Assessing the Quality of the Extraction and Tracking of Sinusoidal Components

Mathieu Lagrange and **Sylvain Marchand**

SCRIME – LaBRI, University of Bordeaux 1

sylvain.marchand@labri.fr

Aims

sinusoidal modeling analysis consists of 2 main stages:

- peak extraction (many estimators, serious SP studies)
- partial tracking (few algorithms, evaluation: open issue)

we wish to:

- evaluate each stage of the analysis chain separately
- initiate the **formalization** for the partial tracking stage
- lay the emphasis on the **evaluation** to (invite others to)
  - compare existing analysis methods
    (objective comparison, in a realistic framework)
  - propose new evaluation methodologies
    (less empirical. . .)

Sinusoidal Modeling

[McAulay and Quatieri (IEEE TASSP 1986)]
[Serra and Smith (Computer Music Journal 1990)]

The audio signal s is given by:

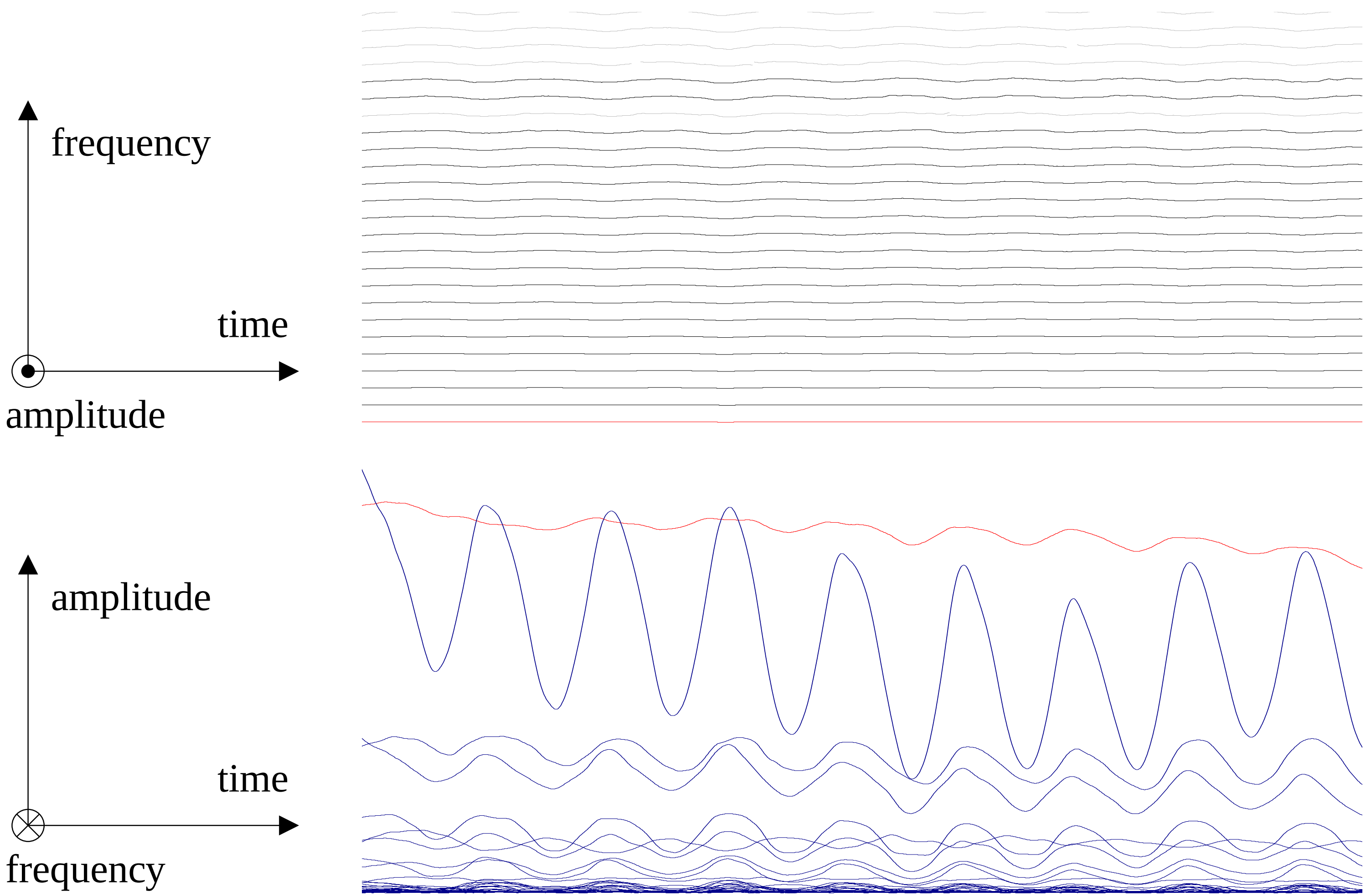
$$s(t) = \sum_{n=1}^{N} a_n(t)\ \cos(\phi_n(t))$$

