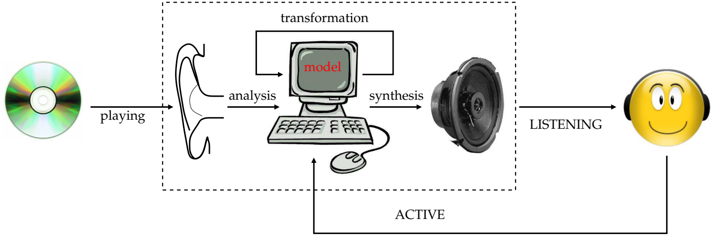
Advances in Spectral Modeling of Musical Sound

Sylvain Marchand

LaBRI – CNRS, Université Bordeaux 1

Active Listening

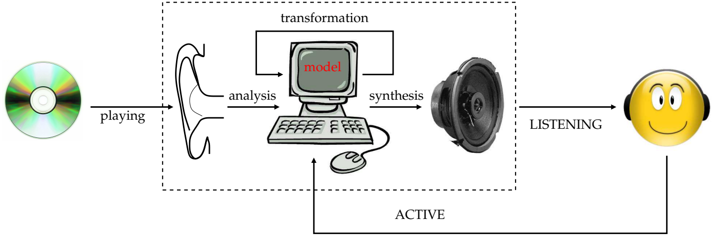


transformation of perceptive parameters

- loudness
- pitch
- timbre
- time scale
- space location

for each sound entity of the musical mix…

Active Listening

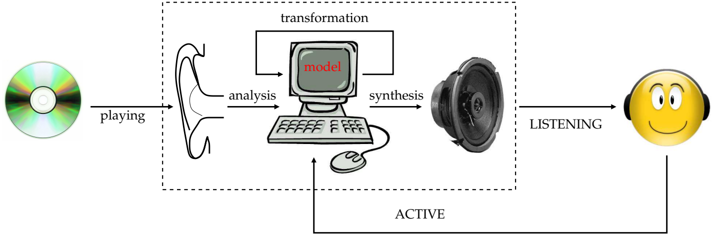


transformation of perceptive parameters

- loudness
- pitch
- timbre
- time scale
- space location

for each sound entity of the musical mix. . .

Active Listening

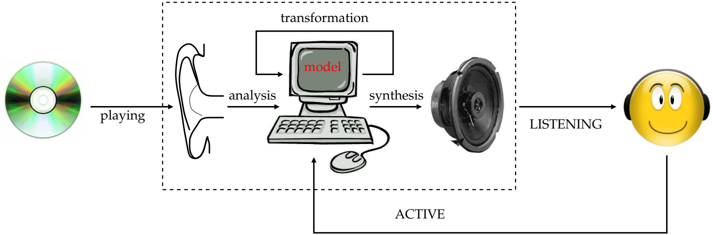


transformation of perceptive parameters

- loudness
- pitch
- timbre
- time scale
- space location

for each sound entity of the musical mix. . .

Active Listening

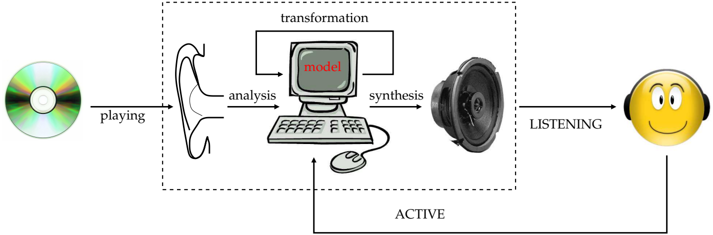


transformation of perceptive parameters

- loudness
- pitch
- timbre
- time scale
- space location

for each sound entity of the musical mix…

Active Listening

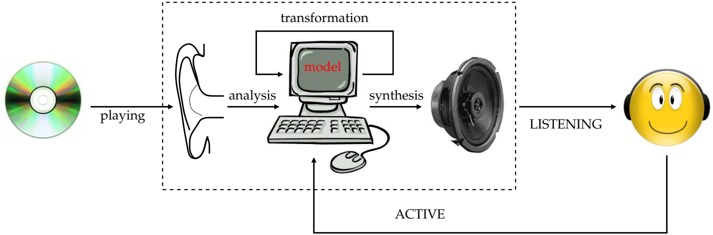


transformation of perceptive parameters

- loudness
- pitch
- timbre
- time scale
- space location

for each sound entity of the musical mix. . .

Active Listening

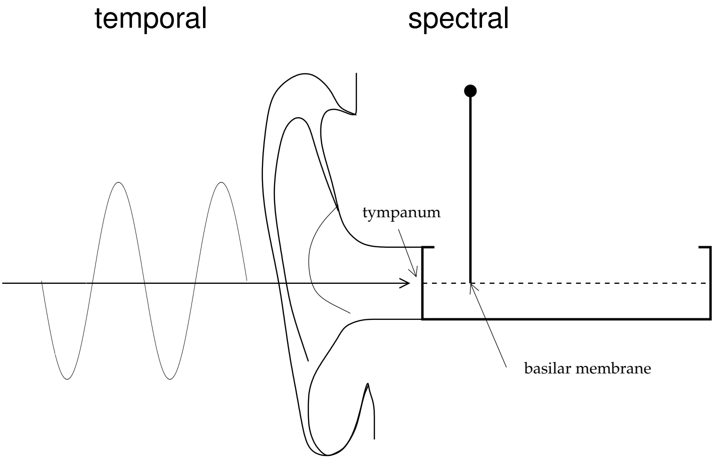


transformation of perceptive parameters

- loudness
- pitch
- timbre
- time scale
- space location

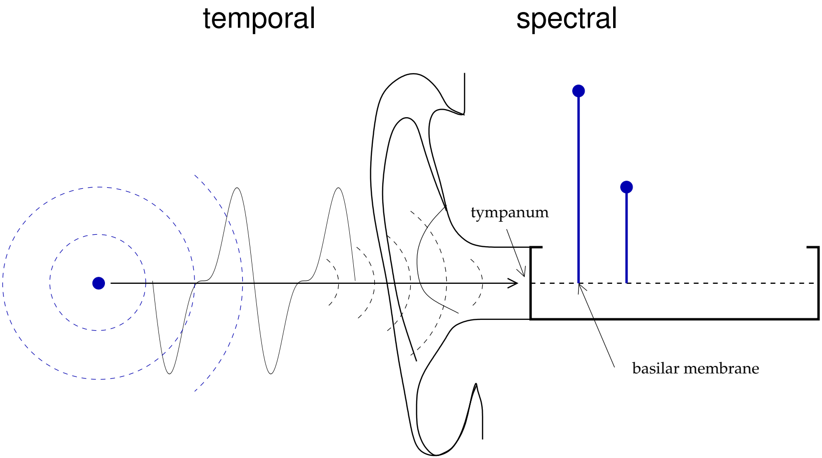
for each sound entity of the musical mix…

Sound



musical sound: complex, polyphonic, noisy, non-stationary. . .

Sound

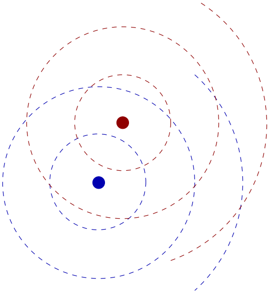


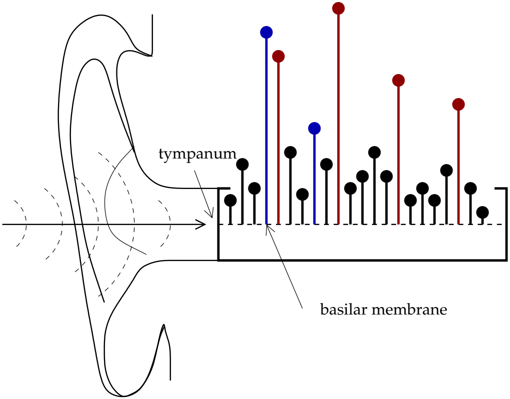
musical sound: complex, polyphonic, noisy, non-stationary…

