INFINITE-STATE SPECTRUM MODEL FOR MUSIC SIGNAL ANALYSIS

Masahiro Nakano[†], *Jonathan Le Roux*[‡], *Hirokazu Kameoka*[‡], *Nobutaka Ono*[†], *Shigeki Sagayama*[†]

[†]Graduate School of Information Science and Technology, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[‡]NTT Communication Science Laboratories, NTT Corporation,
3-1 Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan

ABSTRACT

This paper presents a nonparametric Bayesian extension of nonnegative matrix factorization (NMF) for music signal analysis. Instrument sounds often exhibit non-stationary spectral characteristics. We introduce infinite-state spectral bases into NMF to represent time-varying spectra in polyphonic music signals. We describe our extension of NMF with infinite-state spectral bases generated by the Dirichlet process in a statistical framework, derive an efficient optimization algorithm based on collapsed variational inference, and validate the framework on audio data.

Index Terms— Nonnegative matrix factorization (NMF), Dirichlet process, Collapsed variational Bayes, Nonparametric Bayes

1. INTRODUCTION

Nonnegative matrix factorization (NMF) [1] is an unsupervised decomposition technique allowing the representation of two-dimensional nonnegative data as a linear combination of a few meaningful elementary bases. In particular, NMF has been applied successfully to music spectrograms in audio signal processing, with such applications as automatic music transcription or sound source separation. NMF is able to project all signals that have a similar spectral shape on a single basis, allowing one to represent a variety of phenomena efficiently using a very compact set of spectrum bases.

While real world sounds typically exhibit non-stationary spectral characteristics, the standard NMF implicitly assumes that each elementary component (which is expected to correspond to a single note activation) of a signal under analysis is represented by a ”rank 1” spectrogram. This means that the spectrum of each note is assumed to be constant over time up to a scale factor. Thus, learning an important spectral variability with standard NMF would require to use a large number of bases, and some post-processing to group the bases into single events.

Several approaches have been proposed to overcome this problem. For example, nonnegative matrix factor deconvolution [2] introduces temporal-spectral bases, while [3] considers an extension of NMF where temporal activations are replaced with time-frequency activations based on a source/filter model. Another kind of approaches based on state variations of the spectral patterns has been an active area of research to model sounds whose spectral characteristics evolve other time, such as the piano with attack and release parts. Factorial scaled hidden Markov model [4] or NMF with Markov-chained bases [5] intend to represent time-varying spectra as state transitions through a limited and fixed number of spectral patterns. However, the important problem of the automatic determination of the number of states remained to be solved. For example, while a piano note is likely to be well expressed with a few spectral patterns, one would expect a singing voice to require many. The number of states is thus an important factor to characterize the diversity of each instrument. Hopefully, the adequate number of states should be assigned in response to the tested instrumental sounds.

This paper proposes to extend NMF to represent time-varying spectral patterns of instrument sounds by introducing infinite-state spectral bases. The number of states of basis is adaptively optimized depending on the instrument sounds. Bayesian NMF with deformable bases is explained in Section 2. Section 3 presents the infinite-state spectrum model, an extension of NMF with deformable bases. Section 4 derives an efficient optimization algorithm based on collapsed variational inference. Validation experiments are presented in Section 5.

2. BAYESIAN NONNEGATIVE MATRIX FACTORIZATION WITH DEFORMABLE BASES

NMF applied to audio signal analysis is based on a signal model where the magnitude or power spectrogram $\boldsymbol{Y} = (Y_{\omega,t})_{\Omega\times T} \in \mathbb{R}^{\geq 0,\Omega\times T}$, where $\omega = 1,\ldots,\Omega$ is a frequency bin index, and $t = 1,\ldots,T$ is a time frame index, is factorized into nonnegative parameters, $\boldsymbol{H} = (H_{\omega,d})_{\Omega\times D} \in \mathbb{R}^{\geq 0,\Omega\times D}$ and $\boldsymbol{U} = (U_{d,t})_{D\times T} \in \mathbb{R}^{\geq 0,D\times T}$. This can be written as

$$Y_{\omega,t} \approx \sum_d H_{\omega,d} U_{d,t} \ , \tag{1}$$

where D is the number of bases $\boldsymbol{h}_d = [H_{1,d},\ldots,H_{\Omega,d}]$. The term *component* is used to refer to one basis $\boldsymbol{h}_d$ and its time-varying gain $U_{d,t}$. The bases can be considered as spectral patterns which are frequently observed.