where P is the number of **partials** and

$$\frac{d\phi_n}{dt}(t) = 2\pi f_n(t) \quad i.e. \quad \phi_n(t) = \phi_n(0) + 2\pi \int_0^t f_n(u)\, du$$

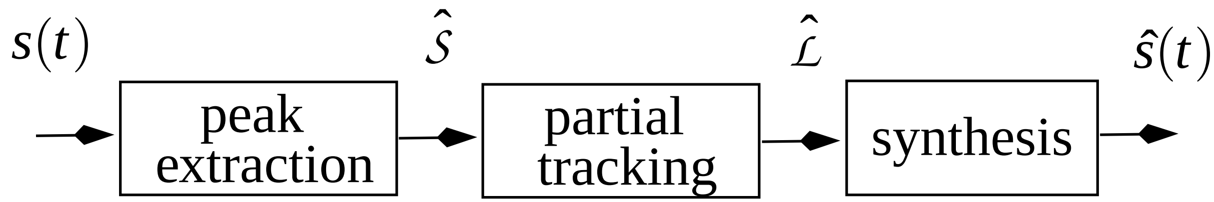
The functions f_n, a_n, and ϕ_n are the instantaneous frequency, amplitude, and phase of the n-th partial, respectively.

Sinusoidal Modeling: Sound Example



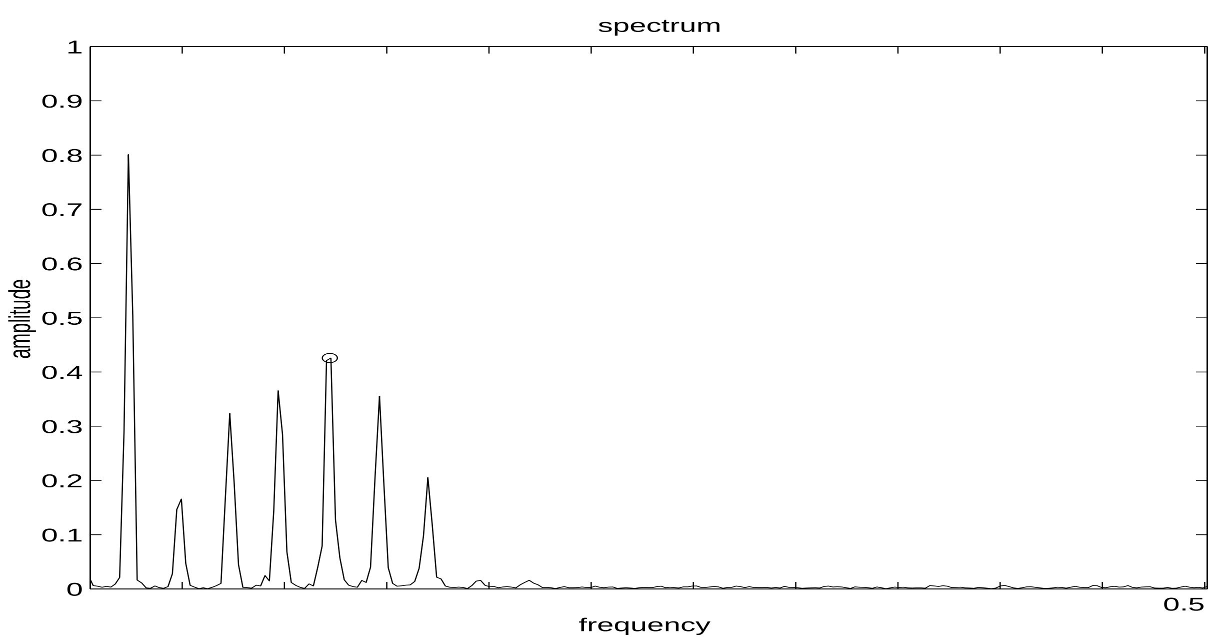
(alto saxophone, 1 second)

