

Infinite Latent Harmonic Allocation:

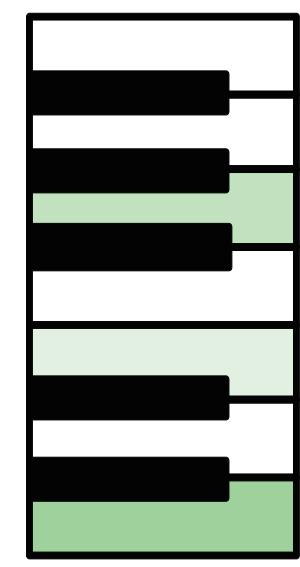
A Nonparametric Bayesian Approach to Multipitch Analysis

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

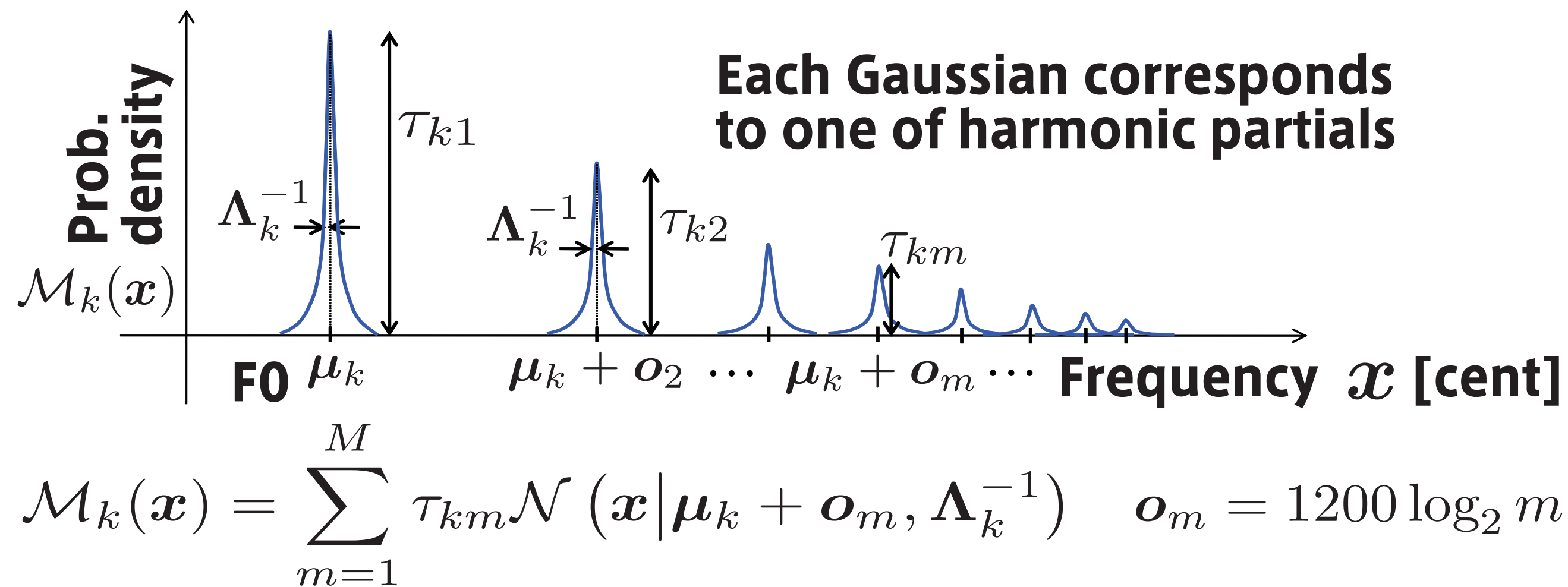


← Modestly confident
← Hardly confident
← Absolutely confident

It is often difficult to make binary decision on the existence of each musical note

