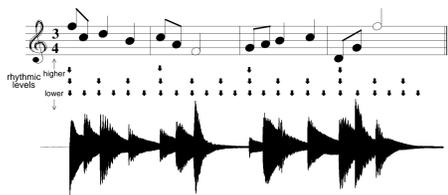


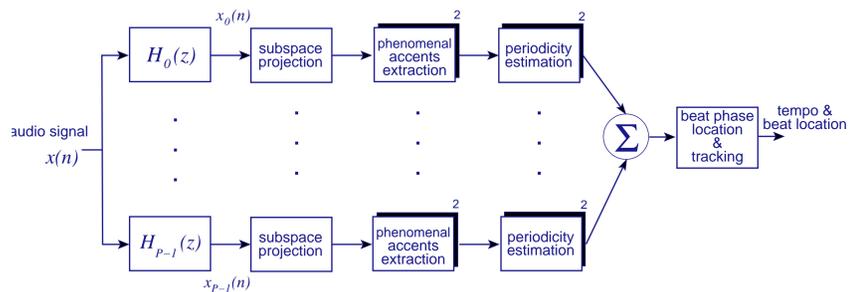
## Introduction

- ★ Actuellement, la plupart des méthodes automatiques pour l'estimation du rythme musical ne fonctionnent que pour certains types de musique possédant des sons très percussifs.
- ★ De nombreuses applications nécessitent l'estimation du rythme : indexation des signaux musicaux, transcription automatique, effets spéciaux, *music information retrieval*.
- ★ Le système proposé vise à trouver le rythme du signal musical, pour tous les genres musicaux en général. Par exemple :



- ★ Ce système effectue d'abord la séparation du signal en une *partie déterministe* et une *partie stochastique*, puis la *détection d'attaques* et finalement l'*estimation de la périodicité*.

- ★ Architecture du système proposé :



## Décomposition en sous-espaces

- ★ D'abord, le signal audio  $x(n)$  est décomposé en sous-bandes  $x_p(n)$  où  $p \in [0, \dots, P-1]$ . Ceci à l'aide d'un banc de filtres en cosinus modulés avec 80dB d'atténuation.

- ★ Dans chaque sous-bande, le signal  $x_p(n)$  est modélisé par une somme de  $M$  sinusoides exponentiellement amorties, où  $w_p(n)$  est un bruit additif

$$x_p(n) = \sum_{i=1}^{M_p} a_i e^{d_i n} \cos(2\pi f_i n + \phi_i) + w_i(n).$$

- ★ Où  $M_p$  est l'ordre du modèle pour la sous-bande  $p$ , les  $a_i \in \mathbb{R}_+^*$  sont les amplitudes, les  $d_i$  sont les coefficients d'amortissement,  $f_i \in [-\frac{1}{2}, \frac{1}{2}]$  sont les fréquences,  $\phi_i \in [-\pi, \pi]$  les phases initiales et  $w_p(n)$  et un bruit additif.