Hopefully, one factorized component should represent a single event, but audio events actually often have varying spectral patterns. We thus propose to consider deformable bases $\boldsymbol{H} = \{(H^{(k)}_{\omega,1})_{\Omega\times K},\ldots,(H^{(k)}_{\omega,D})_{\Omega\times K}\}$, where $(H^{(k)}_{\omega,d})_{\Omega}$ denotes the k-th possible state for the spectral basis of the d-th component. If we let $\boldsymbol{Z} = (Z_{d,t})_{D\times T} \in \mathbb{N}$ denote which spectral basis state of the d-th component is activated at time t, NMF with deformable bases can be written as

$$Y_{\omega,t} \approx \sum_d H^{(Z_{d,t})}_{\omega,d} U_{d,t} \ . \tag{2}$$

Assuming that the generalized Kullback-Leibler divergence is used as the divergence measure, the model can be expressed as the following generative model, similarly to [6]:

$$Y_{\omega,t} = \sum_d C_{\omega,t,d} \ , \ C_{\omega,t,d} \sim \mathrm{Poisson}(C_{\omega,t,d} \mid H^{(Z_{d,t})}_{\omega,d} U_{d,t}) \ .$$

978-1-4577-0539-7/11/$26.00 ©2011 IEEE

It is possible to introduce various prior structures for $\boldsymbol{H}$ and $\boldsymbol{U}$ according to the data and the requirements of the considered applications. For the spectral bases $\boldsymbol{H}$, we here propose to use a Gamma distribution, which is the conjugate prior to the Poisson distribution:

$$H_{\omega,d}^{(k)} \sim \mathrm{Gamma}(a_\omega, b_\omega) \ , \tag{3}$$

the primary motivation for this choice being computational convenience. To promote temporal continuity, often encountered in real-world sounds, we introduce the following prior distribution on the time-varying gains $\boldsymbol{U}$, similarly to [7]:

$$\begin{aligned} W_{d,t} \mid U_{d,t} &\sim \mathrm{Gamma}(\alpha, \alpha U_{d,t}) \ , & (4) \\ U_{d,t+1} \mid W_{d,t} &\sim \mathrm{Gamma}(\beta, \beta W_{d,t}) \ , & (5) \end{aligned}$$

where $\boldsymbol{W} = (W_{d,t})_{D \times T}$ are auxiliary variables. If we set α and β to large values, the prior will enforce the temporal smoothness of $\boldsymbol{U}$.

3. INFINITE-STATE SPECTRUM MODEL

It is important to determine the number of states for each component. For example, a piano note would be often characterized by several spectral patterns such as "attack", "decay", "sustain" and "release". Three or four bases may thus be needed to represent the spectrogram of the piano. As another example, singing voices and stringed instruments feature a particular musical effect, vibrato. Vibrato may need to be expressed using many spectral patterns. Thus, it is desirable to automatically determine the appropriate number of deformable bases for fitting to each instrument sound. To achieve this, we introduce the Dirichlet process (DP) into the deformable bases. DP is popularly used as a nonparametric prior in hierarchical Bayesian specification [8, 9]. Several practical methods for the construction of DP have been derived [8, 10].

Let us first show how to construct it here based on a symmetric Dirichlet prior. We consider the sequence $\{Z_{d,1}, \ldots, Z_{d,T}\}$ of state indices for the d-th component as a sequence of discrete indicator variable, where each $Z_{d,t}$ can take on values $1, \ldots, K$ with proportions given by $\boldsymbol{\pi}_d = \{\pi_{d,1}, \ldots, \pi_{d,K}\}$. The joint distribution of the sequence is multinomial

$$p(Z_{d,1}, \ldots, Z_{d,T} \mid \boldsymbol{\pi}_d) = \prod_{k=1}^{K} \pi_{d,k}^{n_d^{(k)}} \ , \tag{6}$$

where $n_d^{(k)} = \sum_{t'=1}^{T} \delta(Z_{d,t'} - k)$ and $\delta(\cdot)$ denotes the Kronecker-delta function to count the number of times $n_d^{(k)}$ that $Z_{d,t'} = k$ ($t' = 1, \ldots, T$) has been drawn. Let us give the mixing proportions a symmetric Dirichlet prior, which is a conjugate prior of the multinomial distribution, with positive concentration hyperparameter $\boldsymbol{\gamma} = \{\gamma_1, \ldots, \gamma_D\}$:

$$p(\pi_d \mid \gamma_d) \quad \sim \quad \mathrm{Dirichlet}(\gamma_d/K, \ldots, \gamma_d/K) \ . \tag{7}$$

