

Dynamic Nonparametric Bayesian Models for Analysis of Music

Lu Ren¹, David Dunson², Scott Lindroth³ and Lawrence Carin¹

¹Department of Electrical and Computer Engineering

²Department of Statistical Science

³Department of Music

Duke University

Durham, NC, 27708

Email: {lr,lcarin}@ee.duke.edu, dunson@stat.duke.edu, scott.lindroth@duke.edu

Abstract

The dynamic hierarchical Dirichlet process (dHDP) is developed to model complex sequential data, with a focus on audio signals from music. The music is represented in terms of a sequence of discrete observations, and the sequence is modeled using a hidden Markov model (HMM) with time-evolving parameters. The dHDP imposes the belief that observations that are temporally proximate are more likely to be drawn from HMMs with similar parameters, while also allowing for “innovation” associated with abrupt changes in the music texture. The sharing mechanisms of the time-evolving model are derived, and for inference a relatively simple Markov chain Monte Carlo sampler is developed. Segmentation of a given musical piece is constituted via the model inference. Detailed examples are presented on several pieces, with comparisons to other models. The dHDP results are also compared with a conventional music-theoretic analysis.

Key Words: dynamic Dirichlet process; Hidden Markov Model; Mixture Model; Segmentation; Sequential data; Time series.

1. Introduction

The analysis of music is of interest to music theorists, for aiding in music teaching, for analysis of human perception of sounds (Temperley, 2008), and for design of music search and organization tools (Ni et al., 2008). An example of the use of Bayesian techniques for analyzing music may be found in (Temperley, 2007). However, in (Temperley, 2007) it is generally assumed that the user has access to MIDI files (musical instrument digital interface), which means that the analyst knows exactly what notes are sounding when. We are interested in processing the acoustic waveform directly; while the techniques developed here are of interest for music, they are also applicable for analysis of general acoustic waveforms. For example, a related problem which may be addressed using the proposed approach is the segmentation of audio waveforms for automatic speech and speaker recognition (*e.g.*, for labeling different speakers in a teleconference (Fox et al., 2008)).

As motivation we start by considering a well known musical piece: “A Day in the Life” from the Beatle’s album *Sgt. Peppers Lonely Hearts Club Band*. The piece is 5 minutes and 33 seconds long, and the entire audio waveform is plotted in Figure 1. To process these data, the acoustic signal was sampled at 22.05 KHz and divided into 50 ms contiguous frames. Mel frequency cepstral coefficients (MFCCs) (Logan, 2000) were extracted from each frame, these being effective for representing perceptually important parts of the spectral envelope of audio signals (Jensen et al., 2006). The MFCC features are linked to spectral characteristics of the signal over the 50 ms window, and this mapping yields a 40-dimensional vector of real numbers for each

frame. Therefore, after the MFCC analysis the music is converted to a sequence of 40-dimensional real vectors. The details of the model follow below, and here we only seek to demonstrate our objective. Specifically, Figure 2 shows a segmentation of the audio waveform, where the indices on the figure correspond to data subsequences; each subsequence is defined by a set of 75 consecutive 50 ms frames. The results in Figure 2 quantify how inter-related any one subsequence of the music is to all others. We observe that the music is decomposed into clear contiguous segments of various lengths, and segment repetitions are evident. This Beatle’s song is a relatively simple example, for the piece has many distinct sections (vocals, along with clearly distinct instrumental parts). A music-theoretic analysis of the results in Figure 2 indicates that the segmentation correctly captures the structure of the music. In the detailed results presented below, we consider much “harder” examples. Specifically, we consider classical piano music for which there are no vocals, and for which distinct instruments are not present (there is a lack of timbral variety, which makes this a more difficult challenge). We also provide a detailed examination of the quality of the inferred music segmentation, based on music-theoretic analysis.

A typical goal of the music analysis is to segment a given piece, with the objective of inferring inter-relationships among motive and themes within the music. We wish to achieve this task without *a priori* setting the number of segments or their length, motivating a non-parametric framework. A key aspect of our proposed model is an explicit imposition of the belief that the likelihood that two subsequences of music are similar (contained within the same or related segments) increases as they become more proximate temporally.

A natural tool for segmenting or clustering data is the Dirichlet process (DP)(Ferguson, 1973; Blackwell and MacQueen, 1973). In order to share statistical strength across different groups of data, the hierarchical Dirichlet process (HDP) (Teh et al., 2006)

has been proposed to model the dependence among groups through sharing the same set of discrete parameters (“atoms”), and the mixture weights associated with different atoms are varied as a function of the data group. In DP-based mixture models of this form, it is assumed that the data are generated independently and are exchangeable. In the HDP it is assumed that the data groups are exchangeable. However, in many applications data are measured in a sequential manner, and there is information in this temporal character that should ideally be exploited; this violates the aforementioned assumption of exchangeability. For example, music is typically composed according to a sequential organization and the long-term dependence in the time series, known as the distance patterns in music theory, should be accounted for in an associated music model (Paiement et al., 2007; Aucouturier and Pachet, 2007).

The analysis of sequential data has been a longstanding problem in statistical modeling. With music as an example, Paiement et al. (2007) proposed a generative model for rhythms based on the distributions of distances between subsequences; to annotate the changes in mixed music, Plotz et al. (2006) used stochastic models based on the Snip-Snap approach, by evaluating the Snip model for the Snap window at every position within the music. However, these methods are either based on one specific factor (rhythm) of music (Paiement et al., 2007) or need prior knowledge of the music’s segmentation (Plotz et al., 2006). Recently, a hidden Markov model (HMM) (Rabiner, 1989) was used to model monophonic music by assuming all the subsequences are drawn i.i.d. from one HMM (Raphael, 1999); alternatively, an HMM mixture (Qi et al., 2007) was applied to model the variable time-evolving properties of music, within a semi-parametric Bayesian setting. In both of these HMM music models the music was divided into subsequences, with an HMM employed to represent each subsequence; such an approach does not account for the expected statistical relationships between temporally proximate subsequences. By considering one piece

of music as a whole (avoiding subsequences), an infinite HMM (iHMM) (Teh et al., 2006; Ni et al., 2008) was proposed to automatically learn the model structure with countably infinite states. While the iHMM is an attractive model, it has limitations for the music modeling and segmentation of interest here, with this discussed further below.

Developing explicit temporally dependent models has recently been the focus of significant interest. A related work is the dynamic topic model (Blei and Lafferty, 2006; Wei et al., 2007), in which the model parameter at the previous time $t - 1$ is the expectation for the distribution of the parameter at the next time t , and the correlation of the samples at adjacent times is controlled through adjusting the variance of the conditional distribution. Unfortunately, the non-conjugate form of the conditional distribution requires approximations in the model inference.

Recently Dunson (2006) proposed a Bayesian dynamic model to learn the latent trait distribution through a mixture of DPs, in which the latent variable density can change dynamically in location and shape across levels of a predictor. This model has the drawback that mixture components can only be added over time, so one ends up with more components at later times. However, of interest for the application considered here, music has the property that characteristics of a given piece may repeat over time, which implies the possible repetitive use of the same mixture component with time. Based on this consideration, a similar dynamic structure as in (Dunson, 2006) is considered here to extend the hierarchical Dirichlet process (HDP) to incorporate time dependence.

A brief summary of DP and HDP is provided in Section 2.1. The proposed dynamic model structure is described in Section 2.2, with associated properties discussed in Section 2.3. Model inference is described in Section 2.5. Two detailed experimental results are provided in Section 3, followed by conclusions in Section 4.

2. Dynamic Hierarchical Dirichlet Processes

2.1 Background

As indicated in the Introduction, a given piece of music is mapped to a sequence of 40-dimensional real vectors via MFCC feature extraction. The MFCCs are the most widely employed features for processing audio signals, particularly in speech processing. To simplify the HMM mixture models employed here, each 40-dimensional real vector is quantized via vector quantization (VQ) (Gersho and Gray, 1992) (Barnard et al., 2003), and here the codebook is of dimension $M = 16$. For example, after VQ, the continuous waveform in Figure 1 is mapped to the sequence of codes depicted in Figure 3; it is a sequence of this type that we wish to analyze.

The standard tool for analysis of sequential data is the hidden Markov model (HMM) (Rabiner, 1989). For the discrete sequence of interest, given an observation sequence $\mathbf{x} = \{x_t\}_{t=1}^T$ with $x_t \in \{1, \dots, M\}$, the corresponding hidden state sequence is $\mathbf{S} = \{s_t\}_{t=1}^T$, from which $s_t \in \{1, \dots, I\}$. An HMM is represented by parameters $\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{B}, \boldsymbol{\pi}\}$, defined as

- $\mathbf{A} = \{a_{\rho\xi}\}$, $a_{\rho\xi} = Pr(s_{t+1} = \xi | s_t = \rho)$: state transition probability;
- $\mathbf{B} = \{b_{\rho m}\}$, $b_{\rho m} = Pr(x_t = m | s_t = \rho)$: emission probability;
- $\boldsymbol{\pi} = \{\pi_\rho\}$, $\pi_\rho = Pr(s_1 = \rho)$: initial state distribution.