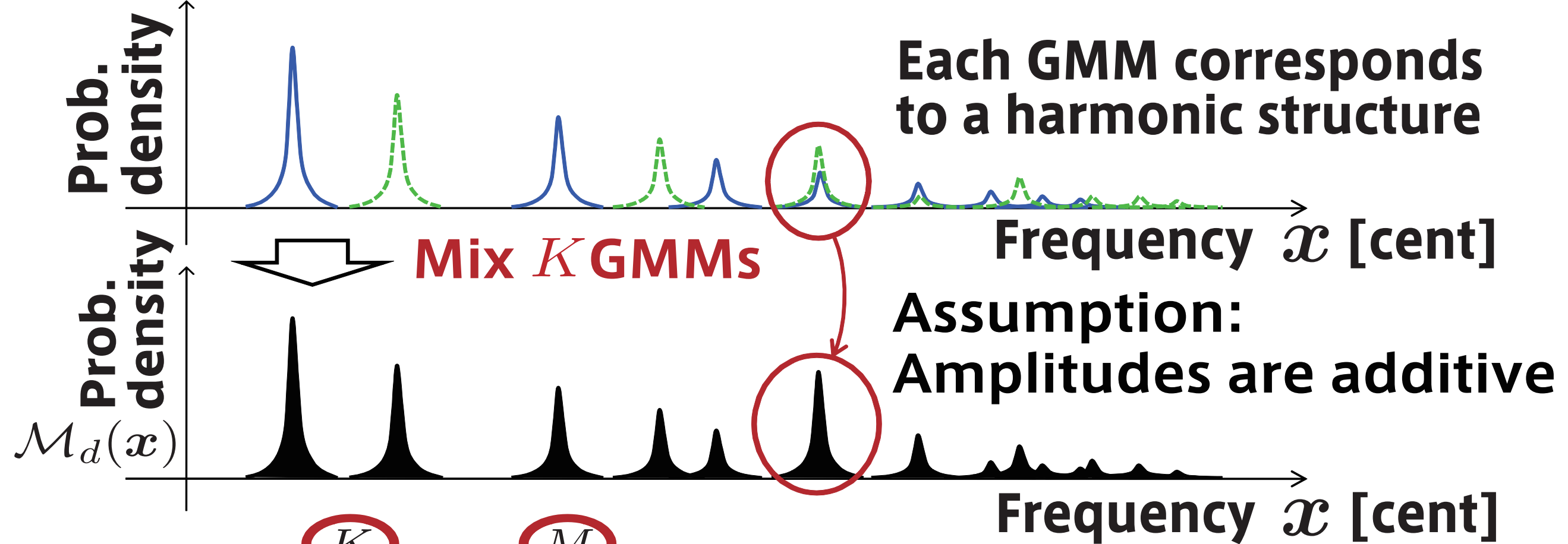
III. Conventional Parametric Models

Finite GMMs for monophonic spectra
PreFest [Goto1999] HC, HTC [Kameoka2004, 2007]



$$\mathcal{M}_k(x) = \sum_{m=1}^M \tau_{km} \mathcal{N}(x | \mu_k + o_m, \Lambda_k^{-1}) \quad o_m = 1200 \log_2 m$$

Nested finite GMMs for polyphonic spectra



$$\mathcal{M}_d(x) = \sum_{k=1}^K \pi_{dk} \sum_{m=1}^M \tau_{km} \mathcal{N}(x | \mu_k + o_m, \Lambda_k^{-1})$$

Mixing weight of sound source k in frame d

Mixing weight of harmonic partial m in sound source k

IV. Proposed Nonparametric Models

Nested infinite GMMs for polyphonic spectra
Model complexities are considered to be infinite

$$\mathcal{M}_d(x) = \sum_{k=1}^{\infty} \pi_{dk} \sum_{m=1}^{\infty} \tau_{km} \mathcal{N}(x | \mu_k + o_m, \Lambda_k^{-1})$$

Manually tuned Completely automated

“Nonparametric” means that the size of parameter space is neither fixed nor limited (it does not mean that there are no parameters)