The critical property of the DP is shown in the conditional probability of $Z_{d,t}$ given the setting of all other indices in the sequence $\{Z_{d,1}, \ldots, Z_{d,t-1}, Z_{d,t+1}, \ldots, Z_{d,T}\}$ (denoted $\mathbf{Z}_{d,-t}$) under $K \to \infty$:

$$p(Z_{d,t} = k \mid \mathbf{Z}_{d,-t}, \gamma_d) = \begin{cases} \dfrac{n_{d,-t}^{(k)}}{T - 1 + \gamma_d} & (k \in \{1, \ldots, K\} \text{ i.e. represented}) \\ \dfrac{\gamma_d}{T - 1 + \gamma_d} & (\text{for all unrepresented } k, \text{combined}) \end{cases}$$

where $n_{d,-t}^{(k)}$ counts the number of times that $Z_{d,t'} = k$ has been drawn in $\boldsymbol{Z}_{d,-t}$. As we can see, $Z_{d,t}$ tends to choose an already popular state. The concentration parameter γ_d controls the tendency to populate a previously unrepresented state. Each component thus tends to keep an adequate number of states depending on the observed signals.

Another formulation of the DP has also been proposed in terms of a stick-breaking construction [10]:

$$V_{d,k} \sim \mathrm{Beta}(1, \gamma_d) \ , \ \pi_{d,k}(\mathbf{V}_d) = V_{d,k} \prod_{j=1}^{k-1} (1 - V_{d,j}) \ , \tag{8}$$

where $\mathbf{V}_d = \{V_{d,1}, V_{d,2}, \ldots\}$. This construction is regarded as breaking pieces off a unit-length stick successively with size determined by variable $V_{d,k} \in [0, 1]$ drawn from $\mathrm{Beta}(1, \gamma_d)$. The k-th broken piece shows the proportion $\pi_{d,k}$.

Note that the present model can be considered as an HMM with uniform transition probabilities. Future work will include the introduction of the hierarchical DP [11] to incorporate transition probabilities in order to more accurately model the succession of spectral patterns. Also note that although the number of components D needs here to be given by hand, it should ideally be determined adaptively according to the input data for example using the techniques described in [12, 13].

4. VARIATIONAL INFERENCE ALGORITHM

Inference algorithms for models with DP prior are mostly based on sampling methods. However, a variational Bayesian (VB) approach is often preferable for large-scale problems. Even though we target a short sequence of music, its spectrogram comprises a large number of parameters. We thus prefer here the framework of VB inference.

Some approaches have been proposed for variational inference involving DP [9, 14]. In the context of VB, DP is usually approximated by finite symmetric Dirichlet prior (FSD) or truncated stick-breaking construction (TSB) [15]. In this paper, we will discuss only the TSB for the sake of brevity. TSB is obtained by using a fixed value K and setting $\forall d, p(V_{d,K} = 1) = 1$; this implies that the mixture proportions $\pi_{d,k}$ are equal to zero for $k > K$. Note that the proposed model does have "infinite-state" spectral bases, and that the finite approximation is only linked to the use of variational inference. As a matter of fact, we do not have to make such an approximation if we use sampling methods, such as Gibbs sampling.

The VB approach in general assumes a factorized form for the posterior distribution. That can be regarded as assuming that the parameters are independent of each other. Here, because of the strong impact of $\boldsymbol{\pi}_d$ on $\mathbf{Z}_d$, it seems difficult to make such an assumption without impeding the performance of the inference. As shown by [14], a collapsed approach based on integrating out $\boldsymbol{\pi}_d$ can be applied to overcome this problem, and we thus use collapsed variational Bayes instead of standard VB. Here, we integrate out $\boldsymbol{\pi}_d$ and then the joint collapsed model is given by