Sound

temporal

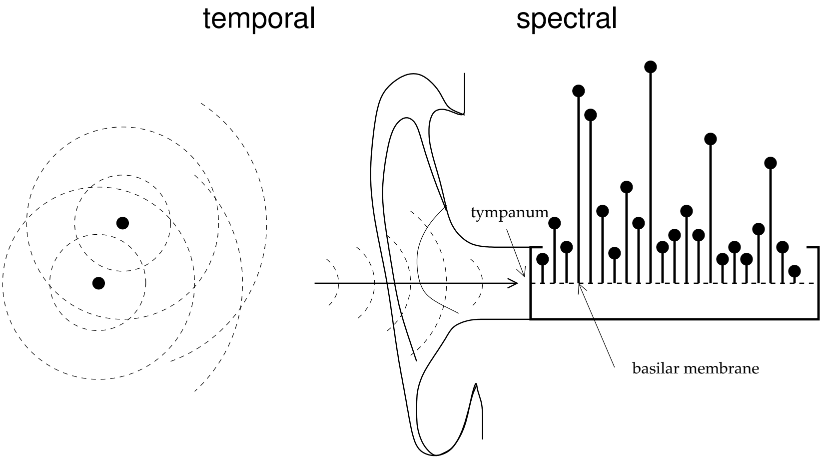
spectral





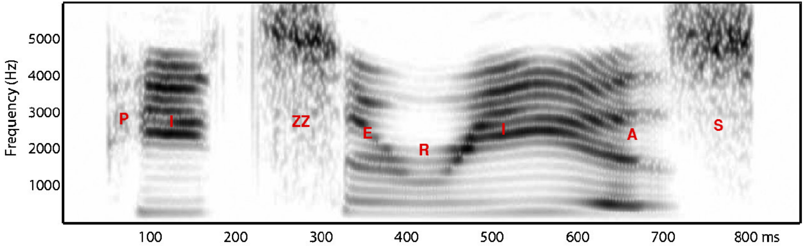
musical sound: complex, polyphonic, noisy, non-stationary. . .

Sound



musical sound: complex, polyphonic, noisy, non-stationary…

Spectrogram



(word "pizzerias" with English pronunciation – from Evan Smith's Ph.D.)

Sinusoidal Modeling – the original sin()...

[McAulay & Quatieri (IEEE Trans. ASSP 1986)]
[Serra & Smith (Computer Music Journal 1990)]

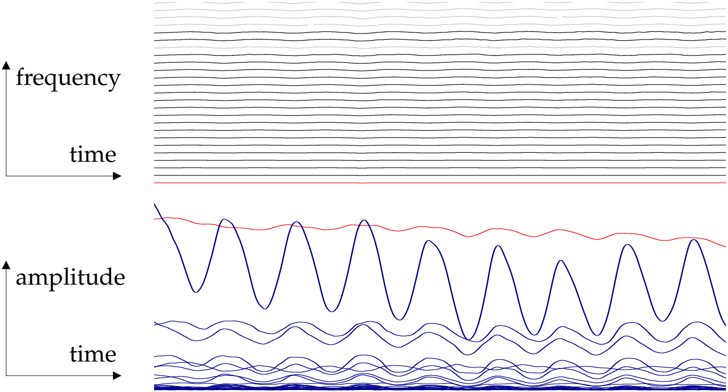
The audio signal s is given by:

$$s(t) = \sum_{p=1}^{P} a_p(t)\ \cos(\phi_p(t)) \quad \text{with} \quad \frac{d\phi_p}{dt}(t) = \omega_p(t)$$

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

The functions a_p, ω_p, and ϕ_p are the instantaneous amplitude, frequency, and phase of the p^{th} partial, respectively.

Trajectories of the Partials



Frequencies and amplitudes, as functions of time,
of the partials of an alto saxophone sound ($\approx$ 1.5 second)

Non-Stationary Case

For one partial ($P = 1$), for one frame (centered on time $t = 0$):

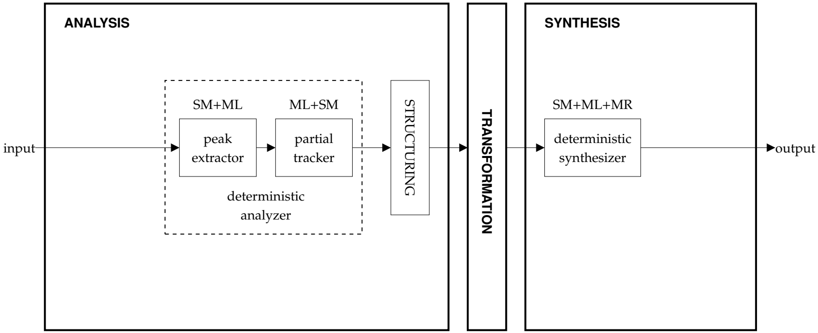
$$s(t) = \exp\left(\underbrace{(\lambda_0 + \mu_0 t)}_{\lambda(t)=\log(a(t))} + j \underbrace{\left(\phi_0 + \omega_0 t + \frac{\psi_0}{2} t^2\right)}_{\phi(t)}\right)$$

$\rightarrow$ one step further in the Taylor expansion
of the (log-)amplitude and phase parameters...

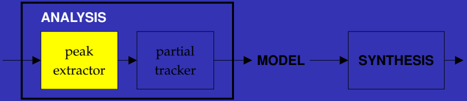
- amplitude $\exp(\lambda_0) = a_0$
- **amplitude modulation** μ_0
- phase ϕ_0
- frequency ω_0
- **frequency modulation** ψ_0

InSpect Software Architecture

- ML: Mathieu Lagrange, MR: Matthias Robine



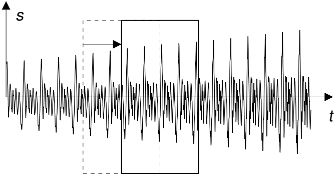
Peak Extraction (Short-Term Analysis)

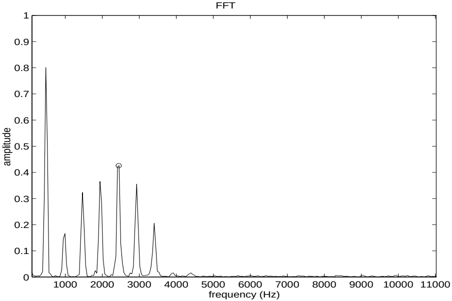


Short-Time Fourier Transform (STFT)

$$S_w(t,\omega) = \int_{-\infty}^{+\infty} s(\tau) w(\tau - t) \exp\left(-j\omega(\tau - t)\right)\, d\tau$$

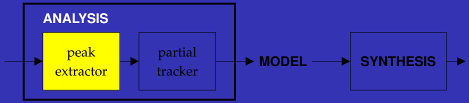
$$s(t) = a_0 e^{j\phi_0} \exp\left(\mu_0 t + j\left(\omega_0 t + \frac{\psi_0}{2} t^2\right)\right)$$





using local maximum *m* of short-term magnitude spectrum

Analysis Window w (*e.g.* Hann window)



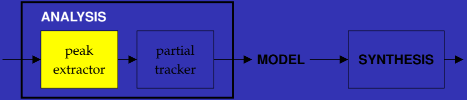
w with finite time support (for the STFT to be computable) and band-limited in frequency (peak $\leftrightarrow$ partial bijection)

$$S_w(0, \omega) = \underbrace{a_0 e^{j\phi_0}}_{s_0} \cdot \Gamma_w(\omega_0 - \omega, \mu_0, \psi_0) \quad \text{where}$$

$$\Gamma_w(\omega, \mu_0, \psi_0) = \int_{-\infty}^{+\infty} w(t) \exp\left(\mu_0 t + j\left(\omega t + \frac{\psi_0}{2} t^2\right)\right) dt$$

- S_w is measured
- a_0 and ϕ_0 are to be estimated
- Γ_w is a complex function of ω, μ, and ψ

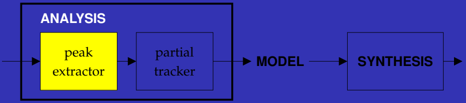
Analysis Methods (STFT-Based)



- difference method (phase vocoder)
- trigonometric estimators
  - arcsin [Marchand (2000)]
  - arccos [Lagrange, Marchand & Rault (2005)]
  - arctan [Betser, Collen, Richard & David (2006)]
- quadratic interpolation [Smith & Serra (1987)]
  generalized [Abe & Smith (2005)]
- spectral reassignment [Auger & Flandrin (1995)]
  generalized [Röbel (2002), Hainsworth (2003)]
- **derivative algorithm** [Desainte-Catherine & Marchand (2000)]
  generalized [Marchand & Depalle (2008)]

→ all these methods, except one (quadratic interpolation), are **equivalent** regarding the estimation of the **frequency**
[Marchand & Lagrange (2006, 2007)]

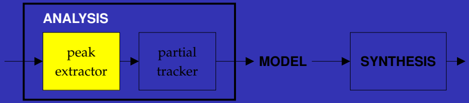
Analysis Methods (STFT-Based)