To model the whole music piece with one HMM (Raphael, 1999), one may divide the sequence into a series of subsequences $\{\mathbf{x}_j\}_{j=1}^J$, with $\mathbf{x}_j = \{x_{jt}\}_{t=1}^T$ and $x_{jt} \in \{1, \dots, M\}$. The joint distribution of the observation subsequences given the model parameters $\boldsymbol{\theta}$ yields

$$p(\mathbf{x}|\boldsymbol{\theta}) = \prod_{j=1}^J \left\{ \sum_{\mathbf{S}_j} \pi_{s_{j,1}} \prod_{t=1}^{T-1} a_{s_{j,t}, s_{j,t+1}} \prod_{t=1}^T b_{s_{j,t}, x_{j,t}} \right\} \quad (1)$$

However, rather than employing a single HMM for a given piece, which is clearly overly-simplistic, we allow the music dynamics to vary with time by letting

$$\mathbf{x}_j \sim F(\boldsymbol{\theta}_j), \quad j = 1, \dots, J, \quad (2)$$

which denotes that the subsequence \mathbf{x}_j is drawn from an HMM with parameters $\boldsymbol{\theta}_j$. In order to accommodate dependence across the subsequences, we can potentially let $\boldsymbol{\theta}_j \sim G$, with $G \sim DP(\alpha_0 G_0)$, where G_0 is a base probability measure having positive mass, and α_0 is a positive real number (Ferguson, 1973). Sethuraman (1994) showed that

$$G = \sum_{k=1}^{\infty} p_k \delta_{\boldsymbol{\theta}_k^*}, \quad p_k = \tilde{p}_k \prod_{i=1}^{k-1} (1 - \tilde{p}_i) \quad (3)$$

where $\{\boldsymbol{\theta}_k^*\}_{k=1}^{\infty}$ represent a set of atoms drawn i.i.d. from G_0 and $\{p_k\}_{k=1}^{\infty}$ represent a set of weights, with the constraint $\sum_{k=1}^{\infty} p_k = 1$; each \tilde{p}_k is drawn i.i.d. from the beta distribution $Be(1, \alpha_0)$. Since in practice the $\{p_k\}_{k=1}^{\infty}$ statistically diminish with increasing k , a truncated stick-breaking process (Ishwaran and James, 2001) is often employed, with a large truncation level K , to approximate the infinite stick breaking process (in this approximation $\tilde{p}_K = 1$). We note that a draw G from a $DP(\alpha_0 G_0)$ is discrete with probability one.

2.2 Nonparametric Bayesian Dynamic Structure

Placing a DP on the distribution of the subsequence-specific HMM parameters, $\boldsymbol{\theta}_j$, allows for borrowing of information across the subsequences, but does not incorporate information that subsequences from proximal times should be more similar. Hence,

we propose a more flexible dynamic mixture model in which

$$\boldsymbol{\theta}_j \sim G_j, \quad G_j = \sum_{k=1}^{\infty} p_{jk} \delta_{\boldsymbol{\theta}_k^*}, \quad \boldsymbol{\theta}_k^* \sim H, \quad (4)$$

where the subsequence-specific mixture distribution G_j has weights that vary with j , represented as \mathbf{p}_j . Including the same atoms for all j allows for repetition in the music structure across subsequences, with the varying weights allowing substantial flexibility. In order to account for relative subsequence positions in a piece, we propose a model that induces dependence between G_{j-1} and G_j by accommodating dependence in the weights.

Also motivated by the problem of accommodating dependence between a sequence of unknown distributions, Dunson (2006) proposed a dynamic mixture of Dirichlet processes. His approach characterized G_j as a mixture of G_{j-1} and an innovation distribution, which is assigned a Dirichlet process prior. The structure allows for the introduction of new atoms, while also incorporating atoms from previous times. There are two disadvantages to this approach in the music application. The first is that the atoms from early times tend to receive very small weight at later times, which does not allow recurrence of themes and grouping of subsequences that are far apart. The second is that atoms are only added as time progresses and never removed, which implies greater complexity in the music piece at later times.

We propose a dynamic HDP (dHDP) with the following structure:

$$G_j = (1 - \tilde{w}_{j-1})G_{j-1} + \tilde{w}_{j-1}H_{j-1} \quad (5)$$

where $G_1 \sim DP(\alpha_{01}G_0)$, H_{j-1} is called an innovation measure drawn from $DP(\alpha_{0j}G_0)$, and $\tilde{w}_{j-1} \sim Be(a_{w(j-1)}, b_{w(j-1)})$. To impose sharing of the same atoms across all

time, $G_0 \sim DP(\gamma H)$. The measure G_j is modified from G_{j-1} by introducing a new innovation measure H_{j-1} , and the random variable \tilde{w}_{j-1} controls the probability of innovation (*i.e.*, it defines the mixture weights).

A draw $G_0 \sim DP(\gamma H)$ may be expressed as

$$G_0 = \sum_{k=1}^{\infty} \beta_k \delta_{\theta_k^*} \quad (6)$$

and the weights are drawn $\boldsymbol{\beta} \sim Stick(\gamma)$, where $Stick(\gamma)$ corresponds to letting $\beta_k = \tilde{\beta}_k \prod_{i=1}^{k-1} (1 - \tilde{\beta}_i)$ with $\tilde{\beta}_k \stackrel{iid}{\sim} Be(1, \gamma)$. Since the same atoms $\theta_k^* \stackrel{iid}{\sim} H$ are used for *all* G_j , it is also possible to share parameters between subsequences widely separated in time; this latter property may be of interest when the music has temporal repetition, as is typical.

The measures G_1, H_1, \dots, H_{J-1} have their own mixture weights with the common expectation equal to $\boldsymbol{\beta}$, yielding

$$G_1 = \sum_{k=1}^{\infty} \zeta_{1k} \delta_{\theta_k^*}, \quad H_1 = \sum_{k=1}^{\infty} \zeta_{2k} \delta_{\theta_k^*}, \quad \dots, \quad H_{J-1} = \sum_{k=1}^{\infty} \zeta_{Jk} \delta_{\theta_k^*} \quad (7)$$

$$\zeta_j \stackrel{ind}{\sim} DP(\alpha_{0j} \boldsymbol{\beta}), \quad j = 1, \dots, J$$

The equivalence between (5)-(6) and (7) follows directly from results in (Teh et al., 2006). Analogous to the discussion at the end of Section 2.1, the different weights $\boldsymbol{\zeta}_j = \{\zeta_{jk}\}_{k=1}^{\infty}$ are independent given $\boldsymbol{\beta}$ since G_1, H_1, \dots, H_{J-1} are independent given G_0 (Teh et al., 2006).

To further develop the dynamic relationship from G_1 to G_J , we extend the mixture structure in (5) from group to group:

$$\begin{aligned}
G_j &= (1 - \tilde{w}_{j-1})G_{j-1} + \tilde{w}_{j-1}H_{j-1} \\
&= \prod_{l=1}^{j-1} (1 - \tilde{w}_l)G_1 + \sum_{l=1}^{j-1} \left\{ \prod_{m=l+1}^{j-1} (1 - \tilde{w}_m) \right\} \tilde{w}_l H_l \\
&= w_{j1}G_1 + w_{j2}H_1 + \dots + w_{jj}H_{j-1}
\end{aligned} \tag{8}$$

where $w_{11} = 1$, $\tilde{w}_0 = 1$, and for $j > 1$ we have $w_{jl} = \tilde{w}_{l-1} \prod_{m=l}^{j-1} (1 - \tilde{w}_m)$, for $l = 1, 2, \dots, j$. It can be easily verified that $\sum_{l=1}^j w_{jl} = 1$ for each j , with w_{jl} the prior probability that parameters for subsequence j are drawn from the l th component distribution, where $l = 1, \dots, j$ indexes G_1, H_1, \dots, H_{j-1} , respectively. Based on the dependent relation induced here, we have an explicit form for each $\{\mathbf{p}_j\}_{j=1}^J$ in (4):

$$\mathbf{p}_j = \sum_{l=1}^j w_{jl} \boldsymbol{\zeta}_l. \tag{9}$$

If all $\tilde{w}_j = 0$, all of the groups share the same mixture distribution related to G_1 and the model reduces to the Dirichlet mixture model described in Section 2.1. If all $\tilde{w}_j = 1$ the model instead reduces to the HDP. In the posterior computation, we treat the $\tilde{\mathbf{w}}$ as random variables and add Beta priors $Be(\tilde{w}_j | a_w, b_w)$ on each \tilde{w}_j with $j = 1, \dots, J - 1$ for more flexibility.

2.3 Sharing Properties

To obtain insight into the dependence structure induced by the dHDP proposed in Section 2.2, this section presents some basic properties. Suppose G_0 is a probability measure on (Ω, \mathcal{B}) , with Ω the sample space of $\boldsymbol{\theta}_j$ and $\mathcal{B}(\Omega)$ the Borel σ -algebra of

subsets of Ω . Then for any $B \in \mathcal{B}(\Omega)$

$$(G_j(B)|G_{j-1}, \tilde{w}_j) \stackrel{\mathcal{D}}{=} G_{j-1}(B) + \Delta_j(B), \quad (10)$$

where $\Delta_j(B) = \tilde{w}_{j-1}\{H_{j-1}(B) - G_{j-1}(B)\}$ is the random deviation from G_{j-1} to G_j .