	Conventional	Proposed
#(sound sources)	Fixed (K)	Infinite
#(harmonic partials)	Fixed (M)	Infinite
Prior on mixing weights of sound sources	Nothing	Noninformative hyperprior + Hierarchical Dirichlet process (HDP)
Prior on mixing weights of harmonic partials	Dirichlet distribution	Noninformative hyperprior + Dirichlet process (DP)
Training method	MAP estimation	Bayesian estimation

II. “Completely” Bayesian Treatment

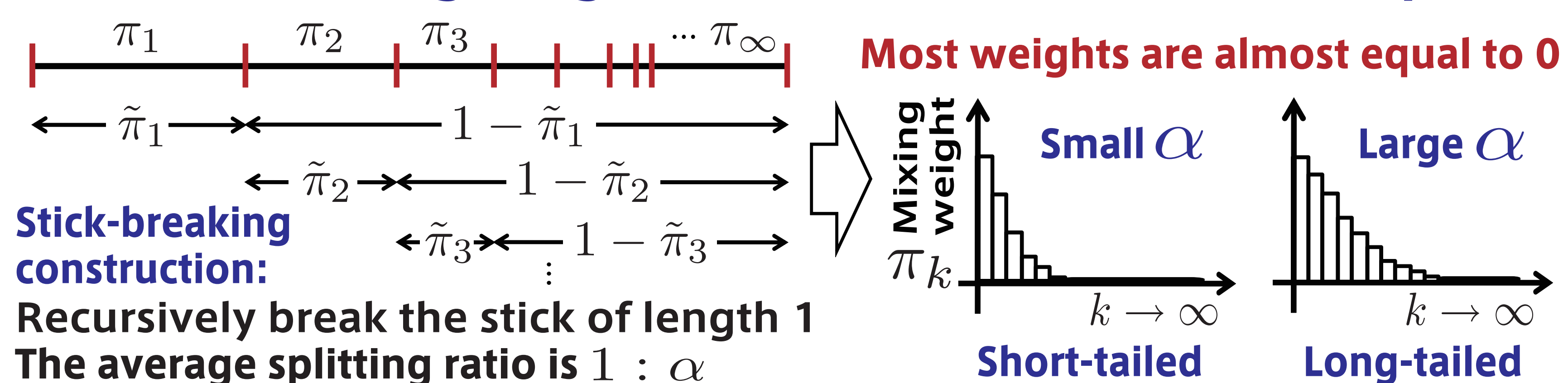
Posterior distributions of all unknown variables (not limited to parameters) should be estimated

	Parameters	Complexity	Hyper-parameters	Robustness
Maximum likelihood estimation (ML)	Point estimates	Manually specified	Nothing	Bad
Maximum a posteriori estimation (MAP)	Point estimates	Manually specified	Manually specified	Modest
Classical Bayesian estimation	Posterior distributions	Manually specified	Manually specified	Good
(1) Nonparametric Bayesian estimation	Posterior distributions	Posterior distributions	—	Excellent
(2) Hierarchical Bayesian estimation	Posterior distributions	—	Posterior distributions	Excellent

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”