- difference method (phase vocoder)
- trigonometric estimators
  - arcsin [Marchand (2000)]
  - arccos [Lagrange, Marchand & Rault (2005)]
  - arctan [Betser, Collen, Richard & David (2006)]
- quadratic interpolation [Smith & Serra (1987)]
  generalized [Abe & Smith (2005)]
- spectral reassignment [Auger & Flandrin (1995)]
  generalized [Röbel (2002), Hainsworth (2003)]
- **derivative algorithm** [Desainte-Catherine & Marchand (2000)]
  generalized [Marchand & Depalle (2008)]

→ all these methods, except one (quadratic interpolation), are **equivalent** regarding the estimation of the **frequency**
[Marchand & Lagrange (2006, 2007)]

Derivative Algorithm



uses the derivatives of the signal

(the derivative of an exponential is an exponential...)

$$s'(t) = (\mu_0 + j(\omega_0 + \psi_0 t)) \cdot s(t)$$

$j\psi_0 t$ is an odd function $\implies$ its spectrum is real...

$$\hat{\omega}_0 = \Im\left(\frac{S'_w}{S_w}(\omega_m)\right)$$

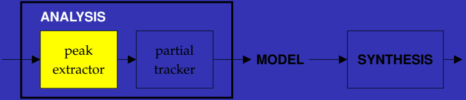
$\hat{\mu}_0$ is given by the real part, $\hat{\psi}_0$ with the second derivative

and finally $\hat{a}_0 = \left|\frac{S_w(\hat{\omega}_0)}{\Gamma_w(0, \hat{\mu}_0, \hat{\psi}_0)}\right|$ and $\hat{\phi}_0 = \angle\left(\frac{S_w(\hat{\omega}_0)}{\Gamma_w(0, \hat{\mu}_0, \hat{\psi}_0)}\right)$

practical issue: get the derivative s' from the signal s

[Marchand & Depalle (2008)]

Derivative Algorithm



uses the derivatives of the signal

(the derivative of an exponential is an exponential...)

$$s'(t) = (\mu_0 + j(\omega_0 + \psi_0 t)) \cdot s(t)$$

$j\psi_0 t$ is an odd function $\implies$ its spectrum is real...

$$\hat{\omega}_0 = \Im\left(\frac{S'_w}{S_w}(\omega_m)\right)$$

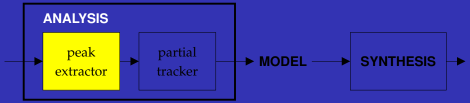
$\hat{\mu}_0$ is given by the real part, $\hat{\psi}_0$ with the second derivative

and finally $\hat{a}_0 = \left|\frac{S_w(\hat{\omega}_0)}{\Gamma_w(0, \hat{\mu}_0, \hat{\psi}_0)}\right|$ and $\hat{\phi}_0 = \angle\left(\frac{S_w(\hat{\omega}_0)}{\Gamma_w(0, \hat{\mu}_0, \hat{\psi}_0)}\right)$

practical issue: get the derivative s' from the signal s

[Marchand & Depalle (2008)]

Derivative Algorithm



uses the derivatives of the signal

(the derivative of an exponential is an exponential. . .)

$$s'(t) = (\mu_0 + j(\omega_0 + \psi_0 t)) \cdot s(t)$$

$j\psi_0 t$ is an odd function $\implies$ its spectrum is real. . .

$$\hat{\omega}_0 = \Im\left(\frac{S'_w}{S_w}(\omega_m)\right)$$

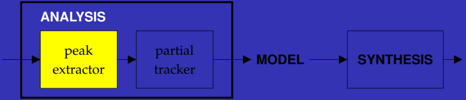
$\hat{\mu}_0$ is given by the real part, $\hat{\psi}_0$ with the second derivative

and finally $\hat{a}_0 = \left|\frac{S_w(\hat{\omega}_0)}{\Gamma_w(0, \hat{\mu}_0, \hat{\psi}_0)}\right|$ and $\hat{\phi}_0 = \angle\left(\frac{S_w(\hat{\omega}_0)}{\Gamma_w(0, \hat{\mu}_0, \hat{\psi}_0)}\right)$

practical issue: get the derivative s' from the signal s

[Marchand & Depalle (2008)]

Derivative Algorithm



uses the derivatives of the signal

(the derivative of an exponential is an exponential. . .)

$$s'(t) = (\mu_0 + j(\omega_0 + \psi_0 t)) \cdot s(t)$$

$j\psi_0 t$ is an odd function $\implies$ its spectrum is real. . .

$$\hat{\omega}_0 = \Im\left(\frac{S'_w}{S_w}(\omega_m)\right)$$

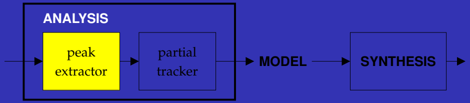
$\hat{\mu}_0$ is given by the real part, $\hat{\psi}_0$ with the second derivative

and finally $\hat{a}_0 = \left|\frac{S_w(\hat{\omega}_0)}{\Gamma_w(0, \hat{\mu}_0, \hat{\psi}_0)}\right|$ and $\hat{\phi}_0 = \angle\left(\frac{S_w(\hat{\omega}_0)}{\Gamma_w(0, \hat{\mu}_0, \hat{\psi}_0)}\right)$

practical issue: get the derivative s' from the signal s

[Marchand & Depalle (2008)]

Derivative Algorithm



uses the derivatives of the signal

(the derivative of an exponential is an exponential...)

$$s'(t) = (\mu_0 + j(\omega_0 + \psi_0 t)) \cdot s(t)$$

$j\psi_0 t$ is an odd function $\implies$ its spectrum is real...

$$\hat{\omega}_0 = \Im\left(\frac{S'_w}{S_w}(\omega_m)\right)$$

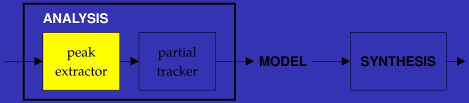
$\hat{\mu}_0$ is given by the real part, $\hat{\psi}_0$ with the second derivative

and finally $\hat{a}_0 = \left|\frac{S_w(\hat{\omega}_0)}{\Gamma_w(0, \hat{\mu}_0, \hat{\psi}_0)}\right|$ and $\hat{\phi}_0 = \angle\left(\frac{S_w(\hat{\omega}_0)}{\Gamma_w(0, \hat{\mu}_0, \hat{\psi}_0)}\right)$

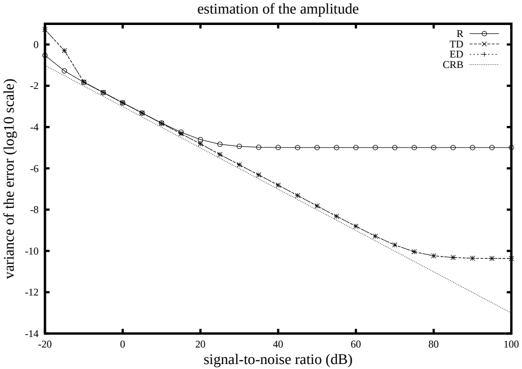
practical issue: get the derivative s' from the signal s

[Marchand & Depalle (2008)]

Performance (Precision)

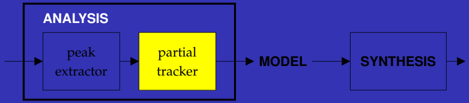


comparison to the Cramér-Rao lower Bound (CRB)
(the best performance achievable by an unbiased estimator,
in presence of Gaussian white noise of given SNR)

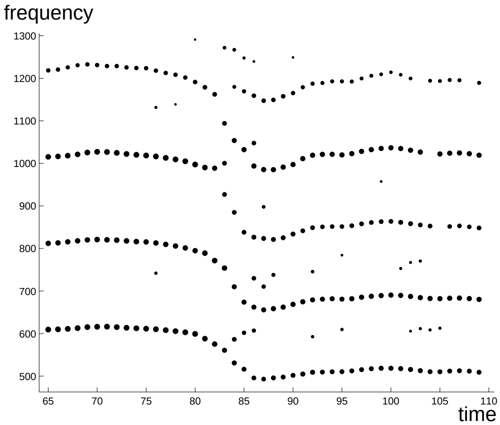


(estimation of the amplitude in the non-stationary case)

Partial Tracking (Long-Term Analysis)