Peak Extraction

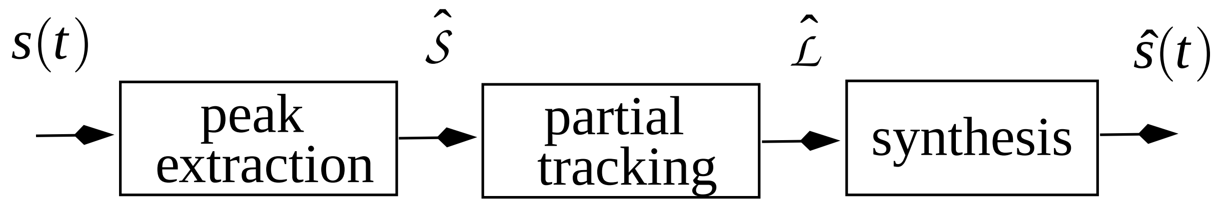


peak extractor: at each frame k, from a short-time signal,

- detects the right number of sinusoidal components;
- estimates the parameters of each component precisely.



Peak Extraction



peak extractor: at each frame k, from a short-time signal,

- detects the right number of sinusoidal components;
- estimates the parameters of each component precisely.

The peak number i of frame number k is defined by four normalized parameters:

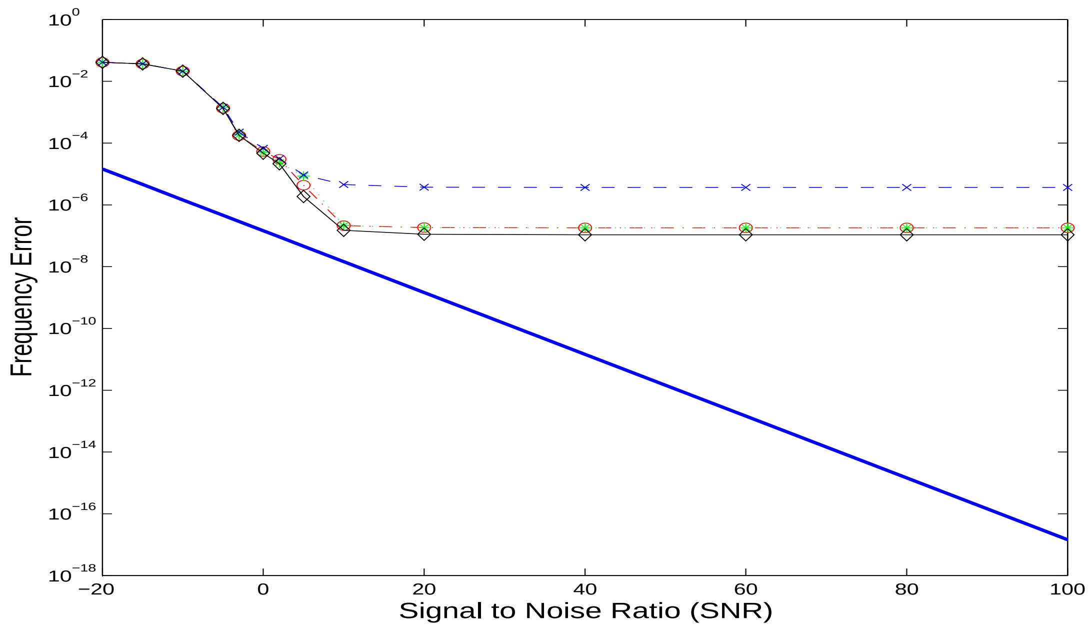
$$p_i^k = \left(f_i^k, a_i^k, \phi_i^k, c_i^k\right)$$

the **frequency** f_i^k, amplitude a_i^k, phase ϕ_i^k, and **confidence** c_i^k.

The set of peaks is $\mathcal{S} = \bigcup_k \mathcal{S}_k$ where $\mathcal{S}_k = \bigcup_i \left\{p_i^k\right\}$.

