Infinite Latent Harmonic Allocation: A Nonparametric Bayesian Approach to Multipitch Analysis

Kazuyoshi Yoshii and Masataka Goto
National Institute of Advanced Industrial Science and Technology (AIST)

I. Why Take Bayesian Approach?
We need a methodology to deal with uncertainty inherent in music analysis.

Example: F0 detection from polyphonic sounds

- Modestly confident: It is often difficult to make binary decision on the existence of each musical note.
- Hardly confident: We need a methodology to deal with uncertainty inherent in music analysis.
- Absolutely confident: It is often difficult to make binary decision on the existence of each musical note.

II. “Completely” Bayesian Treatment
Posterior distributions of all unknown variables (not limited to parameters) should be estimated.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Complexity</th>
<th>Hyper-parameters</th>
<th>Robustness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum likelihood estimation (MLE)</td>
<td>Point estimates</td>
<td>Manually specified</td>
<td>Nothing</td>
</tr>
<tr>
<td>Maximum a posteriori estimation (MAP)</td>
<td>Point estimates</td>
<td>Manually specified</td>
<td>Nothing</td>
</tr>
<tr>
<td>Classical Bayesian estimation</td>
<td>Posterior distributions</td>
<td>Manually specified</td>
<td>Nothing</td>
</tr>
<tr>
<td>(1) Nonparametric Bayesian estimation</td>
<td>Posterior distributions</td>
<td>Manually specified</td>
<td>Nothing</td>
</tr>
<tr>
<td>(2) Hierarchical Bayesian estimation</td>
<td>Posterior distributions</td>
<td>Manually specified</td>
<td>Nothing</td>
</tr>
</tbody>
</table>

The values of F0s? How many F0s? How to optimize prior distributions?

III. Conventional Parametric Models
Finite GMMS for monophonic spectra


Each Gaussian corresponds to one of harmonic partials

\[
M_k(x) = \sum_{m=1}^{r_k} \sum_{k=1}^{\tau_k} \tau_{km} \mathcal{N}(x | \mu_m, \Lambda_m^{-1})
\]

\[
\Lambda_m = \begin{cases} \log m & \text{if } \mu \text{ is fixed} \\
\mu & \text{if } \Lambda \text{ is fixed} \\
\end{cases}
\]

Nest finite GMMS for polyphonic spectra

Each GMM corresponds to a harmonic structure

\[
M_k(x) = \sum_{m=1}^{r_k} \sum_{k=1}^{\tau_k} \tau_{km} \mathcal{N}(x | \mu_m, \Lambda_m^{-1})
\]

IV. Proposed Nonparametric Models
Nested infinite GMMS for polyphonic spectra

Model complexities are considered to be infinite

\[
M_k(x) = \sum_{m=1}^{r_k} \sum_{k=1}^{\tau_k} \tau_{km} \mathcal{N}(x | \mu_m, \Lambda_m^{-1})
\]

Nonparametric means the size of parameter space is not fixed nor limited (it does not mean that there are no parameters).

V. Infinite Latent Harmonic Allocation (ILHA)

Key feature (1): Nonparametric Bayesian formulation

The Dirichlet process (DP) prior can generate the infinite number of mixing weights and lead them to become “sparse”

Prior: \( \pi_k \sim \text{Beta}(1, \alpha) \)

How to optimize influential parameter \( \alpha \)?

Key feature (2): Hierarchical Bayesian formulation

The concentration parameter (hyperparameter) of the DP is assumed to follow a noninformative Gamma hyperprior

Hyperprior: \( \alpha \sim \Gamma(a, b) \)

This naturally represents our situation that we have little knowledge on \( \alpha \).

VI. Comparative Evaluation

Data: Polyphonic audio of piano/guitar performances

6 pieces from RWC-MDB-J-2001: Jazz Music
2 pieces from RWC-MDB-C-2001: Classical Music

23 [s] excerpted from the beginning of each piece

Frequency analysis: Gabor wavelet transform

Evaluation criterion: Frame-level F-measures

The completely automated method (ILHA) yielded very competitive results against carefully tuned conventional methods (PreFest and HTC)