Theorem 1. Under the dHDP (8), for any $B \in \mathcal{B}(\Omega)$ we have:

$$E\{ \Delta_j(B)|G_{j-1}, \tilde{w}_{j-1}, G_0, \alpha_{0j} \} = \tilde{w}_{j-1}\{G_0(B) - G_{j-1}(B)\}, \quad (11)$$

$$V\{ \Delta_j(B)|G_{j-1}, \tilde{w}_{j-1}, G_0, \alpha_{0j} \} = \tilde{w}_{j-1}^2 \frac{G_0(B)(1 - G_0(B))}{(1 + \alpha_{0j})}. \quad (12)$$

The proof is straightforward and is omitted. According to Theorem 1, given the previous mixture measure G_{j-1} and the global mixture G_0 , the expectation of the deviation from G_{j-1} to G_j is controlled by \tilde{w}_{j-1} . Meanwhile, the variance of the deviation is related with both \tilde{w}_{j-1} and the precision parameters α_{0j} given G_0 . In limiting case, we obtain the following: If $\tilde{w}_{j-1} \rightarrow 0$, $G_j \rightarrow G_{j-1}$; If $G_{j-1} \rightarrow G_0$, $E(G_j(B)|G_{j-1}, \tilde{w}_{j-1}, G_0, \alpha_{0j}) \rightarrow G_{j-1}(B)$; If $\alpha_{0j} \rightarrow \infty$, $V(\Delta_j(B)|G_{j-1}, \tilde{w}_{j-1}, G_0, \alpha_{0j}) \rightarrow 0$.

Theorem 2. Conditional on the mixture weights \mathbf{w} , the correlation coefficient of the measures between two adjacent groups $G_{j-1}(B)$ and $G_j(B)$ for $j = 2, \dots, J$ is

$$\begin{aligned} \text{Corr}(G_{j-1}, G_j) &= \frac{E\{G_j(B)G_{j-1}(B)\} - E\{G_j(B)\}E\{G_{j-1}(B)\}}{\left[V\{G_j(B)\}V\{G_{j-1}(B)\}\right]^{1/2}} \\ &= \frac{\sum_{l=1}^{j-1} \frac{w_{jl}w_{j-1,l}}{1+\alpha_{0l}} \cdot (\alpha_{0l} + \gamma + 1)}{\left[\sum_{l=1}^j \frac{w_{jl}^2}{1+\alpha_{0l}} \cdot (\alpha_{0l} + \gamma + 1)\right]^{1/2} \left[\sum_{l=1}^{j-1} \frac{w_{j-1,l}^2}{1+\alpha_{0l}} \cdot (\alpha_{0l} + \gamma + 1)\right]^{1/2}} \end{aligned} \quad (13)$$

The proof is given in the Appendix. Due to the lack of dependence on B , Theorem 2 provides a useful expression for the correlation between the measures, which can provide insight into the dependence structure. To study how the correlation depends on $\tilde{\mathbf{w}}$ and $\boldsymbol{\alpha}_0$, we focus on $Corr(G_1, G_2)$ and (i) in Figure 4(a) we plot the correlation coefficient $Corr(G_1, G_2)$ as a function of \tilde{w}_1 , with the precision parameters γ and $\boldsymbol{\alpha}_0$ fixed at one; (ii) in Figure 4(b) we plot $Corr(G_1, G_2)$ as a function of α_{02} , with $\tilde{w}_1 = 0.5$, $\alpha_{01} = 1$ and $\gamma = 10$; (iii) in Figure 4(c) we consider the plot of $Corr(G_1, G_2)$ as a function of both the variables of \tilde{w}_1 and α_{02} given fixed values of $\gamma = 10$ and $\alpha_{01} = 1$. It is observed that the correlation between adjacent groups increases with smaller \tilde{w} and larger $\boldsymbol{\alpha}_0$. If we assume that $\alpha_{0l} = \alpha$ for $l = 1, \dots, j$, then the correlation coefficient has the simple form

$$Corr(G_{j-1}, G_j) = \frac{\sum_{l=1}^{j-1} w_{jl} w_{j-1,l}}{\left\{ \sum_{l=1}^j w_{jl}^2 \right\}^{1/2} \left\{ \sum_{l=1}^{j-1} w_{j-1,l}^2 \right\}^{1/2}}. \quad (14)$$

2.4 Comparisons with Alternative Models

It is useful to consider relationships between the proposed dHDP and other dynamic nonparametric Bayes models. A particularly relevant connection is to dependent Dirichlet processes (DDPs) (MacEachern, 1999), which provide a class of priors for dependent collections of random probability measures indexed by time, space or predictors. DDPs were applied to time series settings by Rodriguez and Ter Horst (2008). Dynamic DDPs have the property that the probability measure at a given time is marginally assigned a Dirichlet process prior, while allowing for dependence between the measures at different times through a stochastic process in the weights and/or atoms. Most of the applications have relied on the assumption of fixed weights, while allowing the atoms to vary according to a stochastic process. Varying weights

is well motivated in the music application due to repetition in the music piece, and can be accommodated by the order-based DDP (Griffin and Steel, 2006) and the local Dirichlet process (Chung and Dunson, 2009). However, these approaches do not naturally allow long-range dependence and can be complicated to implement. Simpler approaches were proposed by Caron et al. (2008) using dynamic linear models with Dirichlet process components and by Caron, Davy and Doucet (2007) using a dynamic modification of the DP Polya urn scheme. Again, these approaches do not automatically allow long range dependence.

The dHDP can alternatively be characterized as a process that first draws a latent collection of distributions, $\mathcal{H} = \{G_1, H_1, \dots, H_{J-1}\}$, from an HDP, with the HDP providing a special case of the DDP framework. The j th parameter vector, θ_j , is then associated with the l th distribution in the collection \mathcal{H} with probability w_{jl} . This specification simplifies posterior computation and interpretation, while allowing a flexible long range dependence structure. An alternative to the HDP would be to choose a nested Dirichlet process (nDP) (Rodriguez et al., 2008) prior for the collection \mathcal{H} . The nDP would allow clustering of the component distributions within \mathcal{H} ; distributions within a cluster are identical while distributions in different clusters have different atoms and weights. This structure also accommodates long range dependence but in a very different manner that may be both more difficult to interpret and more flexible in allowing different atoms at different times.

2.5 Posterior Computation

There are two commonly-used Gibbs sampling strategies for posterior computation in DPMs. The first relies on marginalizing out the random measure through use of the Polya urn scheme (Bush and MacEachern, 1996), while the second relies on truncations of the stick-breaking representation (Ishwaran and James, 2001). As it is

not straightforward to obtain a generalized urn scheme for the dHDP, we rely on the latter approach, which is commonly-referred to as the blocked Gibbs sampler. The primary conditional posterior distributions used in implementing this approach are listed as follows:

1. The update of \tilde{w}_l , for $l = 1, \dots, J - 1$ from its full conditional posterior distribution, has the simple form:

$$(\tilde{w}_l | \dots) \sim Be \left[a_w + \sum_{j=l+1}^J \delta(r_{j(l+1)} = 1), b_w + \sum_{j=l+1}^J \sum_{h=1}^l \delta(r_{jh} = 1) \right] \quad (15)$$

where $\{\mathbf{r}_j\}_{j=1}^J$ are indicator vectors and $\delta(r_{jl} = 1)$ denotes that $\boldsymbol{\theta}_j$ is drawn from the l th component distribution in (8). In (15) and in the results that follow, for simplicity, the distributions $Be(a_{wj}, b_{wj})$ are set with fixed parameters $a_{wj} = a_w$ and $b_{wj} = b_w$ for all time samples. The function $\delta(\cdot)$ equals 1 if (\cdot) is true and 0 otherwise.