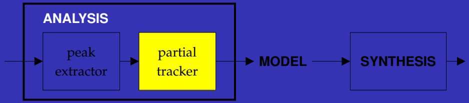


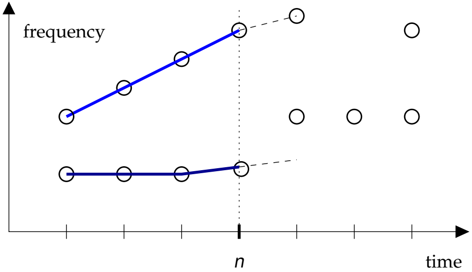
connecting the spectral peaks from frame to frame



to form the trajectories of the partials

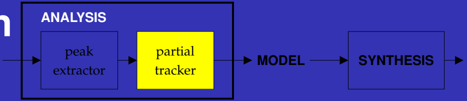
General Principles



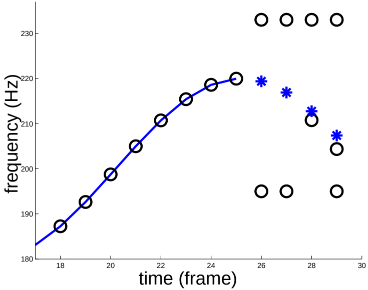


- birth / death concept, zombie state
- scheduling (lowest frequency first, highest amplitude first)
- extrapolation (constant, linear, **linear prediction**)
- connection probability (nearest frequency, **freq. content**)

Using Linear Prediction



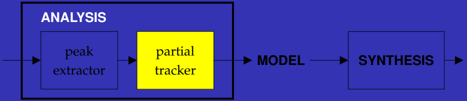
The parameters of the partials are predictable. . .



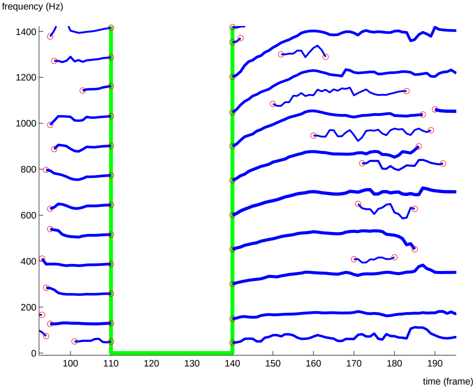
$$\hat{\mathcal{P}}(n) = \sum_{k=1}^{K} c(k)\mathcal{P}(n-k)$$

$c(k)$ coefficients found using the Burg method
[Lagrange, Marchand & Rault (2004, 2007)]

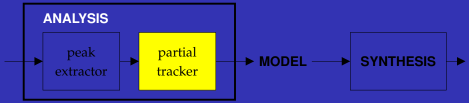
Application to Sound Restoration



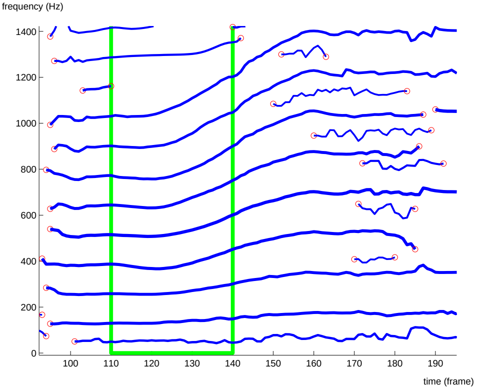
- original sound with a 650-ms gap…
- level-0 (temporal) linear prediction [Kauppinen *et al.* (2001)]
- level-1 linear prediction [Lagrange, Marchand & Rault (2005)]



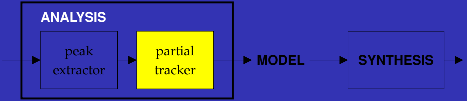
Application to Sound Restoration



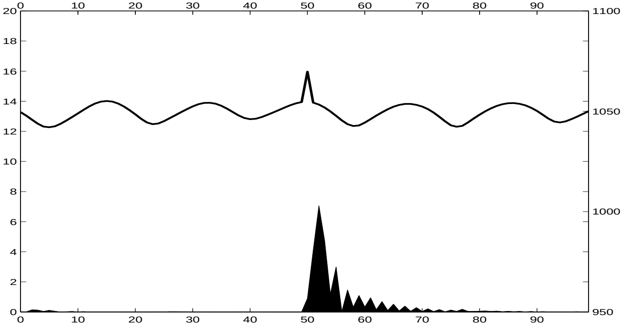
- original sound — with a 650-ms gap...
- level-0 (temporal) linear prediction — [Kauppinen *et al.* (2001)]
- level-1 linear prediction — [Lagrange, Marchand & Rault (2005)]



High-Frequency Content Control



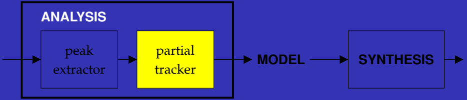
The parameters of the partials are band-limited to 20Hz. . .



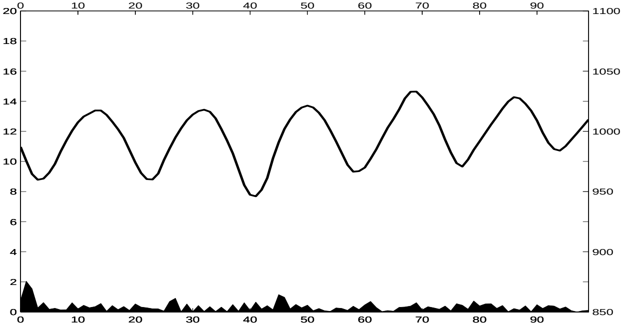
sinusoid with local error (at frame 50)

high-frequency content found using filtering
[Lagrange, Marchand & Rault (2005, 2007)]

High-Frequency Content Control



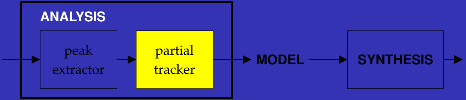
The parameters of the partials are band-limited to 20Hz. . .



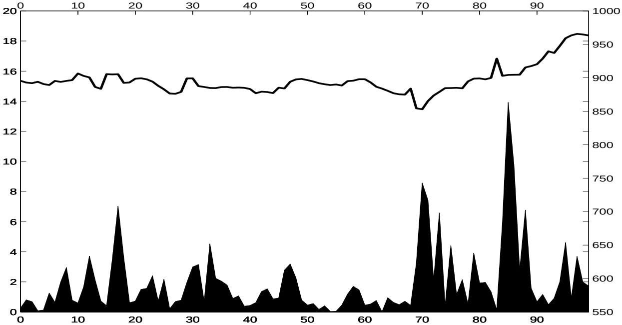
saxophone with vibrato

high-frequency content found using filtering
[Lagrange, Marchand & Rault (2005, 2007)]

High-Frequency Content Control



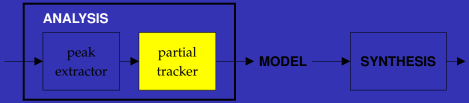
The parameters of the partials are band-limited to 20Hz. . .

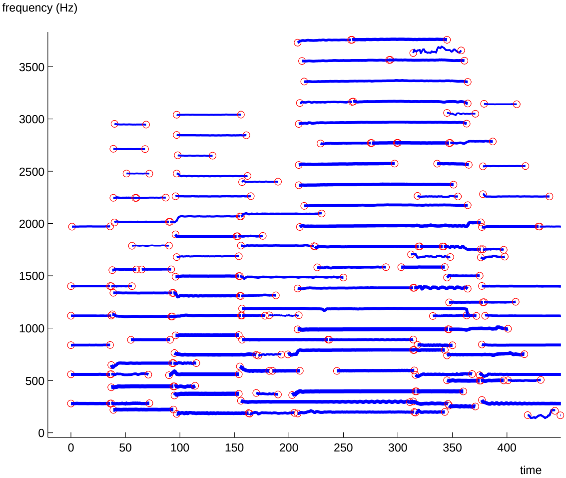


noise (tracked by error)

high-frequency content found using filtering
[Lagrange, Marchand & Rault (2005, 2007)]

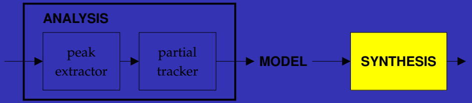
Application to Note Onset Detection





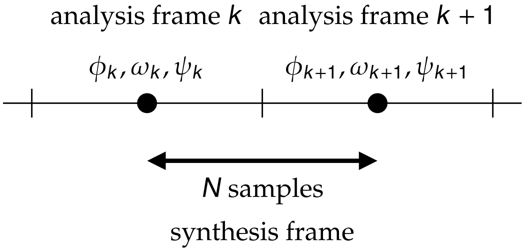
high-frequency content → enhanced onset/offset detection