Peak Extraction: Existing Evaluation Methods

- evaluation of the **frequency** estimation:
  - $s = x + y$, where
    x is a sinusoidal signal and y a Gaussian noise;
  - the variance of the frequency error is plotted
    in function of the Degradation-SNR (D-SNR);
  - comparison to the lower Cramér-Rao Bound (CRB):



Peak Extraction: Existing Evaluation Methods

- evaluation of the **frequency** estimation:
  - $s = x + y$, where
    x is a sinusoidal signal and y a Gaussian noise;
  - the variance of the frequency error is plotted
    in function of the Degradation-SNR (D-SNR);
  - comparison to the lower Cramér-Rao Bound (CRB):

$$\text{D-SNR} = 10 \log 10 \left(\sum_l y^2(l) / \sum_l x^2(l) \right)$$

see e.g. [Hainsworth & Macleod, DAFx 2003]
[Lagrange & Marchand, DAFx 2005]

Peak Extraction: Existing Evaluation Methods

- evaluation of the detection: specific quantities
  - [Keiler & Marchand, DAFx 2002]
    mean number of detected peaks per frame,
    mean number of peaks per frame that are
    in the reference signal but not detected,
    and mean number of peaks that are
    detected but not present in the reference signal.
  - [Hainsworth & Mcleod, DAFx 2003]
    number of *correctly* identified sinusoids,
    and number of falsely detected sinusoids.
  - . . .

Peak Extraction: Proposed Evaluation Methodology

The cost between two **peaks** p_i and p_j is given by:

$$c(p_i, p_j) = (2(f_i - f_j))^2$$

The cost between the two **sets** $\mathcal{S}$ and $\hat{\mathcal{S}}$ is given by:

$c_I \leftarrow 0$
$\mathcal{S}_r \leftarrow \mathcal{S}$; $\hat{\mathcal{S}}_r \leftarrow \hat{\mathcal{S}}$
while $\mathcal{S}_r \neq \emptyset$ **and** $\hat{\mathcal{S}}_r \neq \emptyset$ **do**
    take $p_j \in \hat{\mathcal{S}}_r$ such that $c_j = \max_{p_k \in \hat{\mathcal{S}}_r} c_k$
    find $p_i \in \mathcal{S}_r$ such that $|f_i - f_j| = \min_{k \neq j} |f_k - f_j|$
    $c_I \leftarrow c_I + c(p_i, p_j)$
    $\mathcal{S}_r \leftarrow \mathcal{S}_r - \{p_i\}$; $\hat{\mathcal{S}}_r \leftarrow \hat{\mathcal{S}}_r - \{p_j\}$
end while

UNIVERSITE BORDEAUX 1 Sciences Technologies

Peak Extraction: Proposed Evaluation Methodology

The over and under estimation costs are then defined by:

$$c_O = \sum_{j=1}^{\#\hat{\mathcal{S}}_r} c_j \quad \text{and} \quad c_U = \sum_{i=1}^{\#\mathcal{S}_r} c_i = \#\mathcal{S}_r$$

Finally, we propose a normalized overall cost:

$$C = (c_O + c_U + c_I)/N$$

which reflects both the issues of

- finding the correct number of peaks;
- estimating their parameters precisely.

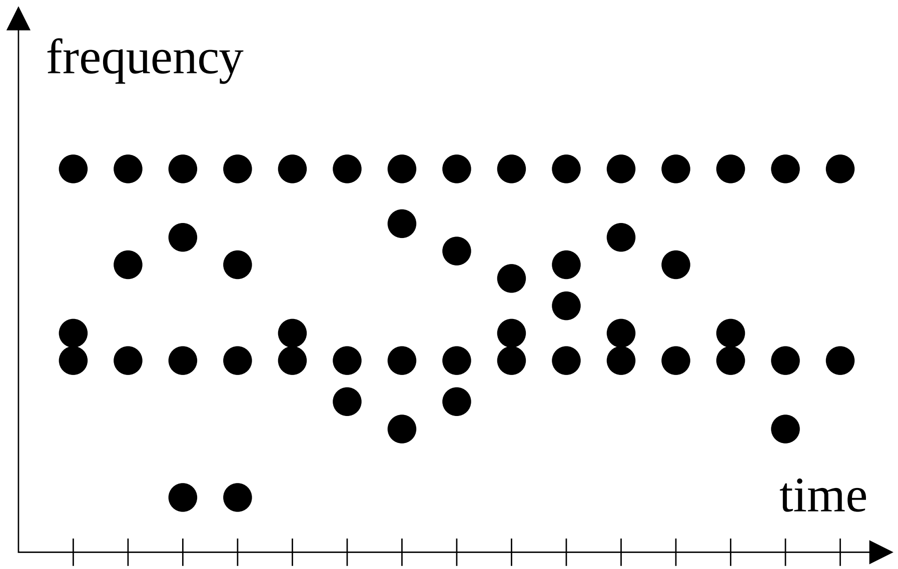
UNIVERSITE BORDEAUX 1 Sciences Technologies

