our proof is organized as follows:

(1) Assuming that the spectral envelope is almost constant over the frequential support of the wavelet, the convolution between h_ν and ψ_{λ_1} can be factorized as a product between the Fourier transform $\hat{h}$ and ψ_{λ_1}, up to a phase shift.[1]
(2) The Q lowest partials in the harmonic comb do not interfere. Thus, we may apply the wavelet ridge theorem to each of them independently.
(3) Combining the two previous results yields a factorization of deformed source and deformed filter in the scalogram.
(4) The harmonicity property implies that the convolution along chromas only applies to the deformed source. In turn, the spectral smoothness property implies that the convolution across octaves only applies to the deformed filter.
(5) We use the wavelet ridge theorem to give the phase of the deformed source after convolution along chromas. The same theorem applies to the deformed filter after convolution across octaves.
(6) The two phase functions extracted in the previous part add up. Hence, after convolution along chromas and across octaves, the analytic amplitude and instantaneous frequency along time are available in closed form.
(7) We use the wavelet ridge theorem a third time to localize in time the relative variations of the two instantaneous frequencies.
(8) The Cartesian equation is derived from the extraction of spiral wavelet ridges.

Assumptions. We work under the following assumptions:

(1) Q large enough to discriminate non-negligible partials,
(2) slowly varying source: $\|\ddot{\theta}/\dot{\theta}\|_\infty \ll \lambda_1/Q$,
(3) slowly varying filter: $\|\ddot{\nu}/\dot{\nu}\|_\infty \ll \lambda_1/Q$,
(4) spectral envelope is almost constant over one semitone: $\|\dot{\hat{h}}/\hat{h}\|_\infty \ll Q/\lambda_1$.
(5) spectral envelope is almost linear over one octave.

In addition, we use the fact that harmonicity implies self-similarity: $e(t) = e(2^n t)$ for all $|t| > 2\pi$ and $n \in \mathbb{N}$. By means of the Poisson summation formula, this rewrites in the Fourier domain as $\hat{e}(\omega) = \hat{e}(2^j \omega)$ for all $|\omega| > 1$ and $j \in \mathbb{N}$.

[1] Interestingly, this result is a *swapped* formulation of the wavelet ridge theorem. Whereas the wavelet ridge theorem assumes that the signal has slowly varying amplitude and fundamental frequency over the support of the wavelet, here we assume that the wavelet has slowly varying amplitude over the support of the signal. Henceforth, the roles of "analytic signal" and "analytic wavelet" are interchanged. The form of the result remains the same.

Wavelet filter bank. Let $\hat{g}(\omega)$ be a real symmetric window of bandwidth $1/Q$ and unit norm. An approximately analytic wavelet is constructed by multiplying $g(t)$ with a sine wave of frequency 1:

$$\psi(t) = g(t)\exp(\mathrm{i}t).$$

Dilated wavelets are written as

$$\psi_{\lambda_1}(t) = \lambda_1 \psi(\lambda_1 t).$$

The role of the wavelet transform is to localize the power spectrum of $e_\theta(t)$ along short-term windows $\psi_{\lambda_1}(t)$ upon which both time warps $\theta(t)$ and $\nu(t)$ can be linearized.

Wavelet ridge theorem. The following theorem provides an approximate closed form for the wavelet transform of a real analytic signal whose amplitude and instantaneous frequency have slow variations.

Theorem. *Let $f(t) = a(t)\cos\varphi(t)$. The wavelet transform of f is equal to*

$$f * \psi_{\lambda_1}(t) = \frac{1}{2} a(t)\exp(\mathrm{i}\varphi(t)) \times \left(\hat{g}\left(1 - \frac{\dot{\varphi}(t)}{\lambda_1}\right) + \varepsilon(t, \lambda_1) \right),$$

where the corrective term satisfies

$$\|\varepsilon(t,\lambda_1)\|_\infty \leq \left\| \frac{\dot{a}(t)}{\lambda_1 a(t)} \right\|_\infty + \left\| \frac{\ddot{a}(t)}{\lambda_1^2 a(t)} \right\|_\infty + \left\| \frac{\ddot{\varphi}(t)}{\lambda_1^2} \right\|_\infty + \sup_{|\omega| \geq |\dot{\varphi}|(t)/\lambda_1} |\hat{g}(\omega)|.$$

The fourth term is negligible if $\dot{\varphi}(t) \geq \lambda_1/Q$.

The proof can be found in Mallat's Wavelet Tour, section 4.4.

0.1. Linearization of filter deformations (or the swapped wavelet ridge theorem).

The first-order Taylor series expansion of νover the support of ψ_{λ_1} leads to an approximate expression of the convolution $[h_\nu * \psi_{\lambda_1}](t)$:

$$[h_\nu * \psi_{\lambda_1}](t) = \int_{-\infty}^{+\infty} h(\nu(t) - \dot{\nu}(t)u)\psi_{\lambda_1}(u)\,\mathrm{d}u$$

The linear change of variables $u' = \dot{\nu}(t) \times u$ yields

$$[h_\nu * \psi_{\lambda_1}](t) = \int_{-\infty}^{+\infty} h(\nu(t) - u')\psi_{\lambda_1}\left(\frac{u'}{\dot{\nu}(t)}\right)\frac{\mathrm{d}u'}{\dot{\nu}(t)}.$$

The dilation factor $\dot{\nu}(t)$ may be turned into a translation in the log-frequency domain. Indeed we have:

$$[h_\nu * \psi_{\lambda_1}](t) = \frac{1}{2\pi}\int_{\infty}^{+\infty} \widehat{h}(\omega)\widehat{\psi_{\nu\dot{(}t)\lambda_1}}$$

0.2. Linearization of source deformations.

0.3. Source-filter factorization in the scalogram.

0.4. Harmonicity and spectral smoothness properties.

0.5. Extraction of instantaneous frequencies.

0.6. Cartesian equation.