2. The full conditional distribution of $\tilde{\zeta}_{lk}$, for $l = 1, \dots, J$ and $k = 1, \dots, K$, is updated under the conjugate prior: $\tilde{\zeta}_{lk} \sim Be \left[\alpha_{0l} \beta_k, \alpha_{0l} (1 - \sum_{m=1}^k \beta_m) \right]$, which is specified in (Teh et al., 2006). The likelihood function associated with each $\tilde{\zeta}_l$ is proportional to $\prod_{k=1}^K \zeta_{lk}^{\sum_{j=l}^J \delta(r_{jl}=1, z_{jk}=1)}$, where \mathbf{z}_j is another indicator vector, with $z_{jk} = 1$ if the subsequence \mathbf{x}_j is allocated to the k^{th} atom ($\boldsymbol{\theta}_j = \boldsymbol{\theta}_k^*$) and $z_{jk} = 0$ otherwise. K represents the truncation level and $\zeta_{lk} = \tilde{\zeta}_{lk} \prod_{m=1}^{k-1} (1 - \tilde{\zeta}_{lm})$. Then the conditional posterior of $\tilde{\zeta}_{lk}$ has the form

$$\begin{aligned} (\tilde{\zeta}_{lk} | \dots) \sim Be & \left[\alpha_{0l} \beta_k + \sum_{j=1}^J \delta(r_{jl} = 1, z_{jk} = 1), \right. \\ & \left. \alpha_{0l} (1 - \sum_{l=1}^k \beta_l) + \sum_{j=1}^J \sum_{k'=k+1}^K \delta(r_{jl} = 1, z_{jk'} = 1) \right] \quad (16) \end{aligned}$$

3. The update of the indicator vector \mathbf{r}_j , for $j = 1, \dots, J$, is completed by generating samples from a multinomial distribution with entries

$$Pr(r_{jl} = 1 | \dots) \propto \tilde{w}_{l-1} \prod_{m=l}^{j-1} (1 - \tilde{w}_m) \prod_{k=1}^K \left\{ \tilde{\zeta}_{lk} \prod_{q=1}^{k-1} (1 - \tilde{\zeta}_{lq}) \cdot Pr(\mathbf{x}_j | \boldsymbol{\theta}_k^*) \right\}^{z_{jk}}, \quad l = 1, \dots, j \quad (17)$$

with $Pr(\mathbf{x}_j | \boldsymbol{\theta}_k^*)$ the likelihood of subsequence j given allocation to the k th atom, $\boldsymbol{\theta}_j = \boldsymbol{\theta}_k^*$. The posterior probability $Pr(r_{jl} = 1)$ is normalized so $\sum_{l=1}^j Pr(r_{jl} = 1) = 1$.

4. The sampling of the indicator vector \mathbf{z}_j , for $j = 1, \dots, J$, is also generated from a multinomial distribution with entries specified as

$$Pr(z_{jk} = 1 | \dots) \propto \prod_{l=1}^j \left\{ \tilde{\zeta}_{lk} \prod_{k'=1}^{k-1} (1 - \tilde{\zeta}_{lk'}) \cdot Pr(\mathbf{x}_j | \boldsymbol{\theta}_k^*) \right\}^{r_{jl}}, \quad k = 1, \dots, K. \quad (18)$$

Other unknowns, including $\{\boldsymbol{\theta}_k^*, \tilde{\beta}_k\}_{k=1}^K$ and precision parameters γ , $\boldsymbol{\alpha}_0$, are updated using standard Gibbs steps. As in (Qi et al., 2007), the component parameters \mathbf{A}_k^* , \mathbf{B}_k^* and $\boldsymbol{\pi}_k^*$ are assumed to be *a priori* independent, with the base measure having a product form with Dirichlet components for each of the probability vectors. The specifics on the specification are shown in the Supplemental materials¹. Since the indicator vector \mathbf{z}_j , for $j = 1, \dots, J$, represents the membership of sharing across all the subsequences, we use this information to segment the music, by assuming that the subsequences possessing the same membership should be grouped together. In order to overcome the issue of label switching that exists in Gibbs sampling, we use the similarity measure $E(\mathbf{z}'\mathbf{z})$ instead of the membership \mathbf{z} in the results. Here $E(\mathbf{z}'\mathbf{z})$ is approximated by averaging the quantity $\mathbf{z}'\mathbf{z}$ from multiple iterations, and in each iteration $\mathbf{z}'_j \mathbf{z}_{j'}$ measures the sharing degree of $\boldsymbol{\theta}_j$ and $\boldsymbol{\theta}_{j'}$ by integrating out the in-

1. <http://sites.google.com/site/bayesianmusic/>

dex of atoms. Related clustering representations of non-parametric models have been considered in (Medvedovic and Sivaganesan, 2002).

3. Experimental Results

To apply the dHDP-HMM proposed in Section 2 to music data, we first complete the prior specification by choosing hyperparameter values. In particular, the prior for $\tilde{\mathbf{w}}$ is chosen to encourage the groups to be shared; consequently, we set the prior $\prod_{j=1}^{J-1} Be(\tilde{w}_j; a_w, b_w)$ with $a_w = 1$ and $b_w = 5$. Since the precision parameters γ and $\boldsymbol{\alpha}_0$ control the prior distribution on the number of clusters, the hyper-parameter values should be chosen carefully. Here we set $Ga(1, 1)$ for γ and each component of $\boldsymbol{\alpha}_0$. Meanwhile, We set the truncation level for DP at $K = 40$.

We recommend running the Gibbs sampler for 100,000 iterations after a 5,000 iteration burn-in based on results applying the diagnostic methods in (Geweke, 1992; Raftery and Lewis, 1992) to multiple chains.

3.1 Statistical Analysis of Beethoven Piece

The music considered below are from particular audio recordings, and may be listened to online². We first consider the movement (“Largo - Allegro”) from the Beethoven’s Sonata No. 17, Op. 31, No. 2 (the “Tempest”). The audio waveform of this piano music is shown in Figure 5. The music is divided into contiguous 100 ms frames, and for each frame the quantized MFCC features are represented by one code from a codebook of size $M = 16$. Each subsequence is of length 60 (corresponding to 6 seconds in total), and for the Beethoven piece considered here there are 83 contiguous subsequences ($J = 83$). The lengths of the subsequences were carefully chosen based on consultation with a music theorist (third author) to be short enough to capture

2. <http://www.last.fm/>

meaningful fine-scale segmentation of the piece. To represent the time dependence inferred by the model, the posterior of indicator \mathbf{r} is plotted in Figure 6(a) to show the mixture-distribution sharing relationship across different subsequences. Figure 6(b) shows the similarity measures $E(\mathbf{z}'\mathbf{z})$ across each pair of subsequences, in which the higher value represents larger probability of the two corresponding subsequences being shared; here \mathbf{z} (see (18)) is a column vector containing one at the position to be associated with the component occupied at the current iteration and zeros otherwise.

For comparison, we now analyze the same music using a DP-HMM (Qi et al., 2007), HDP-HMM (Teh et al., 2006) and an iHMM (Beal et al., 2002; Teh et al., 2006). In the DP-HMM, we use the model in (3), with $F(\boldsymbol{\theta})$ corresponding to an HMM with the same number of states as used in the dHDP; this model yields an HMM mixture model across the music subsequences, and the subsequence order is exchangeable. However, the long time dependence for the music’s coherence is not considered in the components sharing mechanism. For the DP-HMM, we used the same specification of the base measure, H , as in the dHDP-HMM. A Gamma prior $\text{Ga}(1,1)$ is employed as the hyper-prior for the precision parameter α_0 in (3) and the truncation level is also set to 40. The DP-HMM inference was performed with MCMC sampling (Qi et al., 2007). We also consider a limiting case of the dHDP-HMM, for which all innovation weights are zero, with this referred to as an HDP-HMM, with inference performed as in the dHDP, simply with the weights removed. As formulated, the HDP-HMM yields a posterior estimate on the HMM parameters (atoms) for each subsequence, while the DP-HMM yields a posterior estimate on the HMM parameters (atoms) across all of the subsequences. Thus, the HDP-HMM yields an HMM mixture model for each subsequence, and the mixture atoms are shared across all subsequences; for the DP-HMM a single HMM mixture model is learned across all subsequences.

As in Figure 6, we plot the similarity measures $E(\mathbf{z}'\mathbf{z})$ across each pair of subsequences for DP-HMM in Figure 7(a) and also show the same measure from HDP-HMM in Figure 7(b), in which the dynamic structure is removed from dHDP; other variables have the same definition as inferred via the DP-HMM and HDP-HMM. Compared with the result of dHDP in Figure 6(b), we observe a clear difference: although the DP-HMM can also tell the repetitive patterns occurring before the 42th subsequence, the HMM components shared during the whole piece jump from one to the other between the successive subsequences, which makes it difficult to segment the music and understand the development of the piece (*e.g.*, the slow solo part between the 53th and 69th subsequences is segmented into many small pieces in DP-HMM); similar performance is also achieved in the results of HDP-HMM (Figure 7(b)) and the music’s coherence nature is not observed in such modelings.

Additionally, we also compare the dHDP HMM with segmentation results produced by the iHMM (Beal et al., 2002; Teh et al., 2006). With the iHMM, the music is treated as one long sequence (all the subsequences are concatenated together sequentially) and a single HMM with an “infinite” set of states is inferred; in practice, a finite set of states is inferred as probable, as quantified in the state-number posterior. For the piece of music under consideration, the posterior on the number of states across the entire piece is as depicted in Figure 8(a). The inference was performed using MCMC, as in (Teh et al., 2006), with hyper-parameters consistent with the models discussed above.

With the MCMC, we have a state estimation of each observation (codeword, for our discrete-observation model). For each of the subsequences considered by the other models, we employ the posterior on the state distribution to compute the Kullback-Leibler (KL) divergence between every pair of subsequences. Since the KL divergence is not symmetric, we define the distance between two state distribution as:

$D = \frac{1}{2}\{E(D_{KL}(P_1||P_2)) + E(D_{KL}(P_2||P_1))\}$. Based on the collected samples, we use the averaged KL divergence to measure the similarity between any two subsequences and plot it in Figure 8(b). Although such a KL-divergence matrix is a little noisy, we observe a similar time-evolving sharing existing between adjacent subsequences, as inferred by the dHDP. This is because the iHMM also characterizes the music’s coherence since all of the sequential information is contained in one HMM. However, the inference of this relationship requires a post-processing step with the iHMM, while the dHDP infers these relationships as a direct aspect of the inference, also yielding “cleaner” results.