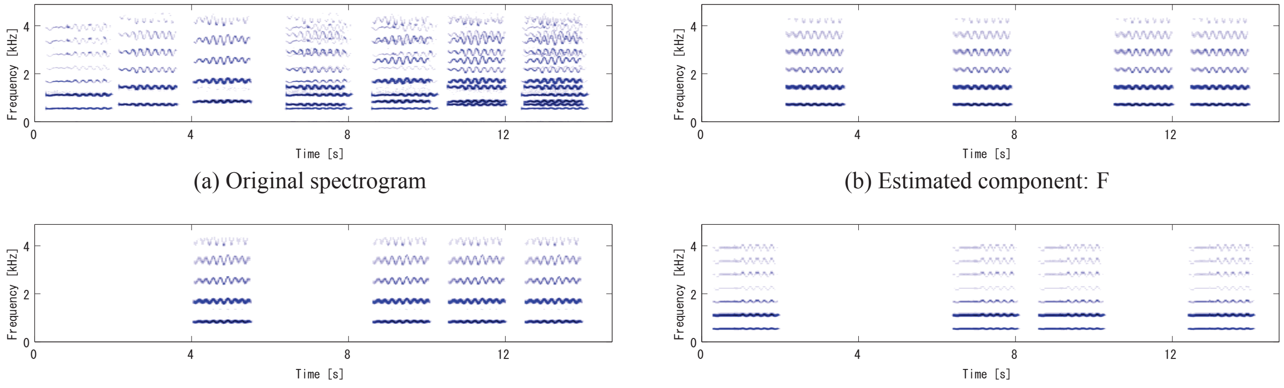
$$\begin{aligned} &P(\boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{H}, \boldsymbol{U}, \boldsymbol{Z}, \boldsymbol{W}) \\ &\quad = p(\boldsymbol{Y} \mid \boldsymbol{C}) p(\boldsymbol{C} \mid \boldsymbol{H}, \boldsymbol{U}, \boldsymbol{Z}) p(\boldsymbol{Z}) p(\boldsymbol{H}) p(\boldsymbol{U}, \boldsymbol{W}) \end{aligned} \tag{9}$$

where $p(\boldsymbol{Z})$ is given by (marginalizing out $\boldsymbol{\pi}_d$ from Eq. (8))

$$p(\boldsymbol{Z}) = \prod_d \prod_k \frac{\Gamma(1 + n_d^{(k)}) \Gamma(\gamma_d + \sum_{k' > k} n_d^{(k')})}{\Gamma(1 + \gamma_d + \sum_{k' \geq k} n_d^{(k')})} \ . \tag{10}$$

We want to maximize the log marginal likelihood $\mathcal{L}(\boldsymbol{Y})$. As it is difficult to do it directly, VB uses a lower bound $\mathcal{B}(\boldsymbol{Y})$ for $\mathcal{L}(\boldsymbol{Y})$ based on the assumption that parameters are independent of each other. The lower bound is given by

$$\mathcal{B}(\boldsymbol{Y}) = \sum_{\boldsymbol{Z}} \int Q(\Xi) \log \frac{P(\boldsymbol{Y}, \boldsymbol{C}, \boldsymbol{H}, \boldsymbol{U}, \boldsymbol{Z}, \boldsymbol{W})}{Q(\Xi)} d\boldsymbol{C} d\boldsymbol{H} d\boldsymbol{U} d\boldsymbol{W}$$



(a) Original spectrogram

(b) Estimated component: F

(c) Estimated component: A♭

(c) Estimated component: D♭

Fig. 1. Original spectrogram of vocal signals (a), estimated model $\sum_k \mathbb{E}[H^{(k)}_{\omega,d}]q(Z_{d,t}=k)\mathbb{E}[U_{d,t}]$ (not $\mathbb{E}[C_{\omega,t,d}]$) of each component (b), (c) and (d). We set the number of bases to $D=3$ and truncated level to $K=30$. Owing to parsimony and temporal smoothness, the proposed method was able to decompose a note with vibrato in an unsupervised way.