- ★ Soit  $\mathbf{H}_p$  la matrice des données de type Hankel

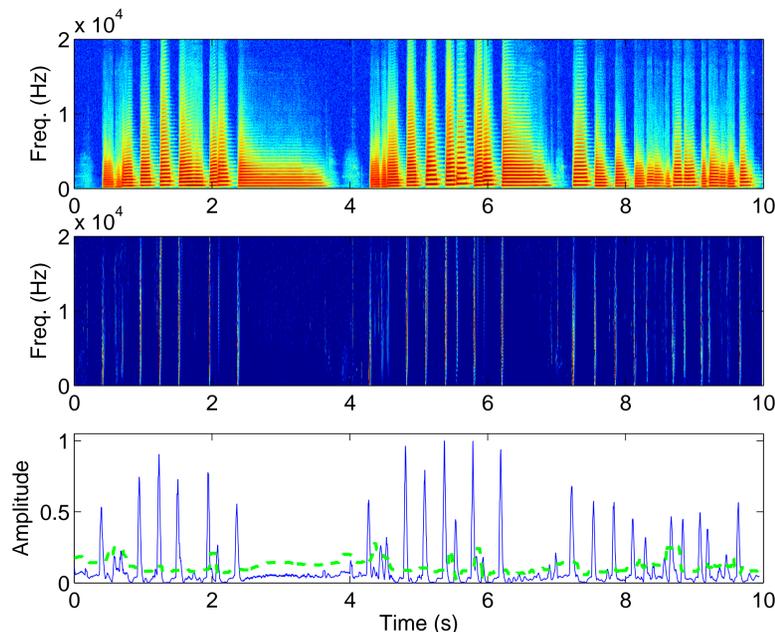
$$\mathbf{H}_p = \begin{bmatrix} x_p(0) & x_p(1) & \dots & x_p(\mathcal{L}-1) \\ x_p(1) & x_p(2) & \dots & x_p(\mathcal{L}) \\ \vdots & \vdots & \ddots & \vdots \\ x_p(\mathcal{L}-1) & x_p(\mathcal{L}) & \dots & x_p(\mathcal{L}-1) \end{bmatrix}.$$

## Décomposition en sous-espaces (suite)

- ★ En décomposant  $\mathbf{H}_p$  en valeurs singulières on obtient  $\mathbf{H}_p = \mathbf{U}_p \mathbf{A}_p \mathbf{U}_p^H$ .
- ★ On calcule la **partie déterministe** du signal en faisant une projection sur le sous-espace signal  $\mathbf{s}_p = \mathbf{U}_p^S \mathbf{U}_p^{S^H} \mathbf{x}_p$ , où  $\mathbf{U}_p^S$  est la matrice formée à partir des  $M_p$  colonnes de  $\mathbf{U}_p$  correspondant aux plus grandes valeurs singulières. Pour obtenir la **partie stochastique** il n'y a pas besoin d'extraire les sinusoides, mais il est nécessaire de faire une projection sur le sous-espace bruit :  $\mathbf{w}_p = \mathbf{x}_p - \mathbf{s}_p$
- ★ La base du sous-espace signal  $\mathbf{U}_p^S$  est calculée de manière itérative.

## Détection d'attaques

- ★ Pour la détection d'attaques, on propose une version améliorée de l'algorithme appelé **Flux Énergétique Spectral (FES)** ou aussi différence spectrale. La détection d'attaques a lieu dans la partie déterministe du signal ainsi que sur la partie stochastique.
- ★ Cette technique repose sur l'idée qu'une attaque est toujours accompagnée d'une variation dans le contenu fréquentiel du signal. Alors, en principe il suffit de dériver le contenu fréquentiel du signal par rapport au temps.
- ★ Soit  $\tilde{S}_p(m, k) = \sum_{n=-\infty}^{\infty} g(Mm-n) s_p(n) e^{-j\frac{2\pi}{N}kn}$  la représentation temps-fréquence de la partie signal  $s_p(n)$ . Par analogie soit  $\tilde{W}_p(m, k)$  la représentation de la partie bruit  $w_p(n)$ . Et  $g(n)$  est une fenêtre glissante de durée finie.
- ★ Le FES est défini comme :  $E_p(l, k) = \sum_m h(l-m) \mathcal{T}\{\tilde{S}_p(m, k)\}$ , où  $h(m)$  est une approximation à un différentiateur idéal  $H(e^{j2\pi f}) \simeq j2\pi f$  et  $\mathcal{T}\{\cdot\}$  est une **transformation psychoacoustique** pertinente qui consiste à faire un filtrage passe-bas et une compression de la dynamique du signal.
- ★ Cette approche utilise un **filtre différentiateur très performant**. Les taux de changement sont intégrés en fréquence et seuillés à l'aide d'un filtre d'ordre pour obtenir la **fonction de détection**  $v_p(m)$ .
- ★ Dans l'image ci-dessous du haut vers le bas : spectrogramme correspondant à un signal de trompette, FES et la fonction de détection.