Piecewise-Polynomial Parameter Models



phase model

$$\phi[n] = \sum_{d=0}^{D} c_d \, n^d$$



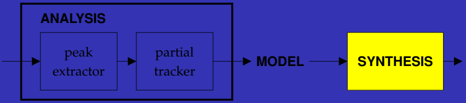
$D + 1$ constraints at the frame boundaries (here $D = 5$):

$$\begin{cases} \phi[0] & = & \phi_k \\ \phi'[0] & = & \underline{\omega}_k \\ \phi''[0] & = & \underline{\psi}_k \end{cases} \quad \text{and} \quad \begin{cases} \phi[N] & = & \phi_{k+1} + 2\pi M \\ \phi'[N] & = & \underline{\omega}_{k+1} \\ \phi''[N] & = & \underline{\psi}_{k+1} \end{cases}$$

$$\Longrightarrow \quad M = e\left[\frac{1}{2\pi}\left((\phi_k - \phi_{k+1}) + (\underline{\omega}_k + \underline{\omega}_{k+1})\frac{N}{2} + (\underline{\psi}_k - \underline{\psi}_{k+1})\frac{N^2}{40}\right)\right]$$

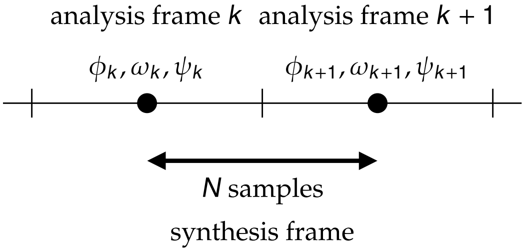
[Girin, Marchand, di Martino, Röbel & Peeters (2003)]

Piecewise-Polynomial Parameter Models



phase model

$$\phi[n] = \sum_{d=0}^{D} c_d \, n^d$$



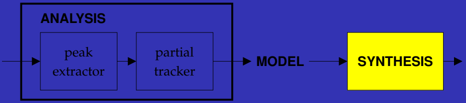
$D+1$ constraints at the frame boundaries (here $D=5$):

$$\begin{cases} \phi[0] &= \phi_k \\ \phi'[0] &= \underline{\omega}_k \\ \phi''[0] &= \underline{\psi}_k \end{cases} \quad \text{and} \quad \begin{cases} \phi[N] &= \phi_{k+1} + 2\pi M \\ \phi'[N] &= \underline{\omega}_{k+1} \\ \phi''[N] &= \underline{\psi}_{k+1} \end{cases}$$

$$\Longrightarrow M = e\left[\frac{1}{2\pi}\left((\phi_k - \phi_{k+1}) + (\underline{\omega}_k + \underline{\omega}_{k+1})\frac{N}{2} + (\underline{\psi}_k - \underline{\psi}_{k+1})\frac{N^2}{40}\right)\right]$$

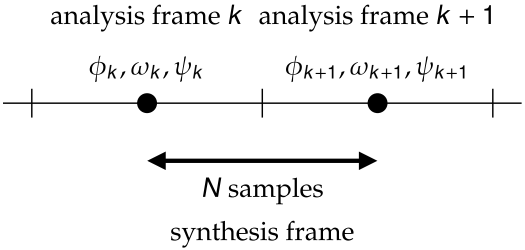
[Girin, Marchand, di Martino, Röbel & Peeters (2003)]

Piecewise-Polynomial Parameter Models



phase model

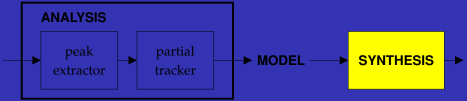
$$\phi[n] = \sum_{d=0}^{D} c_d \, n^d$$

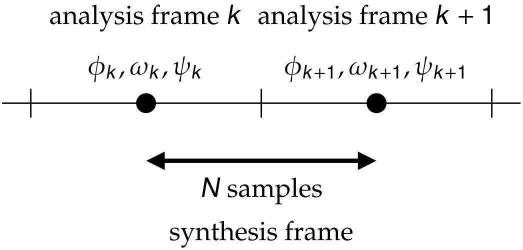


Reconstruction SNRs (synthetic examples)

model order D:	1	3	5
constant	∞	∞	∞
linear	47.19	∞	∞
vibrato only	18.93	88.84	∞
vibrato+tremolo	19.20	88.34	116.93

Piecewise-Polynomial Parameter Models





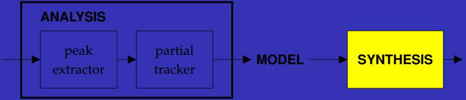
phase model

$$\phi[n] = \sum_{d=0}^{D} c_d \, n^d$$

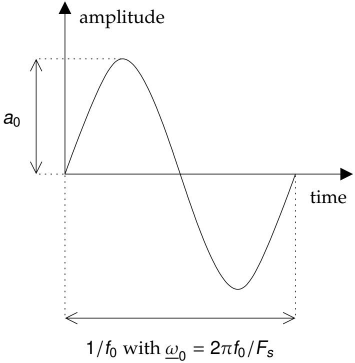
Reconstruction SNRs (natural samples)

model order D:	1	3	5
singing voice	20.14	20.39	20.42
bass	8.71	9.56	9.76
cello	16.35	16.92	17.02
violin	17.68	17.91	17.94

Software Oscillators



- constant amplitude and frequency (order-1 phase)
- recursive formulation of each partial signal



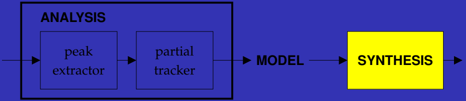
$$s[n] = a_0 \cos(\phi_0 + n\underline{\omega}_0)$$

Digital Resonator

$$\begin{cases} C & = & 2\cos(\underline{\omega}_0) \\ s[0] & = & a_0 \cos(\phi_0) \\ s[1] & = & a_0 \cos(\phi_0 + \underline{\omega}_0) \\ s[n+1] & = & C \cdot s[n] - s[n-1] \end{cases}$$

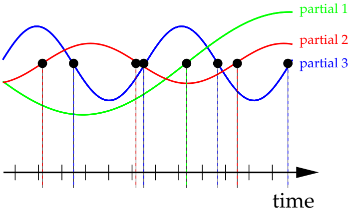
complexity: 1 $\times$, 1 $+$ (optimal. . .)

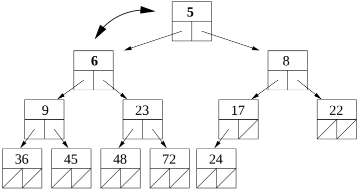
Piecewise-Polynomial Signal Model (PASS)



get the (global) signal without computing any partial signal. . .

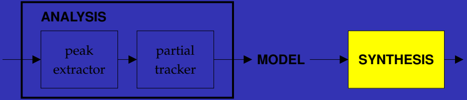
- the partial signals are described by polynomials of degree D
- the sum of polynomials of degree D is a polynomial of degree D
- evaluate this unique polynomial generator of low degree ($D = 2$)
- update frequently the polynomial coefficients of each partial. . .
  - at optimal update times (each quarter cycle)
  - using a specific data structure (optimized heap)





[Robine, Strandh & Marchand (2006)]

Performance (Speed)



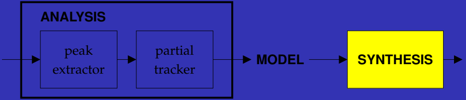
- DR complexity:
  - independent of the mean frequency $\bar{f}$
  - proportional to the sampling frequency F_s
- PASS complexity:
  - proportional to the mean frequency $\bar{f}$
  - roughly independent of the sampling frequency F_s

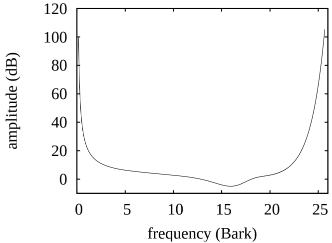
P	$\bar{f}$ (Hz)	F_s (Hz)	DR (s)	PASS (s)
2500	200	44100	3.9	**2.0**
2500	300	44100	3.9	**3.0**
2500	400	44100	**3.9**	4.0

(computation times for 10 seconds of sound)