where $Q(\Xi) = q(\boldsymbol{C})q(\boldsymbol{Z})q(\boldsymbol{H})q(\boldsymbol{U})q(\boldsymbol{W})$. In the following, for the sake of brievity, we only sketch the update rules. Let us first focus on the update rules for $q(\boldsymbol{Z})$. We find the update:

$$q(Z_{d,t}) \propto \exp\left(\mathbb{E}_{\prod_{m\neq t} q(Z_{d,m})}\left[\log p(Z_{d,t} \mid \boldsymbol{Z}_{d,-t})\right]\right) \times \exp\left(\mathbb{E}_{q(\Theta_{Z_{d,t}})q(\boldsymbol{C})}[\log p(\boldsymbol{C}_{t,d} \mid \Theta_{Z_{d,t}})]\right) \tag{11}$$

where $\Theta_{Z_{d,t}} = \{\boldsymbol{h}_d^{(Z_{d,t})}, U_{d,t}\}$. Integrating out parameters often results in expensive computational cost. This can however be avoided by using a Gaussian approximation [14]. This approximation can be applied to random variables $n^{(k)}_{d,-t}, n^{(\geq k)}_{d,-t}$ which are sums over Bernoulli variables. Thanks to the central limit theorem, these sums can be effectively approximated using Gaussian distributions. In order to apply this framework, we use the following second order Taylor expansion: $\mathbb{E}(\log \rho_m) \approx \log(\mathbb{E}[\rho_m]) - \mathbb{V}[\rho_m]/2\mathbb{E}[\rho_m]^2$ for any probabilistic variable ρ_m. We have as for the first term of Eq. (11)

$$\begin{aligned}
&\mathbb{E}_{\prod_{m\neq t} q(Z_{d,m})}\left[\log p(Z_{d,t} \mid \boldsymbol{Z}_{d,-t})\right] \\
&\approx \log\left(1+\mathbb{E}[n^{(k)}_{d,-t}]\right) - \frac{\mathbb{V}[n^{(k)}_{d,-t}]}{2(1+\mathbb{E}[n^{k}_{d,-t}])^2} \\
&+\sum_{j<k}\left\{\log\left(\gamma_d+\mathbb{E}[n^{(>j)}_{d,-t}]\right) - \frac{\mathbb{V}[n^{(>j)}_{d,-t}]}{2(\gamma_d+\mathbb{E}[n^{>j}_{d,-t}])^2}\right\} \\
&-\sum_{j\geq k}\left\{\log\left(1+\gamma_d+\mathbb{E}[n^{(\geq j)}_{d,-t}]\right) - \frac{\mathbb{V}[n^{(\geq j)}_{d,-t}]}{2(1+\gamma_d+\mathbb{E}[n^{\geq j}_{d,-t}])^2}\right\} .
\end{aligned}$$

As for the second term of Eq. (11), we find

$$\begin{aligned}
&\mathbb{E}_{q(\Theta_{Z_{d,t}})q(\boldsymbol{C})}[\log p(\boldsymbol{C}_{t,d} \mid \Theta_{Z_{d,t}})] \\
&= -\sum_{\omega}\mathbb{E}[H^{(Z_{d,t})}_{\omega,d}U_{d,t}] + \sum_{\omega}\mathbb{E}[C_{\omega,t,d}\log H^{(Z_{d,t})}_{\omega,d}U_{d,t}] .
\end{aligned}$$

Next, we find the update for $q(\boldsymbol{C})$:

$$\begin{aligned}
q(C_{\omega,t,\boldsymbol{d}}) &\propto \text{Multinomial}(C_{\omega,t,\boldsymbol{d}} \mid Y_{\omega,t}, \lambda_{\omega,t,\boldsymbol{d}}) , \\
\lambda_{\omega,t,d} &= \frac{\exp\left(\sum_k q(Z_{d,t}=k)\mathbb{E}[\log H^{(k)}_{\omega,d}U_{d,t}]\right)}{\sum_d \exp\left(\sum_k q(Z_{d,t}=k)\mathbb{E}[\log H^{(k)}_{\omega,d}U_{d,t}]\right)} .
\end{aligned}$$

We can then find the following update rule for $q(\boldsymbol{H})$:

$$\begin{aligned}
q(H^{(k)}_{\omega,d}) &\propto \text{Gamma}(\mu^{(k)}_{\omega,d}, \nu^{(k)}_{\omega,d}) , \\
\mu^{(k)}_{\omega,d} &= a_\omega + \sum_t \mathbb{E}[C_{\omega,t,d}]q(Z_{d,t}=k) , \\
\nu^{(k)}_{\omega,d} &= b_\omega + \sum_t q(Z_{d,t}=k)\mathbb{E}[U_{d,t}] ,
\end{aligned} \tag{12}$$

similarly for $q(\boldsymbol{U})$:

$$\begin{aligned}
q(U_{d,t}) &\propto \text{Gamma}(\eta_{d,t}, \tau_{d,t}) , \\
\eta_{d,t} &= \sum_{\omega,k}\mathbb{E}[C_{\omega,t,d}]q(Z_{d,t}=k) + \alpha + \beta , \\
\tau_{d,t} &= \sum_{\omega,k}\mathbb{E}[H^{(k)}_{\omega,d}]q(Z_{d,t}=k) + \alpha\mathbb{E}[W_{d,t}] + \beta\mathbb{E}[W_{d,t-1}] ,
\end{aligned} \tag{13}$$

and finally for $q(\boldsymbol{W})$:

$$\begin{aligned}
q(W_{d,t}) &\propto \text{Gamma}(\phi_{d,t}, \varphi_{d,t}) , \\
\phi_{d,t} = \alpha+\beta \ , &\quad \varphi_{d,t} = \alpha\mathbb{E}[U_{d,t}] + \beta\mathbb{E}[U_{d,t+1}] .
\end{aligned} \tag{14}$$

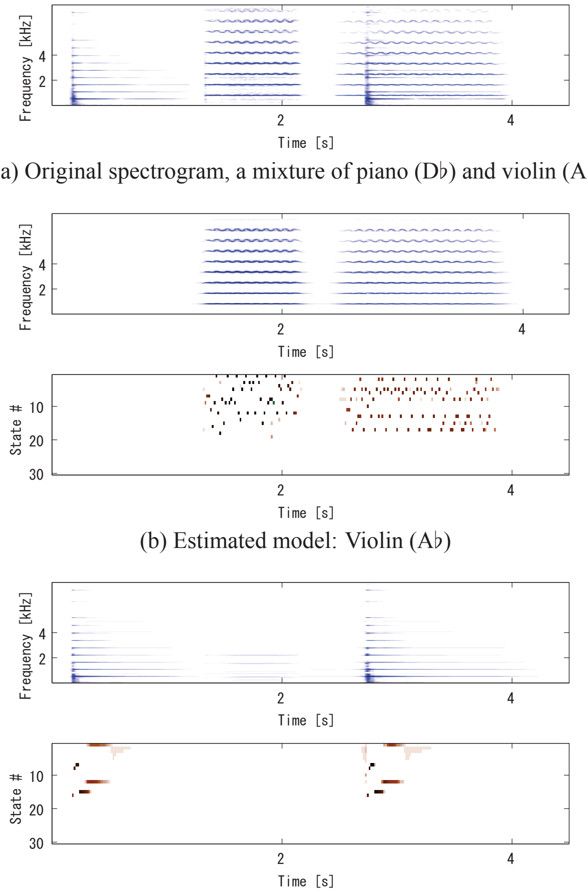
To avoid unexpected local solutions, we propose introducing a weight parameter Υ into $p(\boldsymbol{C} \mid \boldsymbol{H}, \boldsymbol{U}, \boldsymbol{Z})^{\Upsilon}$, and gradually increasing the value of Υ to 1.

5. EXPERIMENTS

We present some results on the application of our algorithm to audio signals, for fully unsupervised sound separation. All data were downmixed to mono and downsampled to 16kHz. The magnitude spectrogram was computed using the short time Fourier transform with 32 ms long Hanning window and with 16 ms overlap.

The treatment of the hyperparameters is important. One possible approach is introducing hyperpriors (such as non-informative priors) on the hyperparameters, which we expect would lead to automatic estimation of the hyperparameters. However, we could instead indicate our wish to fix them so that the model is enforced by sparseness and temporal smoothness. Here, the hyperparameters were set to $a_\omega = b_\omega = 0, \alpha = \beta = 5, \gamma = 1$.