Peak Extraction: Test Cases

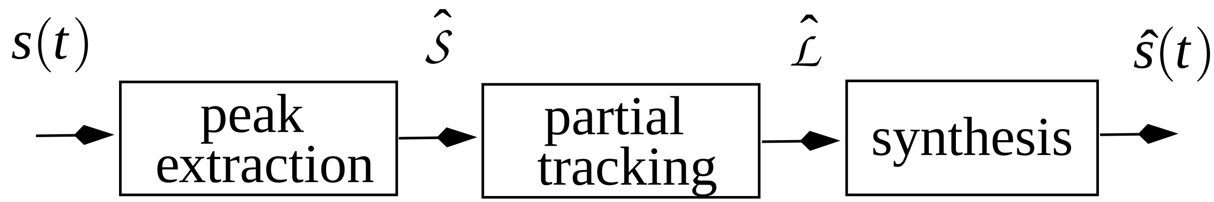
- Precision:
  one sinusoid embedded in a Gaussian noise,
  and C is plotted in function of the SNR;
  thus a perfect extractor will attain the CRB.
- Resolution:
  two sinusoids with the same amplitude,
  and C is plotted in function of the frequency gap
  between the two sinusoids.
- Overall Quality Assessment:
  several sinusoids with **the same amplitude**
  and randomized frequencies / phases,
  C is plotted in function of the number of sinusoids.

Partial Tracking

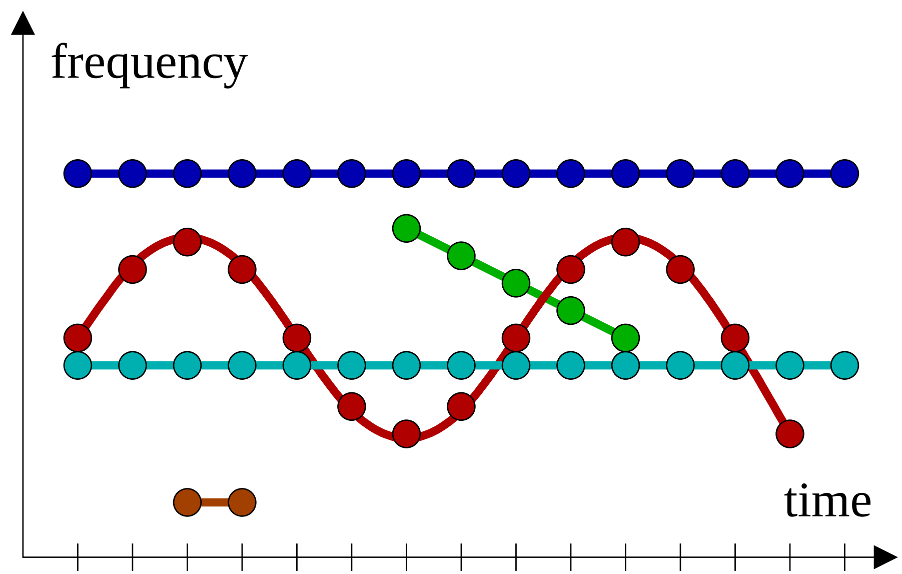
partial tracker: identifies continuities between peaks of successive frames to build a continuous representation of the evolutions of the sinusoidal parameters over time (the $\mathcal{L}$ set).