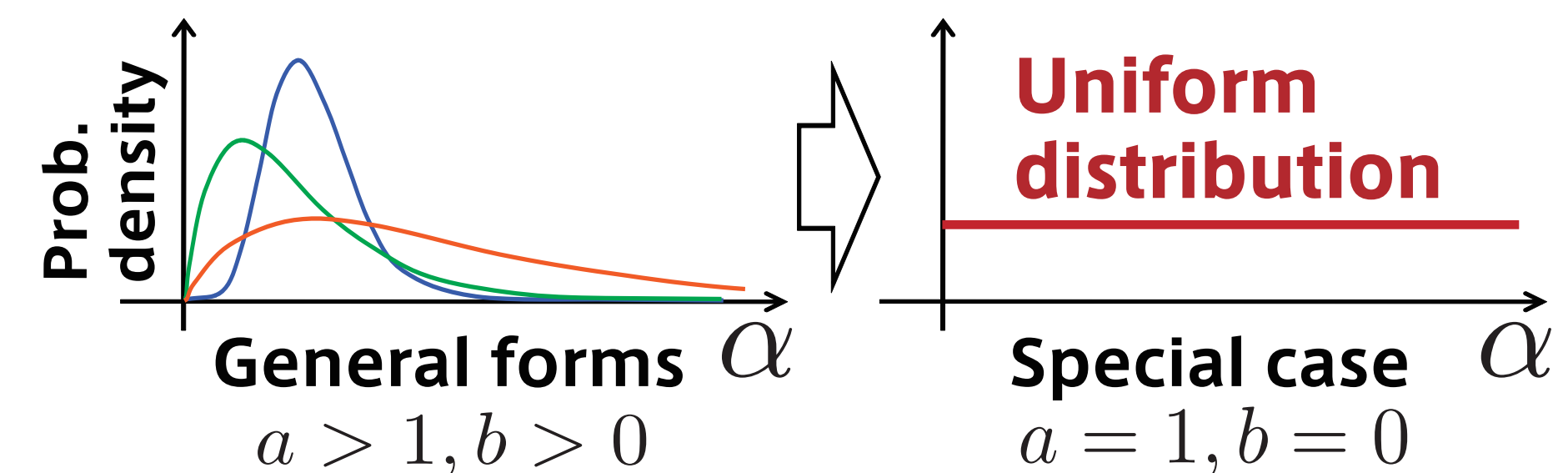
Stick-breaking construction: Recursively break the stick of length 1. The average splitting ratio is $1 : \alpha$

Prior: $\tilde{\pi}_k \sim \text{Beta}(1, \alpha) \Rightarrow$ How optimize influential hyperparameter α ?

Key feature (2): Hierarchical Bayesian formulation

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

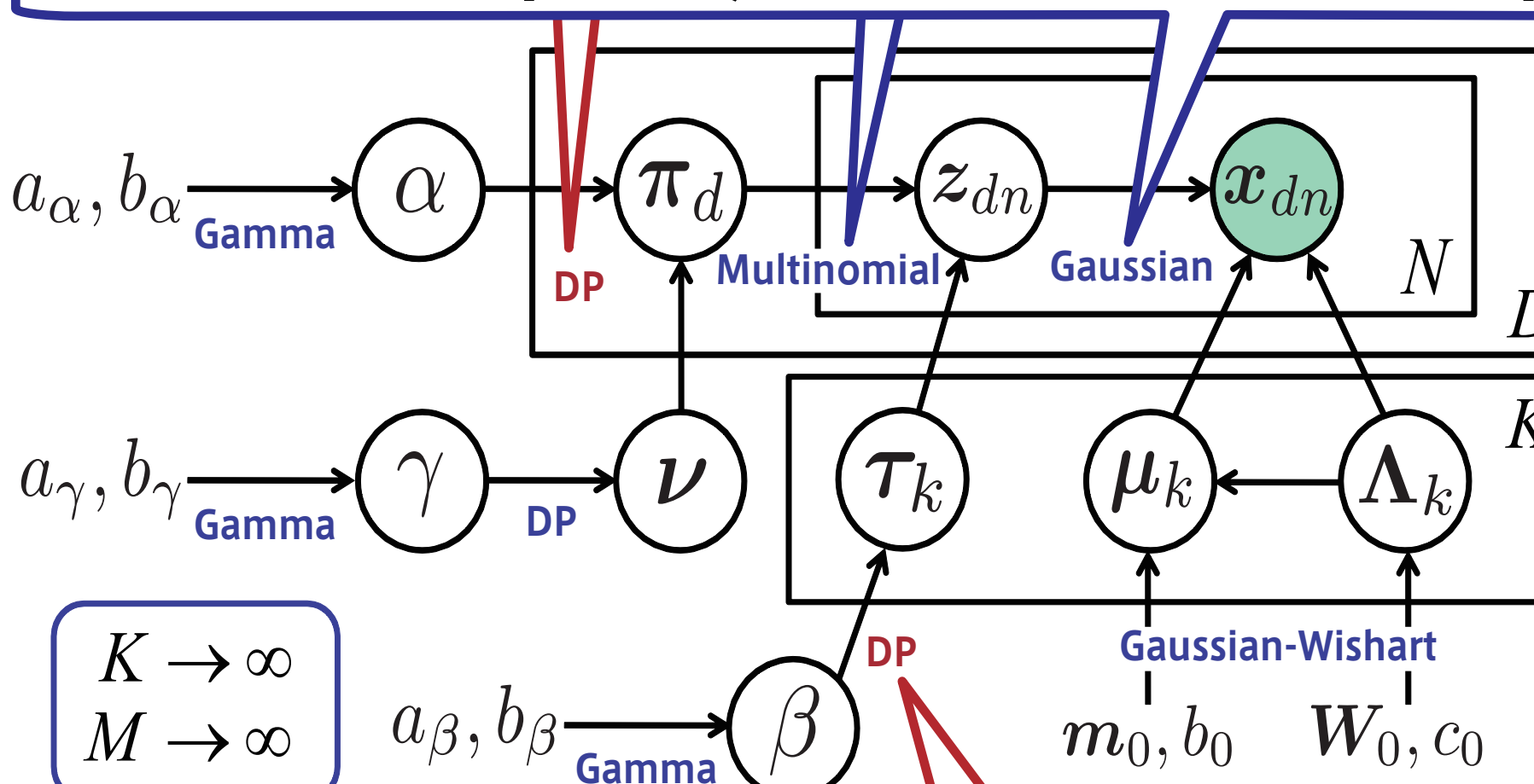
Hyperprior: $\alpha \sim \text{Gam}(a, b)$
 a, b : Shape and rate parameters