3.2 Model Quality Relative to Music Theory

The results of our computational analyses are compared with segmentations performed by a composer, musician and professor of music (third author). This music analysis is based upon reading the musical notes as well as listening to the piece being played. The music-theoretic analysis was performed *independent* of the numerical analysis (performed by the other authors), and then the relationship between the two analyses was assessed by the third author. We did *not* perform a numerical analysis and then subsequently interpret the results; the music analysis and numerical analyses were performed independently, and subsequently compared. The results of this comparison are discussed below.

For this comparison, the temporal resolution of the numerical analysis is increased; in the example presented below 15 discrete observations represent one second of music piece, each subsequence is again of length $T = 60$ (4 second subsequences), and for the Beethoven piece we now have $J = 125$ contiguous frames. All other parameters are unchanged. In Figure 9 it is observed that the model does a good job of segmenting the large sectional divisions found in sonata form: exposition, exposition repeat,

development, and recapitulation (discussed further below). Along the top of the figure, we note a parallel row of circles (white) and ellipses (yellow); these correspond to the “Largo” sections and are extracted well. The first two components of “Largo” (white circles) are exact repeats of the music, and this is reflected in the segmentation. Note that the yellow ellipses are still part of “Largo”, but they are slightly distinct from the yellow circles at left; this is due to the introduction of new key areas extended by recitative passages. The row of white squares correspond to the “main theme” (“Allegro”), and these are segmented properly. The parallel row of blue rectangles corresponds to the second key area, which abruptly changes the rhythmic and melodic texture of the music. Note that the first two of these (centered about approximately sequences 30 and 58) are exact repeats. The third appearance of this passage is in a different key, which is supported by the graph showing slightly lower similarity.

The row of three circles parallel to approximately sequence 30 corresponds to another sudden change in texture characterized by melodic neighbor tone motion emphasizing Neapolitan harmony (A-natural moving to B^b), followed by a harmonic sequence. The rightmost circle, in the recapitulation, is in a different key and consequently emphasizes neighbor motion on D-natural and E^b , and is still found similar to the earlier two appearances.

We also note that the A-natural / B^b neighbor motion is similar to subsequences near subsequence 20, and this may be because subsequence 20 also has strong neighbor tone motion (E-natural to F-natural) in the left hand accompaniment.

Finally, the bottom-right circle in Figure 9 identifies unique material that replaces the recapitulation of the main theme (“Allegro”), and its similarity to the main theme (around sequence 16) moves lower. The arrows at the bottom of Figure 19 identify “Allegro” interjections in the Largo passages, not all of which are in the same key.

3.3 Analysis of Mozart Piece

The above example examined the performance of the dHDP model relative to other competing statistical approaches, and to non-statistical (more-traditional) analysis performed by the third author. Having established the utility of dHDP relative to the other statistical approaches, we now only consider dHDP for the next example: Mozart-K. 333 Movement 1 (sampled with each frame 50 ms long, yielding for this case $J = 139$ subsequences). This is again entirely a piano piece. We now provide a more-complete sense of how the traditional musical analysis was performed.

Above we considered the first movement of Beethoven's Sonata No. 17, Op. 31, No. 2 (the Tempest), and below is considered the first movement of Mozart's Sonata K. 333. Classical sonata movements have a consistent approach to the presentation and repetition of themes as well as a clear tonal structure. The first movement of K. 333 by Mozart frequently appears in music anthologies used in undergraduate courses in music theory and history and often held up as a typical example of sonata form (Burkhart, 2003). The first movement of Op. 31, No. 2 by Beethoven is an example of the composer's self-conscious effort to expand the technical and expressive vocabulary of sonata form, and the music shows a remarkable interplay of convention and innovation.

A classical sonata movement is a ternary form consisting of an Exposition (usually repeated), a Development, and a Recapitulation. The Exposition is subdivided into distinct subsections: a first theme in the tonic key, a second theme in the key of the dominant (or relative major for minor key sonata movements), and a closing theme in the dominant (or relative major). A transition between the first and second themes modulates from the tonic key to the dominant. The closing theme may be followed by a coda to conclude the Exposition in the key of the dominant.

The Development typically draws on fragments from the Exposition themes for melodic material. These are recombined to construct sequential patterns which modulate freely (observing the conventions of Classical harmony). It is not unusual for entirely new themes to be introduced. In most cases, the Development ends with a retransition which extends dominant harmony in preparation for the return to tonic harmony. The Recapitulation presents the first theme again in the tonic key, a modified transition, the second theme, now in the tonic key instead of the dominant, followed by the closing theme and coda, all in the tonic key.

This patterned circulation of themes and key areas gives sonata form a pleasing predictability - the knowledgeable listener can anticipate what is going to happen next - as well as a built-in tension that results from a tonal structure that establishes the tonic key, departs for the dominant key, moves through passages of harmonic instability, and finally releases harmonic tension by a return to the tonic key.

3.3.1 TRADITIONAL ANALYSIS OF K. 333 BY W.A. MOZART

K. 333 closely follows the template described above. Measures 1-10 present the first theme in the tonic key (Bb major). Measures 10-22 present the transition based on the first theme, but modified in such a way that the music cadences on the dominant of the dominant. The second theme appears in the key of the dominant (F major) in mm. 23-30 and is restated in mm. 31-38. The closing theme follows in mm. 38-50, and mm. 50-63 comprise a coda which brings the Exposition to a conclusion in F major, the dominant key.

As is typical for a Mozart a sonata, the first and second themes are clearly distinguished from each other. The first theme is harmonically stable and maintains a consistent texture of melody and accompaniment. In contrast, the second theme juxtaposes several short thematic ideas that introduce dynamic and textural changes,

chromatic inflections, rhythmic syncopations, and virtuosic passage work. The closing theme is distinguished from both the first and second themes by an Alberti bass accompaniment in sixteenth notes and faster melodic motion.

The Development begins in m. 64 with a variation of the first theme in the key of F major. The theme cadences deceptively in the key of F minor in measure 71, which begins a new section cast in an improvisatory character that ends with a chromatic descent to the dominant of the submediant (V/vi) in measure 81. The retransition in mm. 87-93 abruptly introduces dominant harmony and prepares for the return to the tonic key of Bb major.

The Recapitulation begins in measure 94 with a restatement the first theme in the tonic key. Measures 94-103 are an exact restatement of measures 1-10. The transition follows in mm. 104-118. Like the corresponding passage in the Exposition, this passage is based on the first theme, however, it is extended to accommodate a harmonic excursion that cadences on the dominant. The second theme, also in the tonic key, follows in mm. 119-134. Aside from the transposition to the tonic key, this passage is nearly an exact repetition of mm. 23-38, with the restatement of the second theme played an octave higher in mm. 127-134. The closing theme in mm. 134-152 is now stated in the tonic key as expected, however, like the transition, it is extended by a harmonic sequence in mm. 143-146 and by the insertion of entirely new material in mm. 147-151. The coda in measures 152-165 is an exact repetition of mm. 50-63, except now transposed to the tonic key. The thematic/harmonic analysis is summarized in Figure 10.

Tracking themes and key areas is rather simple in K. 333 since it closely adheres to the sonata template. Such an exercise is a typical assignment in an undergraduate music theory course. A more subtle analysis focuses on contrapuntal design as well as on the use of chromaticism at different structural levels. For example, it

is entirely characteristic of Haydn, Mozart, and Beethoven to introduce chromatic melodic embellishments as local events which later serve as a contrapuntal or voice leading “scaffold” projected over many measures, or even over entire sections of a piece. This is seldom audible, even to a sophisticated listener, however, it is a central aspect of compositional technique in the Classical period, one that creates a sense of continuous, organic development across sectional divisions. K. 333 offers an excellent example of this technique³.

The closing theme and coda in the Exposition introduce a chromatic melodic descent based the pitches F-E-Eb-D. The use of chromaticism for local color has been a prominent feature of the second theme, and thus the appearance of the chromatic descent in the closing theme does not seem unusual. The chromatic figure can be seen and heard in mm. 46-47, 50-51, 54-55, and 59-62. The same chromatic descent appears twice in the Development section, the first time projected over measures 64-68, and the second time projected over measures 71-81, the improvisatory passage in the key of F minor. Thus, what appeared to be entirely new music in the Development (mm. 71 ff.) is actually derived from the chromatic melodic descent introduced in the Exposition. This is a perfect example of unity underlying variety.

