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

Paul MAGRON, Roland BADEAU, Bertrand DAVID
Institut Mines-Telecom - Telecom 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 (cells and speech).

Horizontal phase reconstruction

Sinusoidal modeling
STFT of a complex sinusoid \( A, f_0, \phi_0 \), \( \forall (k, t) \in [0, F - 1) \times [0, T - 1) \):
\[
X(k, t) = A e^{2\pi f_0 S + \phi_0} W \left( \frac{k}{F} - f_0 \right),
\]
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:
\[
\phi(k, t) = \phi(k, t - 1) + 2\pi f_0 (t), \text{ where } \phi(k, t) = \angle X(k, t).
\]
Instantaneous frequency estimation at each time frame represents variable frequency signals (vibratos) possible.

\[ f_0 \text{ estimation} \]

Quadratic Interpolated FFT (QIFFT) and decomposition of the frequency range in regions of influence [1]. For the \( \phi \)th peak of magnitude \( A_p \) in channel \( k_p \):
\[
I_p = \frac{A_p k_{p-1} + A_{p-1} k_p}{A_p + A_{p-1}} - \frac{A_p k_{p+1} + A_{p+1} k_p}{A_p + A_{p+1}}.
\]
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 \( n_0 \):
\[
X(k, t) = A w(n_0 - St) e^{2\pi \phi(n_0 - St)}.
\]
Relationship between the phases of two successive frequency channels:
\[
\phi(k, t) = \phi(k - 1, t - \frac{2\pi}{F} (n_0(k) - St)).
\]

\[ n_0 \text{ estimation} \]
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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>IF Error (%)</th>
<th>GL</th>
<th>PU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piano notes</td>
<td>0.38</td>
<td>-6.9</td>
<td>2.5</td>
</tr>
<tr>
<td>Piano pieces</td>
<td>0.36</td>
<td>-12.6</td>
<td>1.7</td>
</tr>
<tr>
<td>String quartets</td>
<td>0.41</td>
<td>-9.7</td>
<td>5.3</td>
</tr>
<tr>
<td>Speech excerpts</td>
<td>0.52</td>
<td>-0.4</td>
<td>0.5</td>
</tr>
</tbody>
</table>

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).

Audio restoration

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

Future research

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

References: