



東京大学  
THE UNIVERSITY OF TOKYO

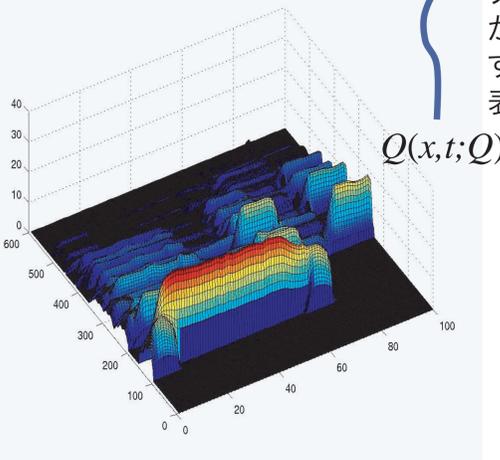
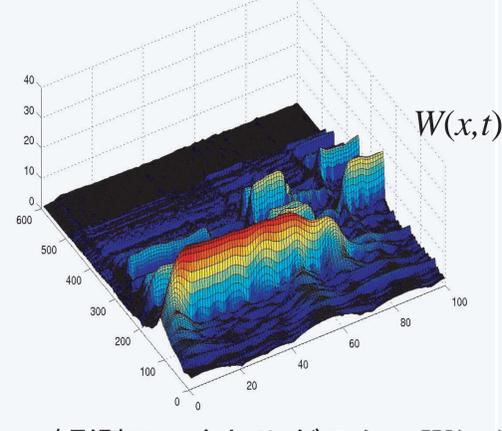
## 音楽音響信号をMIDIデータへ変換する

亀岡 弘和 齊藤 翔一郎 小野 順貴 嵯峨山 茂樹 (東京大学)

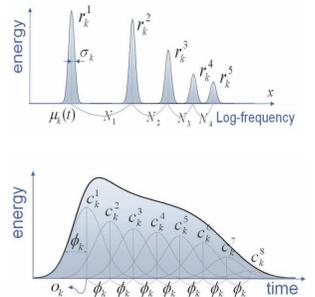
### 背景

i-Podをはじめとした大容量HDDを搭載したオーディオプレイヤーの登場により、音楽は「大量保有、選択視聴」するものへと変わりつつある。そのような流れの中で、楽曲を効率よく手に入れられる音楽版Googleのようなシステムの出現が期待されるのは当然の帰結であるが、現在一般に流通している楽曲のデータ形式は楽曲の波形を表すもので、検索に適しているとはいえない。そこで我々は、楽曲の波形データから楽譜に近い情報量を持つMIDIデータへの全自動変換機能(調波時間構造クラスタリング)と、そのステージにおいて生じた推定誤りや修正すべき箇所を用意にすばやく発見でき且つ負担なく修正できるような機能(Specmurt法)を備えたユーザ支援インターフェースを開発することにした。

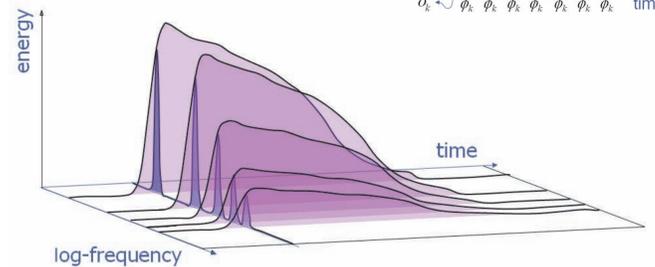
観測スペクトログラム(左図)をパラメトリックな関数(右図)でモデル化し、最適推定することが調波時間構造化クラスタリングのアイデアです。



$Q(x,t;Θ)$ は音響オブジェクトモデルと呼ぶ「調波構造+滑らかなパワーエンベロープ」をなす音源モデルの重みつき和で表されます。



音響オブジェクトモデル



調波構造化クラスタリングの原理

- ◆ 観測スペクトルグラム:  $W(x,t)$
- ◆ Kullback-Leibler尺度(分布間距離)を最小化

$$J = \iint_D W(x,t) \log \frac{W(x,t)}{Q(x,t;Θ)} dxdt$$

$$Q(x,t;Θ) = \sum_{\forall k} \sum_{\forall n} \sum_{\forall y} \frac{c_{k,n,y} r_k^n}{2\pi\sigma_k\phi_k} \exp\left(-\frac{(x-\mu_k(t)-\log n)^2}{2\sigma_k^2} - \frac{(t-o_k-y\phi_k)^2}{2\phi_k^2}\right)$$

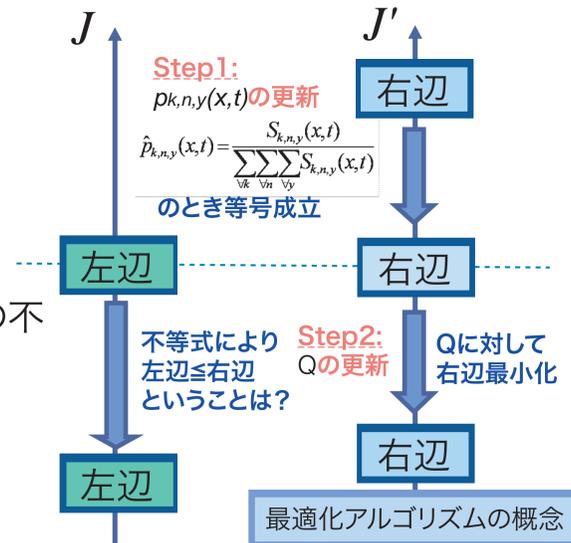
↑ 推定したいパラメータ

$$\Theta = \{\mu_k(t), o_k, w_k, \sigma_k, \phi_k, \{r_k^n\}, \{c_{k,n,y}\}\}$$

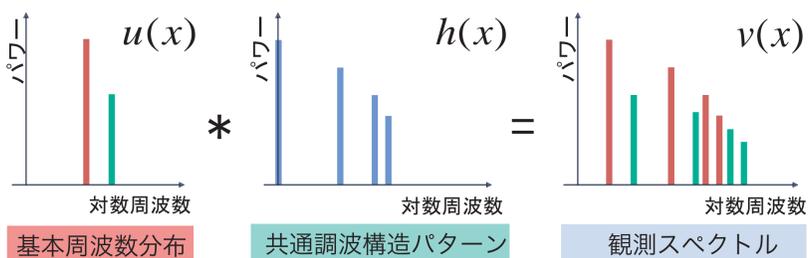
$$\equiv S_{k,n,y}(x,t)$$

- ◆ 対数関数は凹関数なので、任意の重み関数  $p_{k,n,y}(x,t)$  を用いてJensenの不等式が成立
- $$0 \leq p_{k,n,y}(x,t) \leq 1, \sum_{\forall k,n,y} p_{k,n,y}(x,t) = 1$$

$$J = \iint_D W(x,t) \log \frac{W(x,t)}{Q(x,t;Θ)} dxdt \leq \sum_{\forall k} \sum_{\forall n} \sum_{\forall y} \iint_D W(x,t) p_{k,n,y}(x,t) \log \frac{W(x,t) p_{k,n,y}(x,t)}{S_{k,n,y}(x,t;Θ)} dxdt$$



### Specmurt法の原理



$$\frac{V(y)}{H(y)} \xleftarrow{H(y)} V(y)$$

フーリエ変換 / 逆フーリエ変換