A successful dHDP analysis of K. 333 should segment the music in a way that corresponds to sectional divisions of sonata form. Since our performance repeats the Exposition, we would expect dHDP to show strong similarity between the two statements of the first theme, transition, second theme, closing theme, and coda. The Recapitulation presents an interesting challenge. While all thematic materials from the Exposition appear in the Recapitulation, everything from the transition to the end is stated in the tonic key instead of the dominant key. In other words,

3. Analysis of contrapuntal and chromatic details at multiple structural levels was developed by the German theorist, Heinrich Schenker (1868-1935)

the Recapitulation has strong melodic similarity to the Exposition, but the notes are different. The Development offers another challenge. While this section begins with a variation of the first theme, the improvisation that follows is (seemingly) entirely new music. If anything, dHDP analysis might show the similarity of the improvisation to the closing theme because both passages make use of Alberti bass figuration in sixteenth notes. A truly remarkable analysis would catch the projection of chromatic details over long passages in the Development section.

3.3.2 SEGMENTATION BY DHDP ANALYSIS OF K. 333

Before beginning the analysis of Figure 11, it should be emphasized that precise linkage between music-theoretic analysis and statistical analysis is difficult, since for the latter the music is divided into a series of contiguous 4 second blocks (these blocks do not in general line up precisely with music-theoretic segments in the music). This makes detailed analysis of some passages more difficult, particularly when several small segments appear in close succession. Having said this, dHDP analysis segments the music appropriately (based on the third author).

Considering the annotations in Figure 11, the vertical arrow at the bottom identify unaccompanied melodic transitions in the right hand or sudden changes to soft dynamics, which are generally distinguished by the dHDP. The first row of white circles (near the top) correspond to the beginning of the second theme, characterized by the distinctive chordal gesture in the key of the dominant, and this decomposition or relationship appears to be accurate. We note that the third appearance of this gesture in the recapitulation is in a different key, and the similarity is correspondingly lower. An example of an “error” is manifested in the row of white rectangles. These correspond to the closing theme, and the left two rectangles (high correlation between each) are correct, but the right rectangle does not have a corresponding high

correlation inside; it is therefore not recognized in the recap, when it appears in a different key (tonic). The results in Figure 11 show a repeated high degree of similarity that is characteristic of Mozart piano sonatas; the consistent musical structure is occasionally permeated by exquisite details, such as a phase transition (these, again, identified by the arrows at the bottom).

The large sectional divisions between the Exposition, Development, and Recapitulation are easily seen in Figure 11. This figure also marks the beginnings of the first theme, second theme, closing theme, and coda within the Exposition. The beginning of the transition section is not distinguished from the first theme in Figure 11, despite the clear cadence that separates the first theme and transition. On the other hand, dHDP isolates a brief passage that occurs in the middle of the transition (m. 14, beat 4 - m. 16). This passage is characterized by a sudden change in dynamics and register. Other examples of local segmentation appear at the end of the transition and the beginning of the second theme (mm. 22-23), when the right hand is unaccompanied by the left. Here Figure 11 shows a prominent orange band denoting less similarity with the music immediately preceding and following this passage, which is entirely consistent with the musical texture. The figure marks the restatement of the second theme (m. 31) and isolates the final measures of the coda when the musical texture thins out at the Exposition cadence. The sudden change of texture and dynamics within the closing theme (mm. 46-48) is clearly separated from the main part of the closing theme in the figure. Even smaller segments comprising a few notes are marked. These segments isolate moments between phrases when the right hand plays quietly, unaccompanied by the left hand. The dHDP analysis of the Exposition repeat precisely replicates the segmentation described above.

The Development is represented as a single block, though the beginning of the improvisatory section in F minor (m. 71) appears to be marked by a prominent green

band, indicating less similarity with the music immediately preceding and following this moment. Figure 11 marks the retransition with several small segments, however, the resolution of the figure makes it difficult to correlate these segments with particular moments in the music. Figure 11 clearly marks the Recapitulation with its return to the first theme in the tonic key. As before, the beginning of the transition goes unnoticed, however, dHDP again segments the transition passage associated with a sudden change in register and dynamics (mm. 110, beat 4 -112).

The end of the transition and beginning of the second theme (mm. 118-119) is marked by a prominent orange/yellow band (Figure 11) indicating less similarity, just as was seen at the same moment in the Exposition (mm. 22-23). The figure does not mark the restatement of the second theme as it did in the Exposition, however, this may be a consequence of misalignment between the music playback and the analysis, as discussed above. The closing theme is segmented appropriately, and the sudden change of texture and dynamics in mm. 142-146 is segmented apart from the rest of the closing theme, just as we saw in mm. 46-48 in the Exposition. Note that Figure 11 clearly shows this passage has been extended to five measures in the Recapitulation compared to three measures in the Exposition. The figure segments the coda in the same way we saw in the Exposition, including its isolation of the final cadence.

In sum, dHDP analysis has segmented the music remarkably well. Parallel passages which appear throughout the movement are represented the same way each time they occur. Even the omissions are consistent, such as the lack of segmentation of the transition from the first theme. The results are summarized in Figure 12.

3.3.3 QUALITY OF SIMILARITY DEFINED BY dHDP ANALYSIS OF K. 333

The dHDP analysis shows a high degree of similarity of most thematic materials in the movement. For example, the first theme, transition, second theme, and coda

are all marked with the highest degree of similarity to each other across the entire movement. The dHDP analysis does not appear to recognize the differences in note successions in these passages.

Figure 11 does indicate moments of dissimilarity. For example, the closing theme (beginning in m. 38) is marked as dissimilar from anything else in the movement. Recall that the closing theme introduced a new Alberti bass accompaniment in sixteenth notes which helps set this music apart. However, the reappearance of the closing theme in the Recapitulation is not represented as similar to the closing theme in the Exposition. Perhaps the transposition of the closing theme to the tonic key in the Recapitulation obscures the similarity, but this does not explain why the closing theme in the Recapitulation is marked as highly similar to the first and second themes, transition, and development throughout the rest of the movement, no matter what key they are in.

Several incidental details are marked with a high degree of similarity to each other while being dissimilar to the rest of the movement. These are normally moments when the music suddenly becomes quiet or features isolated groups of notes played by the right hand without accompaniment. Examples of this can be seen along the horizontal axis at the very top of Figure 11 and include the pickups to m. 1, m. 46, the pickups to m. 64, the pickups to m. 94, mm. 118-119, m. 130, mm. 142-146, and the final cadence in m. 165.

Finally, we observe that dHDP analysis is not suitable for revealing the projection of chromatic details at a larger structural level. We note, however, that dHDP analysis did mark the first appearance of the descending chromatic melodic figure in mm. 46-48 of the closing theme.

From these results we may suppose that similarity in dHDP analysis is more strongly associated with dynamics, texture, and register than with melody and har-

mony. This raises an important point. While dynamics can be specified in the musical score, it is up to the musician to interpret these markings in performance. It is possible that dHDP analysis would represent another interpretation of the same piece differently.

For brevity, we only provide the detailed music-theoretic analysis, with comparison to dHDP, for the Mozart piece. However, the same detailed analysis was used to yield the conclusions above with respect to the Beethoven piece. That analysis is provided online as supplemental material⁴. We reiterate that the music-theoretic analysis of the type summarized in Figure 10 was performed *independent* of the statistical analysis, with comparisons performed subsequently.

4. Conclusions

The dynamic hierarchical Dirichlet process (dHDP) has been developed for analysis of sequential data, with a focus on analysis of audio data from music. The framework assumes a parametric representation $F(\boldsymbol{\theta})$ to characterize the statistics of the data observed at a single point in time. The parameters $\boldsymbol{\theta}$ associated with a given point in time are assumed to be drawn from a mixture model, with in general an infinite number of atoms, analogous to the Dirichlet process. The mixture models at time $t-1$ and time t are inter-related statistically. The model is linked to the hierarchical Dirichlet process (Teh et al., 2006) in the sense that the initial mixture model and the subsequent time-dependent mixtures are drawn from the same discrete distribution. This implies that the underlying atoms in the $\boldsymbol{\theta}$ space associated with the aforementioned mixtures are the same, and what is changing with time are the mixture weights. The model has the following characteristics: (*i*) with inferred probabilities, the underlying

4. <http://sites.google.com/site/bayesianmusic/>

parameters associated with data at adjacent times are the same; and *(ii)* since the same underlying atoms are used in the mixtures at all times, it is possible that the same atoms may be used at temporally distant time, allowing the capture of repeated patterns in temporal data. The underlying sharing properties (correlations) between observations at adjacent times have also been derived. Inference has been performed in an MCMC setting.

Examples have been presented on three musical pieces: a relatively simple piece from the Beatles, as well as two more complicated classical pieces. The classical pieces are more difficult to analyze because there are no vocals, and a single instrument is generally used, and therefore the segmentation of such data is more subtle. The results of the classical-piece segmentations have been analyzed for their connection to music analysis. In this connection it is felt that the results are promising. While there were mistakes in the analysis of the Beethoven and Mozart pieces considered, there is a great deal of accuracy as well. The results clearly reveal meaningful characteristics about Beethoven and Mozart.

The dHDP analysis effectively segments the two classical compositions by Mozart and Beethoven at both the large-scale and local levels. Segmentation appears to be related to musical dynamics, texture, and register. The dHDP analysis of similarity is far more successful in the Beethoven sonata than in the work by Mozart. It may be that the greater variety of musical textures, dynamics, and registral placement in Op. 31, No. 2 yield more gradations of similarity in the graph. The dHDP analysis makes several plausible similarity connections, though there are inconsistencies as well. The greatest deficiency in the dHDP analysis of similarity is the apparent inability to track note successions (*i.e.*, themes) and key areas as a basis for comparison.