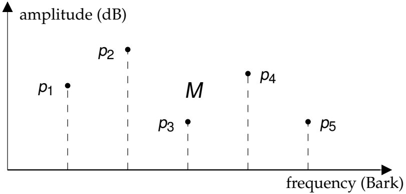
P	$\bar{f}$ (Hz)	F_s (Hz)	DR (s)	PASS (s)
4000	300	22050	**3.2**	6.6
4000	300	44100	**6.3**	6.6
4000	300	96000	13.7	**6.6**

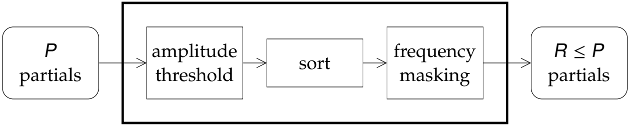
Taking Advantage of Psychoacoustics





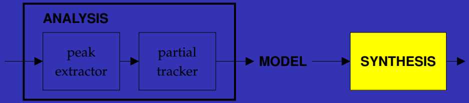
[Lagrange & Marchand (2001)]

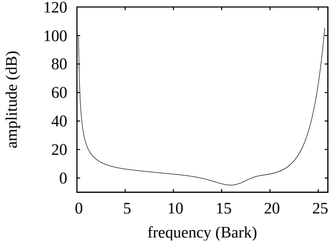




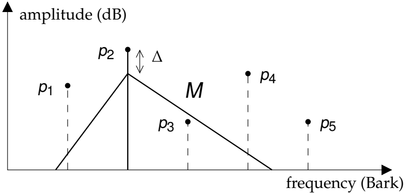
- amplitude threshold
- frequency masking (mask M)
- using a specific data structure (skip-list)

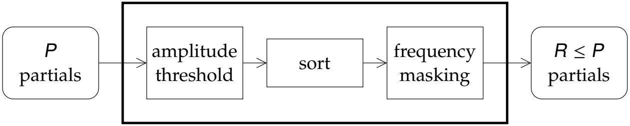
Taking Advantage of Psychoacoustics





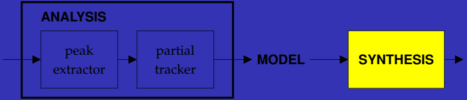
[Lagrange & Marchand (2001)]

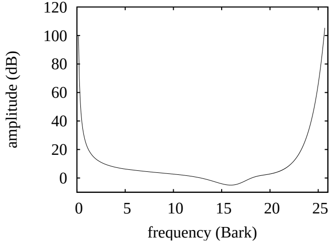




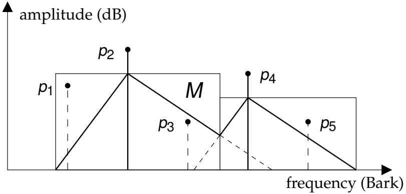
- amplitude threshold
- frequency masking (mask M)
- using a specific data structure (skip-list)

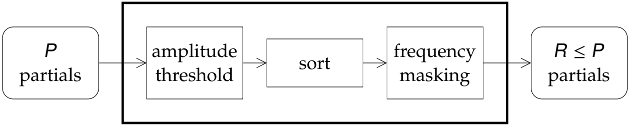
Taking Advantage of Psychoacoustics





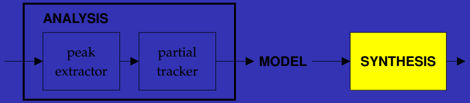
[Lagrange & Marchand (2001)]

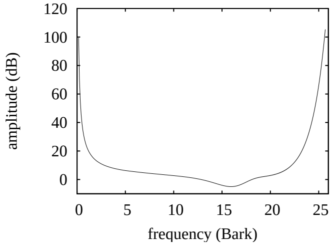




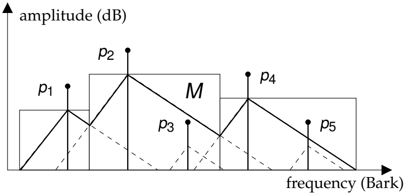
- amplitude threshold
- frequency masking (mask M)
- using a specific data structure (skip-list)

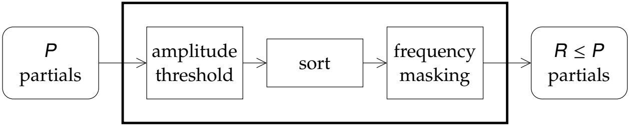
Taking Advantage of Psychoacoustics





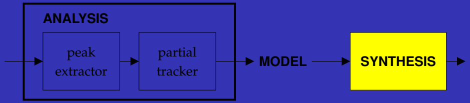
[Lagrange & Marchand (2001)]

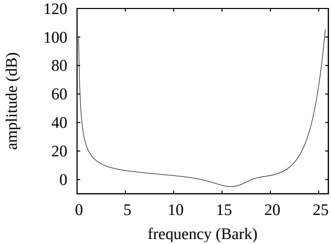




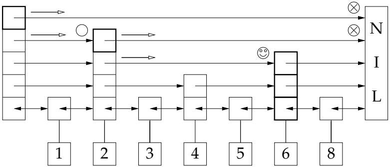
- amplitude threshold
- frequency masking (mask M)
- using a specific data structure (skip-list)

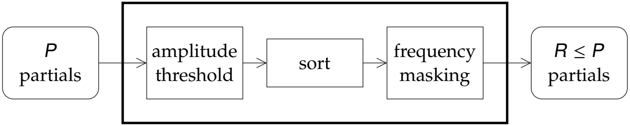
Taking Advantage of Psychoacoustics





[Lagrange & Marchand (2001)]

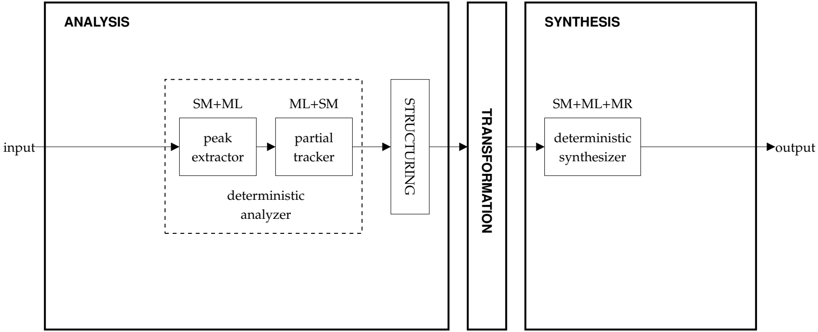




- amplitude threshold
- frequency masking (mask M)
- using a specific data structure (skip-list)

Hierarchic Modeling

- ML: Mathieu Lagrange, MR: Matthias Robine



- Martin Raspaud

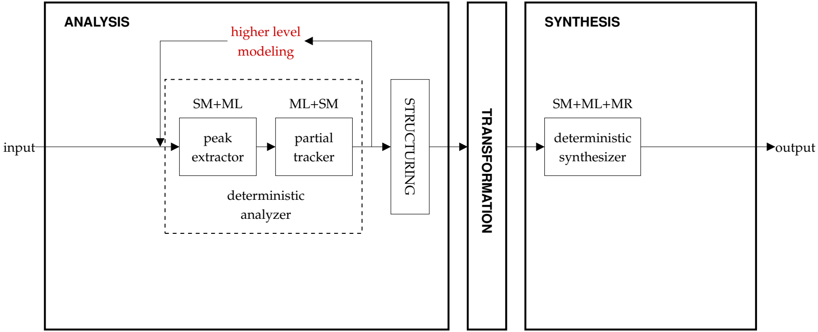
polynomials+sinusoids as model parameters

$$\mathcal{P}(t) = \Pi(t) + \sum_{p=1}^{P} a_p(t)\sin(\phi_p(t))$$

[Raspaud, Marchand & Girin (2005)]

Hierarchic Modeling

- ML: Mathieu Lagrange, MR: Matthias Robine



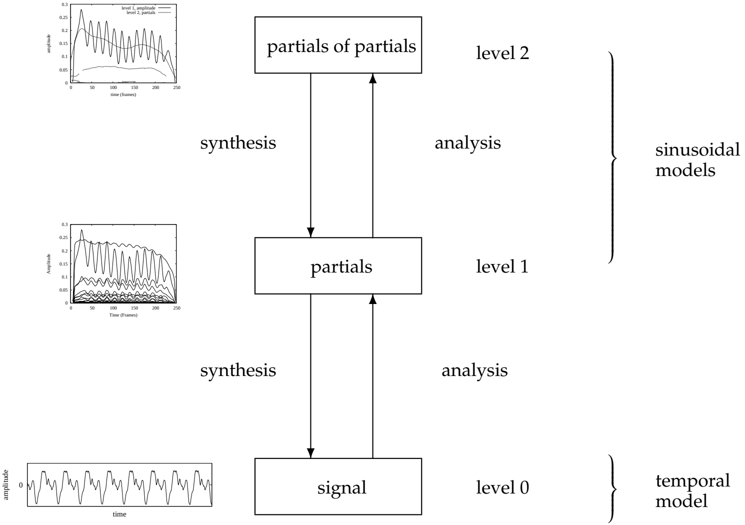
- Martin Raspaud

polynomials+sinusoids as model parameters

$$\mathcal{P}(t) = \Pi(t) + \sum_{p=1}^{P} a_p(t) \sin(\phi_p(t))$$

[Raspaud, Marchand & Girin (2005)]