Partial Tracking

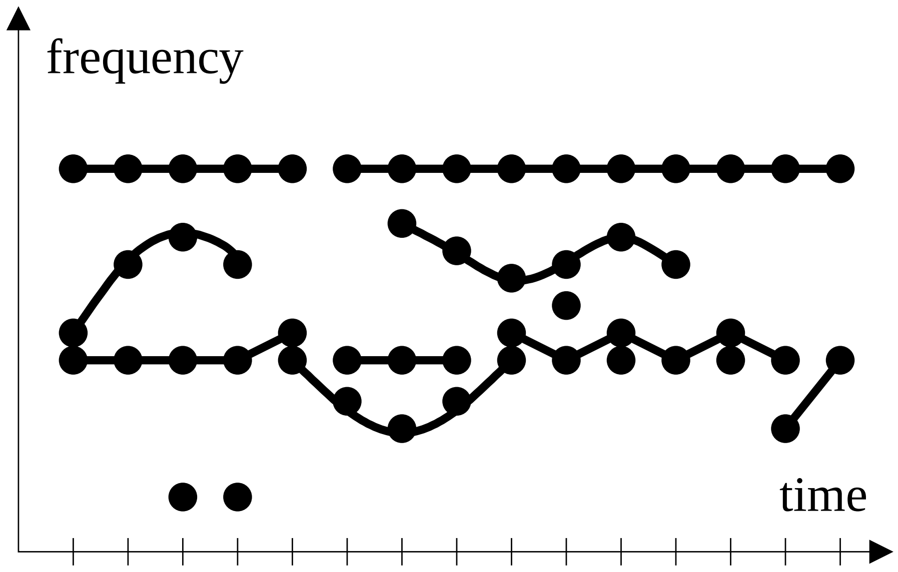


partial tracker: identifies continuities between peaks of successive frames to build a continuous representation of the evolutions of the sinusoidal parameters over time (the $\mathcal{L}$ set).

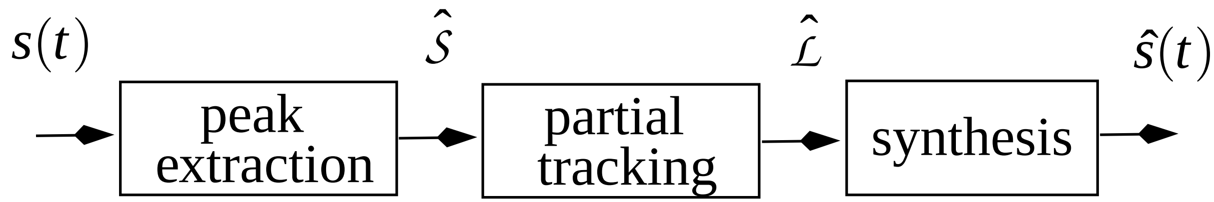


Partial Tracking

partial tracker: identifies continuities between peaks of successive frames to build a continuous representation of the evolutions of the sinusoidal parameters over time (the $\mathcal{L}$ set).



Partial Tracking



partial tracker: identifies continuities between peaks of successive frames to build a continuous representation of the evolutions of the sinusoidal parameters over time (the $\mathcal{L}$ set). A partial is a vector of peaks with successive time indices:

$$P_n(m) = \{F_k(m), A_k(m), \Phi_k(m)\}, \quad \forall m \in [b_n, \cdots, b_n + l_n - 1]$$

where P_n is the partial number n, of length l_n,
and that appeared (was born) at frame index b_n;
$P_n(m)$ is the peak of time index m of this partial.

The set of partials is $\mathcal{L} = \bigcup_{n=1}^{N} P_n$.

Partial Tracking: Existing Evaluation Methods

subjective and non automatic:

- visual examination of the trajectories of the partials:
  may illustrate specific properties, but not compare algorithms...
- listening tests:
  also evaluate the peak extraction and synthesis stages, but one stage of the sinusoidal modeling chain may favor another stage...

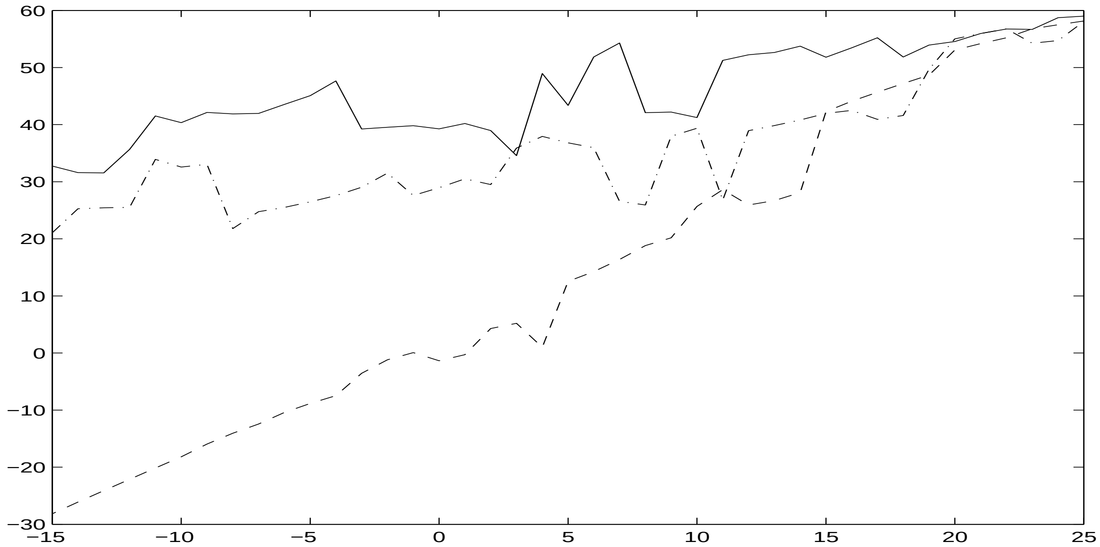
objective and automatic:

- Reconstruction Signal-to-Noise Ratio (R-SNR):

$$\text{R-SNR} = 10 \log 10 \left(\sum_l (s(l) - \hat{s}(l))^2 / \sum_l s^2(l) \right)$$

Partial Tracking: Existing Evaluation Methods

- R-SNR plotted in function of the D-SNR:



(evaluation of the management of crossing partials)
[Lagrange, Marchand & Rault, IEEE ICASSP 2004 & 2005]

- specific factors: $R_d = n_d / n_e$ and $R_f = n_f / n_e$,
  where n_d is the number of detected tracks, n_f is the number of false tracks, and n_e is the number of expected tracks.
  [Satar-Boroujeni & Shafai, DAFx 2005]

Partial Tracking Evaluation. . .

How to compare two **sets** of **partials**?

Partial Tracking Evaluation. . .

How to compare two **sets** of **partials**?

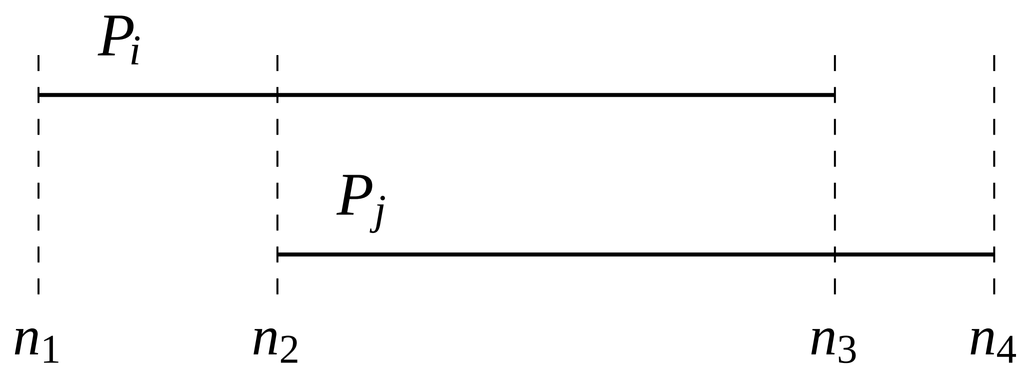
each partial being a vector of peaks. . .

Partial Tracking Evaluation. . .

How to compare two **sets** of **partials**?

each partial being a vector of peaks. . .
each peak having several parameters. . .

Partial Tracking: Proposed Evaluation Methodology



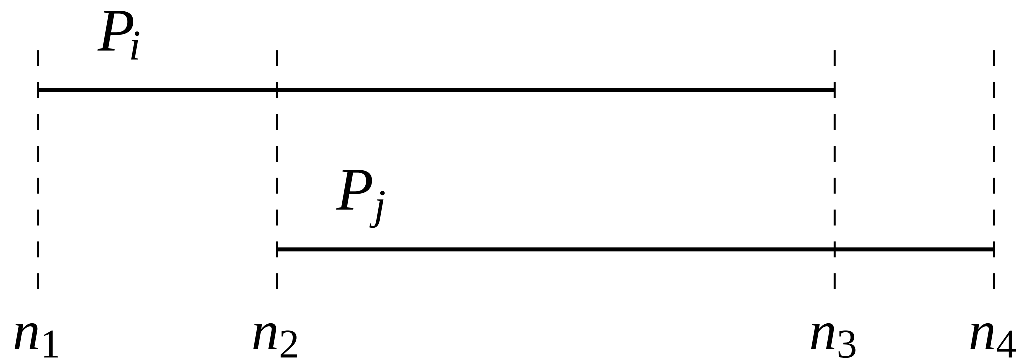
Considering that $b_i \leq b_j$, we define four frame indices:

$$\begin{aligned} n_1 &= b_i \\ n_2 &= \min(b_j, b_i + l_i) \\ n_3 &= \min(\max(b_j, b_i + l_i), b_j + l_j) \\ n_4 &= \max(b_i + l_i, b_j + l_j) \end{aligned}$$