Despite these shortcomings, dHDP analysis is instructive for musicians, perhaps especially so for composers (these are observations of the third author, who is a

composer and musician). In K. 333, Mozart articulates form through themes and tonal structure. Beethoven articulates form in Op. 31, No. 2 through themes that are linked to emphatic gestures, as well as through a detailed tonal design. This is not to say that one is better than the other. There are works by Beethoven that will likely yield results similar to K. 333, and Mozart has composed works that may result in results that are as varied as the results of Op. 31, No. 2. Nonetheless, dHDP analysis of K. 333 and Op. 31, No. 2 illustrates general tendencies of the two composers that are commonly acknowledged by musicians and audiences alike.

Concerning future research, for large data sets the MCMC inference engine employed here may not be computationally tractable. The graphical form of the dHDP is applicable to more-approximate inference engines, such as a variational Bayesian (VB) analysis (Blei and Jordan, 2004). We intend to examine VB inference in future studies, and to examine its relative advantages in computational efficiency compared to its inference accuracy (relative to MCMC). Additionally, our model was motivated by a stick-breaking construction of DP; however, it is also of interest to consider a Chinese restaurant/franchise Teh et al. (2006) representation, which may have advantages for interpretation and inference.

Appendix

Proof of Theorem 2.

According to (8), $G_j = w_{j1}G_1 + \sum_{l=2}^j w_{jl}H_{l-1}$, where $w_{jl} = \tilde{w}_{l-1} \prod_{m=l}^{j-1} (1 - \tilde{w}_m)$.

Then given $\{w_{j1}, \dots, w_{jj}\}$ and the base distribution H , the expectation of G_j is:

$$\begin{aligned} E\{G_j(B)\} &= w_{j1}E\{G_1(B)\} + \sum_{l=2}^j w_{jl}E\{H_{l-1}(B)\} \\ &= \sum_{l=1}^j w_{jl}H(B) \end{aligned} \quad (19)$$

Since given G_0 , the variance of $G_j(B)$ is: $V\{G_j(B)|G_0(B)\} = \sum_{l=1}^j (\frac{w_{jl}^2}{\alpha_{0l}+1})G_0(B)\{1 - G_0(B)\}$. Then we can get the variance of $G_j(B)$ with the expectation of $G_0(B)$ as follows:

$$\begin{aligned} V\{G_j(B)\} &= E\left[V(G_j(B)|G_0(B))\right] + V\left[E(G_j(B)|G_0(B))\right] \\ &= E\left[\sum_{l=1}^j \left(\frac{w_{jl}^2}{\alpha_{0l}+1}\right)G_0(B)(1 - G_0(B))\right] + V\left[\sum_{l=1}^j w_{jl}G_0(B)\right] \\ &= \sum_{l=1}^j \frac{w_{jl}^2}{\alpha_{0l}+1}E\left[G_0(B) - G_0^2(B)\right] + V\left[\sum_{l=1}^j w_{jl}G_0(B)\right] \\ &= \sum_{l=1}^j \frac{w_{jl}^2}{\alpha_{0l}+1}\left[H(B) - (V(G_0(B)) + H^2(B))\right] + \sum_{l=1}^j w_{jl}^2V[G_0(B)] \quad (20) \\ &= \sum_{l=1}^j w_{jl}^2\left[\left(1 - \frac{1}{1 + \alpha_{0l}}\right)V[G_0(B)] + \frac{H(B)[1 - H(B)]}{1 + \alpha_{0l}}\right] \\ &= \sum_{l=1}^j w_{jl}^2\left[\frac{\alpha_{0l}}{1 + \alpha_{0l}} \cdot \frac{1}{1 + \gamma}H(B)[1 - H(B)] + \frac{H(B)[1 - H(B)]}{1 + \alpha_{0l}}\right] \\ &= \sum_{l=1}^j \frac{w_{jl}^2}{1 + \alpha_{0l}}\left(\frac{\alpha_{0l} + \gamma + 1}{1 + \gamma}\right)H(B)[1 - H(B)] \end{aligned}$$

Additionally given G_0 we can get:

$$\begin{aligned}
& E\{G_j(B)G_{j-1}(B)\} - E\{G_j(B)\}E\{G_{j-1}(B)\} \\
&= E\left[\{w_{j1}G_1(B) + \dots + w_{jj}H_{j-1}(B)\}\{w_{j-1,1}G_1(B) + \dots + w_{j-1,j-1}H_{j-2}(B)\}\right] \\
&\quad - E\{w_{j1}G_1(B) + \dots + w_{jj}H_{j-1}(B)\}E\{w_{j-1,1}G_1(B) + \dots + w_{j-1,j-1}H_{j-2}(B)\} \\
&= w_{j1}w_{j-1,1}V\{G_1(B)\} + \sum_{l=2}^{j-1} w_{jl}w_{j-1,l}V\{H_{l-1}(B)\} \\
&= \sum_{l=1}^{j-1} \frac{w_{jl}w_{j-1,l}}{1 + \alpha_{0l}} \cdot \frac{\alpha_{0l} + \gamma + 1}{1 + \gamma} H(B)[1 - H(B)]
\end{aligned} \tag{21}$$

From the above analysis, the correlation coefficient of the distributions between the adjacent groups defined in (13) can be formularized as follows:

$$\text{Corr}(G_{j-1}(B), G_j(B)) = \frac{\sum_{l=1}^{j-1} \frac{w_{jl}w_{j-1,l}}{1 + \alpha_{0l}} \cdot (\alpha_{0l} + \gamma + 1)}{\left[\sum_{l=1}^j \frac{w_{jl}^2}{1 + \alpha_{0l}} \cdot (\alpha_{0l} + \gamma + 1) \right]^{1/2} \left[\sum_{l=1}^{j-1} \frac{w_{j-1,l}^2}{1 + \alpha_{0l}} \cdot (\alpha_{0l} + \gamma + 1) \right]^{1/2}} \tag{22}$$

References

- J. J. Aucouturier and F. Pachet. The influence of polyphony on the dynamical modelling of musical timbre. *Pattern Recognition Letters*, 28(5):654–661, 2007.
- K. Barnard, P. Duygulu, D. Forsyth, N.D. Freitas, D.M. Blei, and M.I. Jordan. Matching words and pictures. *Journal of Machine Learning Research*, 3:1107–1135, 2003.
- M.J. Beal, Z. Ghahramani, and C.E. Rasmussen. The infinite hidden Markov model. In *Neural Information Processing Systems(NIPS)*, 2002.

- D. Blackwell and J.B. MacQueen. Ferguson distributions via Polya urn Schemes. *Ann. Statist.*, 1(2):353–355, 1973.
- D.M. Blei and M.I. Jordan. Variational methods for the Dirichlet process. In *Proceedings of the International Conference on Machine Learning*, 2004.
- D.M. Blei and J.D. Lafferty. Dynamic topic models. In *Proceedings of the International Conference on Machine Learning*, 2006.
- C. Burkhardt. *Anthology for Music Analysis*. Schirmer Books, 2003.
- C.A. Bush and S.N. MacEachern. A semiparametric Bayesian model for randomised block designs. *Biometrika*, 83(2):275–285, 1996.
- Y. Chung and D.B. Dunson. The local Dirichlet process. *Annals of the Institute of Statistical Mathematics*, to appear, 2009.
- D.B. Dunson. Bayesian dynamic modeling of latent trait distributions. *Biostatistics*, 7(4):551–568, 2006.
- T.S. Ferguson. A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1:209–230, 1973.
- E.B. Fox, E.B. Sudderth, M.I. Jordan, and A.S. Willsky. An HDP-HMM for systems with state persistence. *Proc. 25th International Conference on Machine Learning (ICML)*, 2008.
- A. Gersho and R.M. Gray. *Vector Quantization and Signal Compression*. Springer, 1992.
- J. Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Stat.*, 4:169–193, 1992.

- J.E. Griffin and M.F.J. Steel. Order-based dependent Dirichlet processes. *Journal of the American Statistical Association*, 101:179–194, 2006.
- H. Ishwaran and L.F. James. Gibbs sampling methods for Stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173, 2001.
- J.H. Jensen, M.G. Christensen, M.N. Murthi, and S.H. Jensen. Evaluation of MFCC estimation techniques for music similarity. In *Proceedings of the 14th European Signal Processing*, 2006.
- B. Logan. Mel frequency cepstral coefficients for music modeling. In *International Symposium on Music Information Retrieval*, 2000.
- S.N. MacEachern. Dependent nonparametric process. In *ASA Proceeding of the Section on Bayesian Statistical Science*, 1999.
- M. Medvedovic and S. Sivaganesan. Bayesian infinite mixture model based clustering of gene expression profiles. *Bioinformatics*, 18(9):1194–1206, 2002.
- K. Ni, J. Paisley, L. Carin, and D. Dunson. Multi-task learning for analyzing and sorting large databases of sequential data. *IEEE Trans. Signal Processing*, 56:3918–3931, 2008.
- J.-F. Paiement, Y. Grandvalet, S. Bengio, and D. Eck. A generative model for rhythms. *NIPS'2007 Music, Brain & Cognition Workshop*, 2007.
- T. Plotz, G.A. Fink, P. Husemann, S. Kanies, K. Lienemann, T. Marschall, M. Martin, L. Schillingmann, M. Steinrucken, and H. Sudek. Automatic detection of song changes in music mixes using stochastic models. *18th International Conference on Pattern Recognition (ICPR'06)*, 2006.

