# Phase reconstruction of spectrograms with linear unwrapping: application to audio signal restoration


Paul MAGRON, Roland BADEAU, Bertrand DAVID
Institut Mines-Télécom - Télécom ParisTech - CNRS/LTCI - Paris, France


## Introduction

### Problem

Many audio signal processing techniques act on spectrogram-like representations in the Time-Frequency (TF) domain.

- The phase information is often discarded, lost or partially missing.
- We need to reconstruct the phase to re-synthesize time-domain signals.

### Our approach

Explicit STFT phase calculation of mixtures of sinusoids:
→ *Horizontal unwrapping* over time frames (temporal coherence).
Similar calculation on impulse signals:
→ *Vertical unwrapping* over frequency channels (spectral coherence).
Dynamic estimation of instantaneous frequencies:
→ *Non-stationary signals* (cellos and speech).

## Horizontal phase reconstruction

### Sinusoidal modeling

STFT of a complex sinusoid $(A,\ f_0,\ \phi_0)$, $\forall (k,t) \in [0, F-1] \times [0, T-1]$:

$$X(k,t) = A e^{2i\pi f_0 S t + i\phi_0} W\left(\frac{k}{F} - f_0\right),$$

with:

- $S$ the time shift between successive frames,
- $W$ the discrete time Fourier transform of the analysis window $w$.

### Phase relationship

Relationship between phases of adjacent bins:

$$\phi(k,t) = \phi(k,t-1) + 2\pi S f_0(t), \text{ where } \phi(k,t) = \angle X(k,t).$$

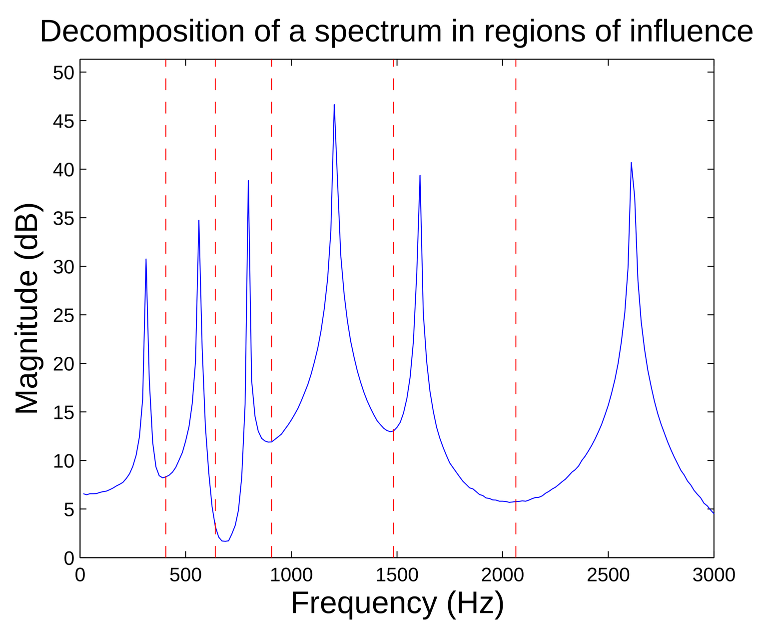
Instantaneous frequency estimation at each time frame makes representing variable frequency signals (vibratos) possible.

### $f_0$ estimation

Quadratic Interpolated FFT (QIFFT) and decomposition of the frequency range in *regions of influence* [1]. For the $p$-th peak of magnitude $A_p$ in channel $k_p$:

$$I_p = \left[\frac{A_p k_{p-1} + A_{p-1} k_p}{A_p + A_{p-1}}; \frac{A_p k_{p+1} + A_{p+1} k_p}{A_p + A_{p+1}}\right].$$

The greater $A_p$ is relatively to $A_{p-1}$ and $A_{p+1}$, the wider $I_p$ is.




## Onset phase reconstruction

### Impulse model

STFT of an impulse signal centered at time $n_0$:

$$X(k,t) = A w(n_0 - St) e^{-2i\pi \frac{k}{F}(n_0 - St)}.$$

Relationship between the phases of two successive frequency channels:

$$\phi(k,t) = \phi(k-1,t) - \frac{2\pi}{F}(n_0(k) - St).$$

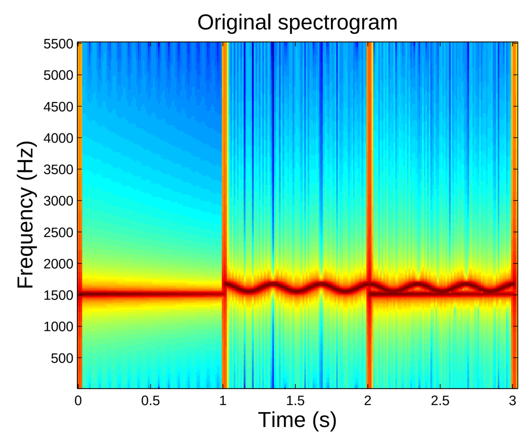
### $n_0$ estimation

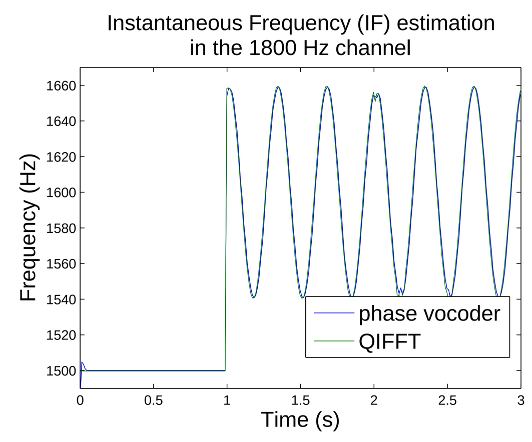
The shape of $|X|$ is similar to that of $w$:

- Least square estimation method : $n_0$-LS.
- Alternatively, quadratic interpolation over time frames : $n_0$-QI.

## Experimental results

### Horizontal phase reconstruction




- Frequency error between phase vocoder and QIFFT estimates.
- Phase retrieval with the Griffin Lim (GL) algorithm [2] or with the proposed Phase Unwrapping (PU) method.
- Signal reconstruction quality measured with the SDR (in dB).

| Dataset | IF Error (%) | GL | PU |
|---|---|---|---|
| Piano notes | 0.38 | −6.9 | **2.5** |
| Piano pieces | 0.36 | −12.6 | **1.7** |
| String quartets | 0.41 | −9.7 | **5.3** |
| Speech excerpts | 0.52 | −0.4 | **0.5** |

### Onset phase reconstruction

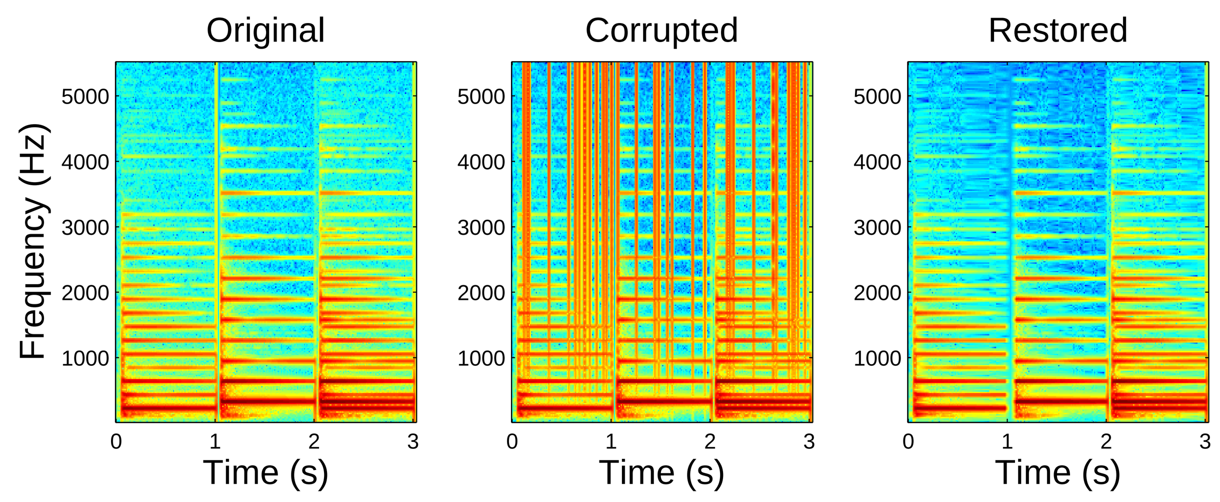
Test on piano signals with several strategies: impulse model-based $n_0$ estimation, random phases (Rand), zero-phases (0) and alternating partial phases between 0 and $\pi$ (Alt).

| Method | GL | $n_0$-LS | $n_0$-QI | Rand | 0 | Alt |
|---|---|---|---|---|---|---|
| SDR (dB) | −7.9 | −4.0 | **−2.6** | −4.3 | −4.7 | −3.5 |

$n_0$-estimation with QI: best result, neat percussive attack.

## Audio restoration

Temporal signals are corrupted with clicks.
Restoration in the TF domain: linear interpolation of the log-magnitude:




Below, phase restoration with the PU method or the GL algorithm. Alternatively: traditional method (AR) or High Resolution NMF (HRNMF)[3].

| Dataset | AR | HRNMF | GL | PU |
|---|---|---|---|---|
| Piano notes | 11.4 | **16.9** | 8.6 | **11.7** |
| Piano pieces | 4.3 | **10.9** | 5.9 | **7.1** |
| String quartets | 8.2 | **10.6** | 6.6 | **7.1** |
| Speech excerpts | 8.3 | **10.9** | 8.9 | **9.4** |

- PU performs better than AR and GL.
- HRNMF: best performance, but this method is not blind: the model is learned on the non-corrupted data.

## Future research

- Improve onset phase estimation (transient modeling).
- Exploit observed phase data for inferring missing bins.
- Use time-invariant parameters (phase offset between partials).
- Source separation framework: mixture phase can be exploited.


References:

[1] J. Laroche and M. Dolson, Improved phase vocoder time-scale modification of audio, IEEE Trans. on SAP, 1999.

[2] D. Griffin and J. S. Lim, Signal estimation from modified short-time Fourier transform, IEEE Trans. on ASSP, 1984.

[3] R. Badeau and M. D. Plumbley, Multichannel high resolution NMF for modelling convolutive mixtures of non-stationary signals in the time-frequency domain, IEEE Trans. on ASLP, 2014.