At first, we generated synthetic data, a mixture of vocal signals taken from RWC database (RWC-MDB-I-2001 No.45) [16]. The sequence is composed of 3 notes (Major chord: D♭, F and A♭ have



(a) Original spectrogram, a mixture of piano (D♭) and violin (A♭)

(b) Estimated model: Violin (A♭)

(c) Estimated model: Piano (D♭)

Fig. 2. Original spectrogram (a), estimated model (b) and (c). Estimated model: $\sum_k \mathbb{E}[H^{(k)}_{\omega,d}]q(Z_{d,t} = k)\mathbb{E}[U_{d,t}]$ (top) and $q(Z_{d,t} = k)\mathbb{E}[U_{d,t}]$ (bottom) of each component. We set the number of bases to $D = 2$ and truncated level to $K = 30$. The number of states was automatically optimized depending on the instrumental sounds.

overlapping harmonic components): first, each note is played alone in turn, then all the combinations of two notes are played and finally all notes are played simultaneously. "synthetic" means that each note is played four times and they are synthesized from the same signal. The result is shown in Fig. 1. The proposed method was confirmed to appropriately decompose a note with vibrato with deformable bases.

Next, we generated as test data (Fig. 2), a mixture of the piano (D♭: RWC-MDB-I-2001 No. 1) and the violin (A♭: RWC-MDB-I-2001 No. 15). Each note is played alone in turn, then two notes are played simultaneously. Each note is played twice and each time synthesized from a different signal. The results are shown in Fig. 2. The number of states is determined automatically depending on the instruments.

6. CONCLUSION

We presented a nonparametric Bayesian extension of NMF, which we call infinite-state spectrum model. The proposed model applied to audio signals represents the time-varying spectrum of each note event in polyphonic music. We derived an efficient optimization algorithm based on collapsed variational inference, and presented experimental results showing that our model is fitted for the modeling of non-stationary audio signals. In the future, we will extend this model to an infinite-state Markov chain spectrum model with hierarchical DP [11]. A remaining important problem for audio signal analysis based on NMF is that of the automatic determination of the number of components [12, 13]: future work will include the introduction of an Indian buffet process prior into our model to overcome this problem.

7. REFERENCES

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, Oct. 1999.

[2] Paris Smaragdis, "Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs," in *Proc. ICA*, 2004, pp. 494–499.

[3] R. Hennequin, R. Badeau, and B. David, "NMF with time-frequency activations to model non stationary audio events," *IEEE Trans. on Audio, Speech, and Language Process.*, 2010.

[4] A. Ozerov, C. Févotte, and M. Charbit, "Factorial scaled hidden Markov model for polyphonic audio representation and source separation," in *Proc. WASPAA*, 2009.

[5] M. Nakano, J. Le Roux, H. Kameoka, Y. Kitano, N. Ono, and S. Sagayama, "Nonnegative matrix factorization with markov-chained bases for modeling time-varying patterns in music spectrograms," in *Proc. LVA/ICA*, 2010, pp. 149–156.

[6] C. Févotte and A. T. Cemgil, "Nonnegative matrix factorizations as probabilistic inference in composite models," in *Proc. EUSIPCO*, 2009, vol. 47, pp. 1913–1917.

[7] O. Dikman and A. T. Cemgil, "Unsupervised single-channel source separation using Bayesian NMF," in *Proc. WASPAA*, 2009, pp. 93–96.

[8] C. Antoniak, "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems," *Annals of Statistics*, vol. 2, no. 6, pp. 1152–1174, 1974.

[9] D. M. Blei and M. I. Jordan, "Variational inference for Dirichlet process mixtures," *Journal of Bayesian Analysis*, vol. 1, no. 1, pp. 121–144, 2005.

[10] J. Ishwaran and L. James, "Gibbs sampling methods for stick-breaking priors," *Journal of the American Statistical Association*, pp. 161–174, 2001.

[11] Y. Teh, M. Jordan, M. Beal, and D. Blei, "Hierarchical Dirichlet processes," *Advances in neural information processing systems*, vol. 17, pp. 1385–1392, 2004.

[12] V. Y. F. Tan and C. Fevotte, "Automatic relevance determination in nonnegative matrix factorization," in *Proc. SPARS*, 2009.

[13] M. Hoffman, D. Blei, and P. Cook, "Bayesian nonparametric matrix factorization for recorded music," in *Proc. ICML*, 2010.

[14] K. Kenichi, MaxWelling, and Y. Whye Teh, "Collapsed variational dirichlet process mixture models," in *Proc. IJCAI*, 2007.

[15] J. Sethuraman, "A constructive definition of Dirichlet priors," *Statistica Sinica*, vol. 4, pp. 639–650, 1994.

[16] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "RWC music database: Popular, classical, and jazz music database," in *Proc. ISMIR*, 2002, pp. 287–288.