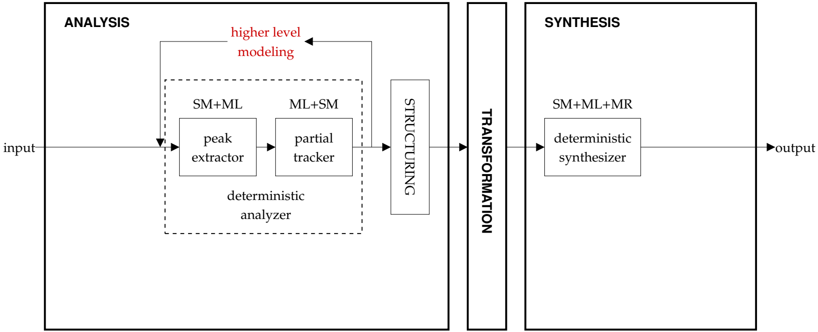
Hierarchic Sinusoidal Modeling



[Marchand and Raspaud (2004)]

Stochastic Modeling

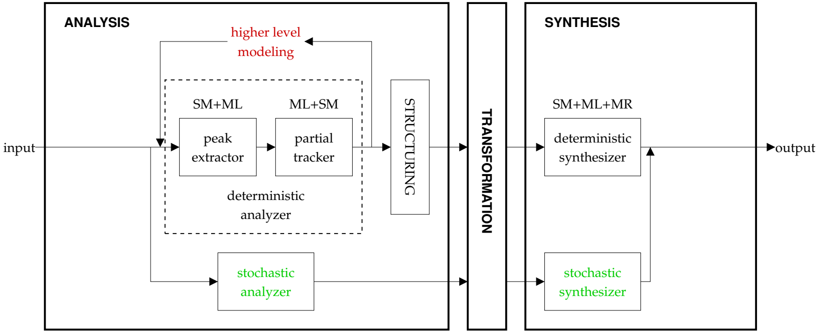
- ML: Mathieu Lagrange, MR: Matthias Robine



- Martin Raspaud
- Guillaume Meurisse

Stochastic Modeling

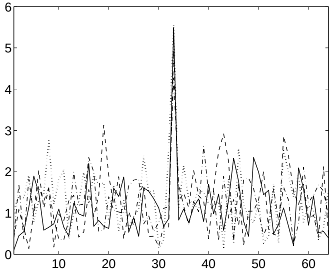
- ML: Mathieu Lagrange, MR: Matthias Robine

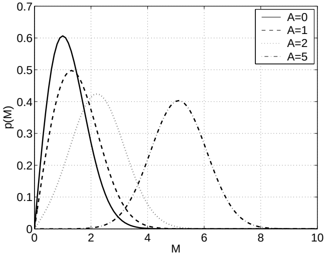


- Martin Raspaud
- Guillaume Meurisse

Towards an Unified Sinusoids+Noise Model

- distribution of the magnitude M in time (and frequency)





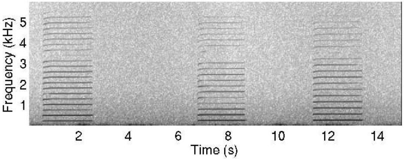
- sinusoid of amplitude A: constancy (Gaussian distribution)
- noise of deviation σ: variability (Rayleigh distribution)
- sinusoid+noise: → Rice distribution

$$p_{A,\sigma}(M) = \frac{M}{\sigma^2} \exp\left(\frac{-\left(M^2 + A^2\right)}{2\sigma^2}\right) I_0\left(\frac{AM}{\sigma^2}\right)$$

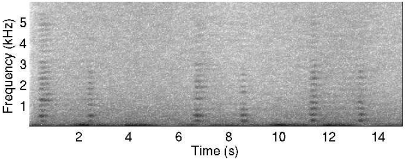
Analysis Method and Resynthesis Results

analysis (moments method):
from the measured normalized mean μ',
estimate the SNR $\gamma = \frac{A^2}{2\sigma^2}$ (then σ) by solving

$$\mu'(\gamma) = \frac{\sqrt{\pi}}{2\sqrt{1+\gamma}}\left[(1+\gamma)I_{e_0}\left(\frac{\gamma}{2}\right) + \gamma\, I_{e_1}\left(\frac{\gamma}{2}\right)\right]$$



original (saxophone+wind)

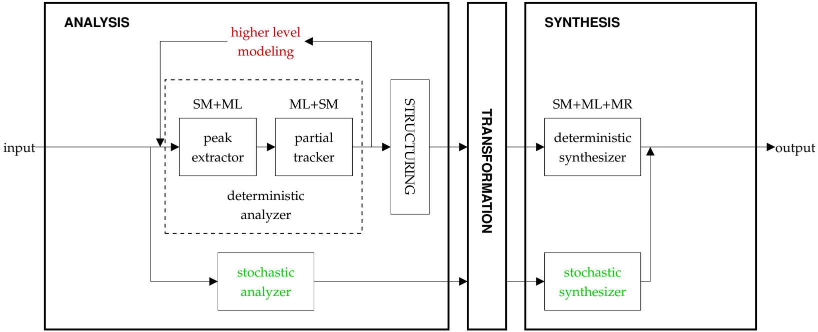


resynthesis (σ only)

[Meurisse, Hanna & Marchand (2006)]

Spatial Modeling

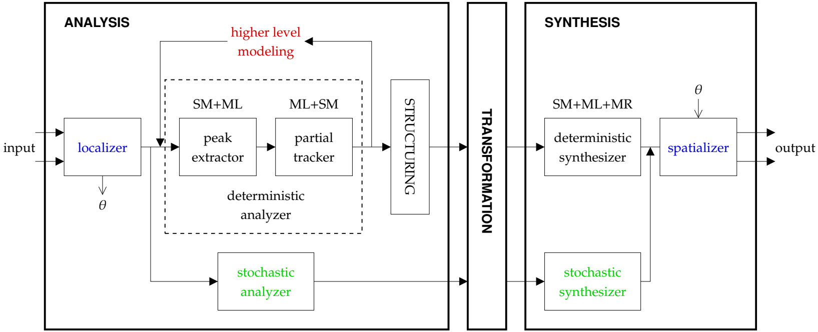
- ML: Mathieu Lagrange, MR: Matthias Robine



- Martin Raspaud
- Guillaume Meurisse
- Joan Mouba

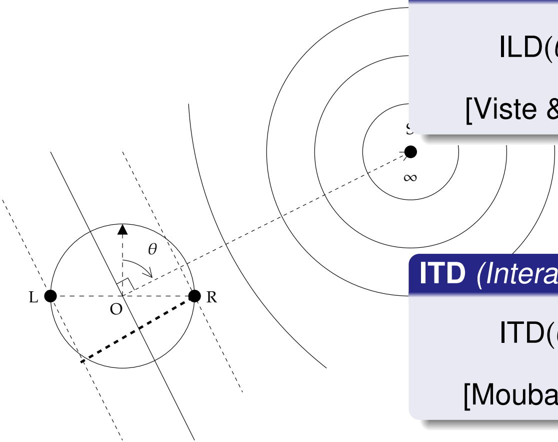
Spatial Modeling

- ML: Mathieu Lagrange, MR: Matthias Robine



- Martin Raspaud
- Guillaume Meurisse
- Joan Mouba

Sound Propagation and Head Model



ILD *(Interaural Level Difference)*

$$\mathrm{ILD}(\theta, f) = \alpha(f)\frac{1}{c}\sin(\theta)$$

[Viste & Evangelista (2003)]