Generate the infinite number of mixing weights (probabilities) of harmonic partials

Select one of M partials in one of K sources

Generate frequency x from the selected partial



Generate the infinite number of mixing weights (probabilities) of sound sources

This naturally represents our situation that we have little knowledge on α

How train the model?
Collapsed Variational Bayes

Full joint distribution
 $p(X, Z, \pi, \tau, \mu, \Lambda, \alpha, \nu, \gamma, \beta)$
 $= p(X|Z, \mu, \Lambda) p(Z|\pi, \tau) p(\pi|\alpha, \nu) p(\nu|\gamma) p(\tau|\beta) p(\mu, \Lambda) p(\alpha) p(\gamma) p(\beta)$

Marginal distribution
 $p(X, Z, \alpha, \nu, \gamma, \beta) = p(X|Z) p(Z|\alpha, \nu, \beta) p(\nu|\gamma) p(\alpha) p(\beta)$

True posterior
 $p(Z, \alpha, \nu, \gamma, \beta | X)$

Iterative approximation via the VB-EM algorithm

Hand-tuned prior + Temporal modeling + MAP estimation Noninfo. hyperprior + Nonparametric Bayes + Bayesian estimation

Hand-tuned prior + MAP estimation Noninfo. prior + Bayesian est.

VII. Conclusion

Our contributions

We proposed an ultimate mixture-model-based method for multipitch analysis

This is the first attempt to apply the nonparametric Bayesian framework to multipitch analysis

Future directions

We plan to use this framework in a wide range of applications such as content-based clustering of musical pieces and musical structure analysis

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)

	PreFest	HTC	LHA	iLHA
J No.1	75.8	79.0	70.7	82.2
J No.2	78.5	78.0	69.1	77.9
J No.6	70.4	78.3	49.8	71.2
J No.7	83.0	86.0	70.2	85.5
J No.8	85.7	84.4	55.9	84.6
J No.9	85.9	89.5	68.9	84.7
C No.30	76.0	83.6	81.4	81.6
C No.35	72.8	76.0	58.9	79.6
Total	79.4	82.0	65.8	81.7