## Estimation de la périodicité

- ★ La sortie  $v_p(m)$  du module de détection est un peigne d'impulsions. Le but consiste à estimer la périodicité des pics. On propose **trois méthodes** pour l'estimation de la périodicité. Deux techniques fréquentielles : la **somme spectrale** et le **produit spectral** et une technique temporelle : l'**autocorrélation**.

- **Techniques spectrales** : le spectre de  $v_p(m)$  est compressé par un facteur  $k$ , puis les spectres obtenus sont ajoutés ou multipliés. La fréquence fondamentale se trouve ainsi considérablement renforcée

$$S(e^{j2\pi f}) = \sum_{k=1}^K |V_p(e^{j2\pi k f})|^2 \quad \text{et} \quad \mathcal{P}(e^{j2\pi f}) = \prod_{k=1}^K |V_p(e^{j2\pi k f})|^2 \quad \text{pour } f < \frac{1}{2M}.$$

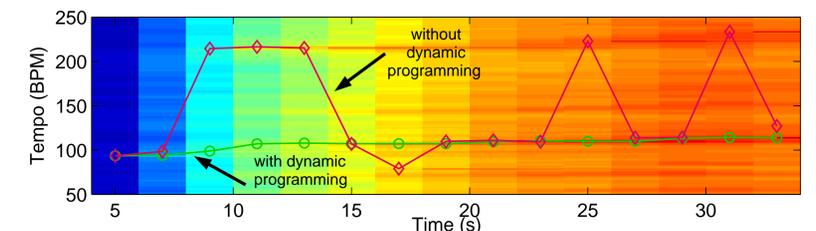
- **Technique temporelle** : l'autocorrélation  $r_p(k)$  de la fonction de détection  $v_p(m)$  est définie par

$$r_p(\tau) = \sum_m v_p(m+\tau) v_p(m).$$

- ★ On calcule la somme de toutes les fonctions d'estimation de la périodicité  $\sum_p S_p$ ,  $\sum_p \mathcal{P}_p$  et  $\sum_p r_p$ , le maximum indique le tempo  $\mathbb{T}$  du signal musical traité.

## Localisation de la phase du beat et son tracking

- ★ Pour trouver la phase du *beat*, on calcule l'**intercorrélation** entre la somme de fonctions de détection  $v(m) = \sum_p v_p(m)$  et un **train d'impulsions** artificielles  $\llcorner\llcorner(k)$  possédant un tempo  $\mathbb{T}$ .
- ★ Le *tracking* se fait à l'aide d'un algorithme de **programmation dynamique** (PD) dans le temps qui garantit un déroulement stable du tempo.
- ★ L'algorithme de PD corrige aussi les problèmes liés au doublement du tempo. On fixe au préalable l'étendue entre 80 BPM  $\rightsquigarrow$  160 BPM comme la plage de tempo préférée.



## Results

- ★ Pour l'estimation du tempo, la performance de l'algorithme a été testée sur une **base contenant 489 morceaux** annotés à la main. Ces pièces musicales couvrent une grande variété de genres.
- ★ **Taux global de succès de 95.2%**. En plus, l'algorithme est robuste au bruit perturbant les signaux d'entrée, avec 5dB de RSB il montre un taux global de succès de 87.1%.
- ★ La partie du *beat tracking* a seulement été testé de façon subjective.

## Conclusions

- ★ La méthode présentée atteint une haute performance pour l'estimation du *beat* musical.
- ★ On a développé un algorithme de détection d'attaques très efficace et de basse complexité.
- ★ L'étape suivante consistera à renforcer l'algorithme pour réussir face aux cas très compliqués (par exemple, les passages des cordes dans la musique classique).