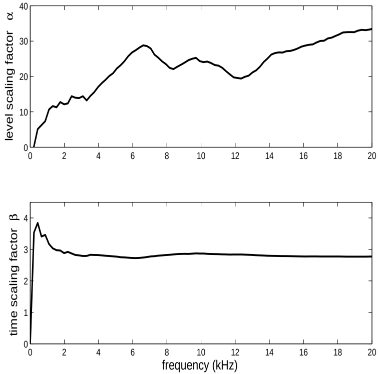
ITD *(Interaural Time Difference)*

$$\mathrm{ITD}(\theta, f) = \beta(f)\frac{r}{c}\sin(\theta)$$

[Mouba & Marchand (2008)]

r: head radius, c: sound celerity

Sound Propagation and Head Model



ILD *(Interaural Level Difference)*

$$\mathrm{ILD}(\theta, f) = \alpha(f)\frac{1}{c}\sin(\theta)$$

[Viste & Evangelista (2003)]

ITD *(Interaural Time Difference)*

$$\mathrm{ITD}(\theta, f) = \beta(f)\frac{r}{c}\sin(\theta)$$

[Mouba & Marchand (2008)]

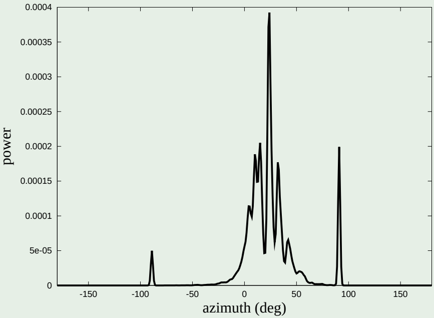
r: head radius, c: sound celerity

Localization Algorithm

Duplex Theory [Lord Rayleigh (1907)]:

- ITD prominent at low frequencies (less absorbed),
- ILD crucial for high frequencies (phase ambiguity).

histogram obtained with a real source positioned at $\theta = 30^\circ$

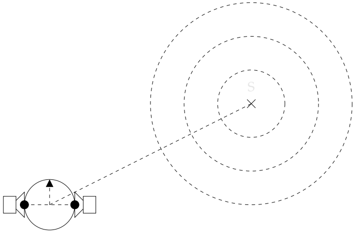


Spatialization Algorithm

- binaural spatialization

$$S_L(f) = H_L(f) \cdot S(f) \text{ where } H_L(f) = 10^{+\mathrm{ILD}(\theta,f)/2} e^{+jf\mathrm{ITD}(\theta,f)/2}$$
$$S_R(f) = H_R(f) \cdot S(f) \text{ where } H_R(f) = 10^{-\mathrm{ILD}(\theta,f)/2} e^{-jf\mathrm{ITD}(\theta,f)/2}$$

- transaural multi-diffusion



$$S_1 = -S \cdot (H_R \cdot H_{R1} - H_L \cdot H_{R2})/\Delta$$
$$S_2 = -S \cdot (H_L \cdot H_{L2} - H_R \cdot H_{L1})/\Delta$$
$$\text{where } \Delta = H_{L1} \cdot H_{R2} - H_{L2} \cdot H_{R1}$$

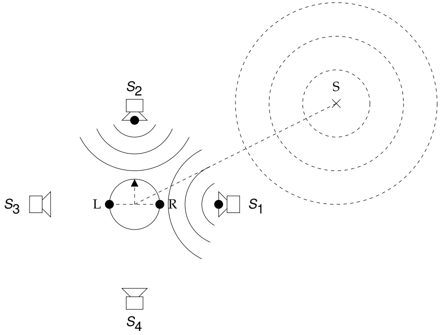
[Mouba & Marchand (2008)]

Spatialization Algorithm

- binaural spatialization

$$S_L(f) = H_L(f) \cdot S(f) \text{ where } H_L(f) = 10^{+\mathrm{ILD}(\theta,f)/2} e^{+jf\mathrm{ITD}(\theta,f)/2}$$
$$S_R(f) = H_R(f) \cdot S(f) \text{ where } H_R(f) = 10^{-\mathrm{ILD}(\theta,f)/2} e^{-jf\mathrm{ITD}(\theta,f)/2}$$

- transaural multi-diffusion



$$S_1 = -S \cdot (H_R \cdot H_{R1} - H_L \cdot H_{R2})/\Delta$$
$$S_2 = -S \cdot (H_L \cdot H_{L2} - H_R \cdot H_{L1})/\Delta$$
$$\text{where } \Delta = H_{L1} \cdot H_{R2} - H_{L2} \cdot H_{R1}$$

[Mouba & Marchand (2008)]

Conclusions and Perspectives

- modeling → transients
  - sinusoidal
  - hierarchic
  - stochastic
  - spatial → elevation angle
- analysis
  - short-term
    - derivative algorithm → optimization
    - equivalence of phase-based methods
  - long-term → evaluation
    - using linear prediction
    - frequency content control
- synthesis → hybrid method
  - polynomial approach
  - software oscillators
  - taking advantage of psychoacoustics

Conclusions and Perspectives

- modeling → transients
  - sinusoidal
  - hierarchic
  - stochastic
  - spatial → elevation angle
- analysis
  - short-term
    - derivative algorithm → optimization
    - equivalence of phase-based methods
  - long-term → evaluation
    - using linear prediction
    - frequency content control
- synthesis → hybrid method
  - polynomial approach
  - software oscillators
  - taking advantage of psychoacoustics

More (Mid-Term) Perspectives. . .

- separation of sound entities (structuring)
  - semi-blind (using perceptive cues)
    - common onset
    - correlated evolutions
    - spectral structure (harmonic sources)
    - spatial location
  - informed
    - using side information
    - using watermarking
- **active listening**
  - loudness
  - pitch
  - timbre
  - time scale
  - **spatial location**

More (Mid-Term) Perspectives. . .

- separation of sound entities (structuring)
  - semi-blind (using perceptive cues)
    - common onset
    - correlated evolutions
    - spectral structure (harmonic sources)
    - spatial location
  - informed
    - using side information
    - using watermarking
- **active listening**
  - loudness
  - pitch
  - timbre
  - time scale
  - **spatial location**

More (Mid-Term) Perspectives. . .

- separation of sound entities (structuring)
  - semi-blind (using perceptive cues)
    - common onset
    - correlated evolutions
    - spectral structure (harmonic sources)
    - spatial location
  - informed
    - using side information
    - using watermarking
- **active listening**
  - loudness
  - pitch
  - timbre
  - time scale
  - **spatial location**

More (Mid-Term) Perspectives. . .

- separation of sound entities (structuring)
  - semi-blind (using perceptive cues)
    - common onset
    - correlated evolutions
    - spectral structure (harmonic sources)
    - spatial location
  - informed
    - using side information
    - using watermarking
- **active listening**
  - loudness
  - pitch
  - timbre
  - time scale
  - **spatial location**

More (Mid-Term) Perspectives. . .

- separation of sound entities (structuring)
  - semi-blind (using perceptive cues)
    - common onset
    - correlated evolutions
    - spectral structure (harmonic sources)
    - spatial location
  - informed
    - using side information
    - using watermarking
- **active listening**
  - loudness
  - pitch
  - timbre
  - time scale
  - **spatial location**

More (Mid-Term) Perspectives...

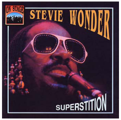
- separation of sound entities (structuring)
  - semi-blind (using perceptive cues)
    - common onset
    - correlated evolutions
    - spectral structure (harmonic sources)
    - spatial location
  - informed
    - using side information
    - using watermarking
- **active listening**
  - loudness
  - pitch
  - timbre
  - time scale
  - **spatial location**

More (Mid-Term) Perspectives. . .

- separation of sound entities (structuring)
  - semi-blind (using perceptive cues)
    - common onset
    - correlated evolutions
    - spectral structure (harmonic sources)
    - spatial location
  - informed
    - using side information
    - using watermarking
- **active listening**
  - loudness
  - pitch
  - timbre
  - time scale
  - spatial location

More (Mid-Term) Perspectives. . .

- separation of sound entities (structuring)
  - semi-blind (using perceptive cues)
    - common onset
    - correlated evolutions
    - spectral structure (harmonic sources)
    - spatial location
  - informed
    - using side information
    - using watermarking
- **active listening**
  - loudness
  - pitch
  - timbre
  - time scale
  - **spatial location**



More (Mid-Term) Perspectives. . .

- separation of sound entities (structuring)
  - semi-blind (using perceptive cues)
    - common onset
    - correlated evolutions
    - spectral structure (harmonic sources)
    - spatial location
  - informed
    - using side information
    - using watermarking
- **active listening**
  - loudness
  - pitch
  - timbre
  - time scale
  - **spatial location**