(b_i being the birth index of partial i, and l_i its length).
When a partial k is not active, its parameters are null:

$$A_i(k) = F_i(k) = 0 \;\text{ if } k < b_i \text{ or } k \geq b_i + l_i$$

Partial Tracking: Proposed Evaluation Methodology



The cost between two **partials** P_i and P_j is given by:

$$c\left(P_i, P_j\right) = \underbrace{\sum_{k=n_2}^{n_3} {\color{blue}2} \left|F_i(k) - F_j(k)\right| \cdot \left|A_i(k) - A_j(k)\right|}_{\text{cost of the common part}}$$

$$+ \underbrace{\sum_{k=n_1}^{n_2} \max(A_i(k), A_j(k)) + \sum_{k=n_3}^{n_4} \max(A_i(k), A_j(k))}_{\text{unmatched boundaries: cumulated amplitudes}}$$

Partial Tracking: Proposed Evaluation Methodology

The cost between the two **sets** $\mathcal{L}$ and $\hat{\mathcal{L}}$ is given by:

$c_I \leftarrow 0$
$\mathcal{L}_r \leftarrow \mathcal{L}$
$\hat{\mathcal{L}}_r \leftarrow \hat{\mathcal{L}}$
while $\mathcal{L}_r \neq \emptyset$ **and** $\hat{\mathcal{L}}_r \neq \emptyset$ **do**
  take $P_j \in \hat{\mathcal{L}}_r$ such that $A_j = \max_{P_k \in \hat{\mathcal{L}}_r} A_k$
  find $P_i \in \mathcal{L}_r$ such that $c(P_i, P_j) = \min_{P_k \in \mathcal{L}_r} c(P_k, P_j)$
  $c_I \leftarrow c_I + c(P_i, P_j)$
  $\mathcal{L}_r \leftarrow \mathcal{L}_r - \{P_i\}$
  $\hat{\mathcal{L}}_r \leftarrow \hat{\mathcal{L}}_r - \{P_j\}$
end while

the overall amplitude of a partial P_k being: $A_k = \sum_{i=b_k}^{b_k+l_k-1} A_k(i)$.

Partial Tracking: Proposed Evaluation Methodology

The over and under estimation costs are then defined by:

$$c_O = \sum_{j=1}^{\#\hat{\mathcal{L}}_r} A_j \quad \text{and} \quad c_U = \sum_{i=1}^{\#\mathcal{L}_r} A_i$$

Finally, we propose a normalized overall cost:

$$C = (c_O + c_U + c_I)/N$$

which reflects the quality of the tracking.

Partial Tracking: Test Cases

protocol:

- a set of partials $\mathcal{L}$ is considered as a reference;
- $\mathcal{L}$ is converted back into a set of peaks $\mathcal{S}$;
- degradations may be applied to $\mathcal{S}$:
  - adding / removing peaks,
  - degrading the parameters of partials;
- the resulting set of peaks is used as the input of the tracker, which produces another set of partial $\hat{\mathcal{L}}$;
- finally, $\hat{\mathcal{L}}$ is compared to the reference $\mathcal{L}$.

polyphonic case: (simplified)

simulated by accumulating several sound sources into $\mathcal{S}$.

Conclusion and Future Work

- neither a review of existing methods…
- nor the proposal of a new method…

instead, we propose an **evaluation** methodology / protocol:

- objective and automatic
- application independent
- does not compare time-signals
- each stage of the chain can be evaluated separately
- reasonable but still empirical…
- **a first proposal, as a starting point…**
- **results?**

Conclusion and Future Work

DAFx'07 contest on sinusoidal analysis / synthesis

Participants will have the opportunity to submit:

- peak extractors,
- partial trackers,
- evaluation methodologies / protocols,
- reference sounds. . .

The resulting survey will be (semi-)automatically generated and published in the proceedings.

Conclusion and Future Work

DAFx'07 contest on sinusoidal analysis / synthesis

Participants will have the opportunity to submit:

- peak extractors,
- partial trackers,
- evaluation methodologies / protocols,
- reference sounds...

The resulting survey will be (semi-)automatically generated and published in the proceedings.

$\Longrightarrow$ people interested?

(Roland Badeau, Michaël Betser, Philippe Depalle, Stephen Hainsworth, Florian Keiler,

Mathieu Lagrange, Sylvain Marchand, Axel Röbel, Xavier Serra, Julius Smith, Udo Zölzer...)