- Y. Qi, J.W. Paisley, and L. Carin. Music analysis using hidden Markov mixture models. *IEEE Transactions on Signal Processing*, 55(11):5209–5224, 2007.
- L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- A.E. Raftery and S. Lewis. How many iterations in the Gibbs sampler? *Bayesian Stat.*, 4:763–773, 1992.
- C. Raphael. Automatic segmentation of acoustic musical signals using hidden Markov models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):360–370, 1999.
- A. Rodriguez, D.B. Dunson, and A.E. Gelfand. The nested Dirichlet process (with discussion). *Journal of the American Statistical Association*, 103:1131–1144, 2008.
- Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei. Hierarchical Dirichlet processes. *JASA*, 101(476):1566–1581, 2006.
- D. Temperley. *Music and Probability*. MIT Press, 2007.
- D. Temperley. A probabilistic model of melody perception. *Cognitive Science*, 32:418–444, 2008.
- X. Wei, J. Sun, and X. Wang. Dynamic mixture models for multiple time series. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2007.



Figure 1: The audio waveform of the Beatles music.

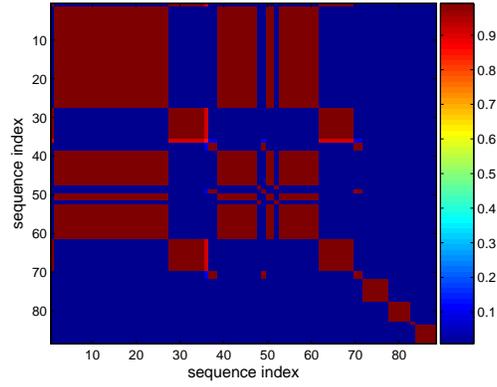


Figure 2: Segmentation of the audio waveform in Figure 1.

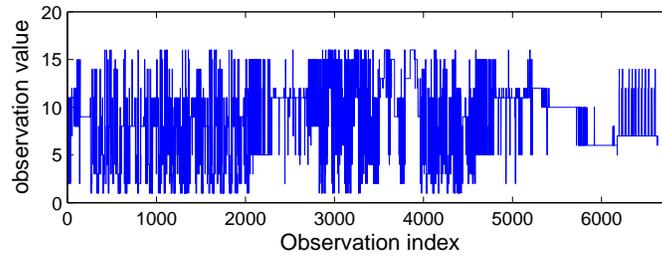
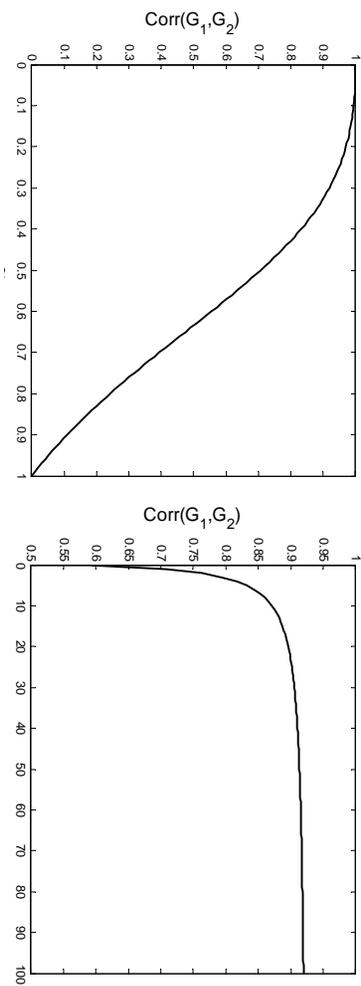
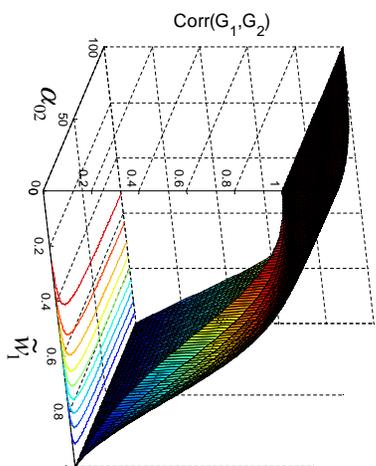


Figure 3: Sequence of code indices for the waveform in Figure 1, using a codebook of dimension $M = 16$.



(a)



(c)

Figure 4: (a) $Corr(G_1, G_2)$ as a function of \tilde{w}_1 with γ and α fixed. (b) $Corr(G_1, G_2)$ as a function of α_{02} , with γ , α_{01} and \mathbf{w} fixed. (c) $Corr(G_1, G_2)$ as a function of both \tilde{w}_1 and α_{02} , with the values of γ and α_{01} fixed.

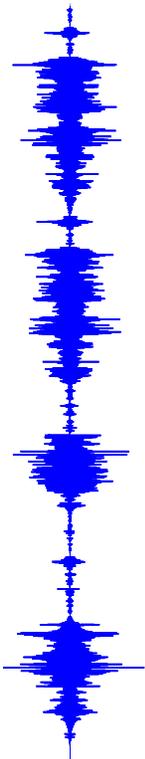


Figure 5: Audio waveform of the first movement of Op. 31, No. 2.

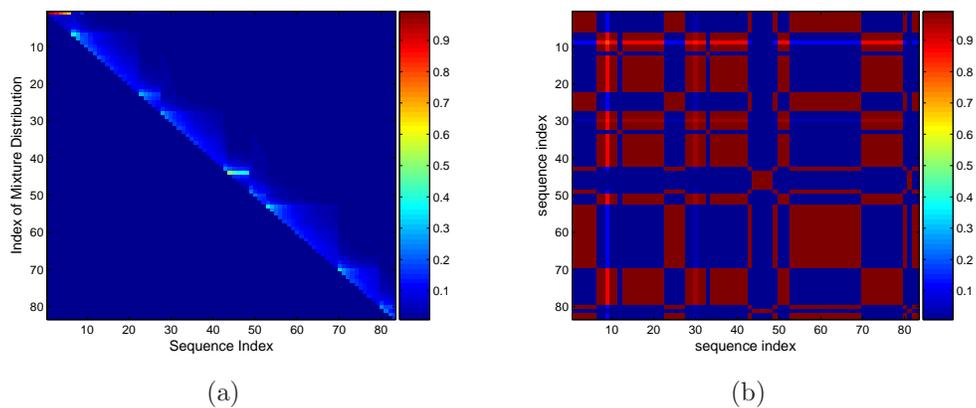


Figure 6: Results of dHDP HMM modeling for the Sonata No.17. (a) The posterior distribution of indicator variable \mathbf{r} . (b) The similarity matrix $E[\mathbf{z}'\mathbf{z}]$.

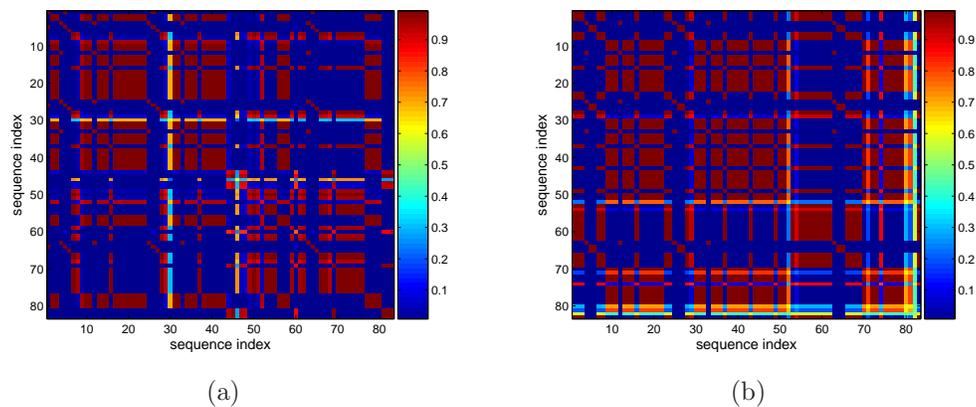
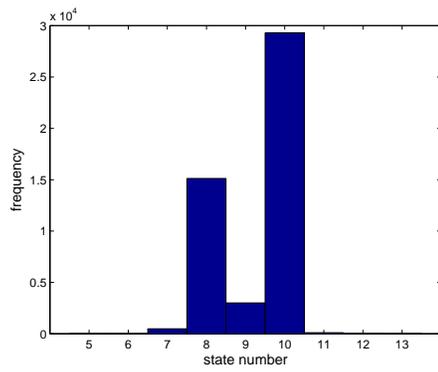
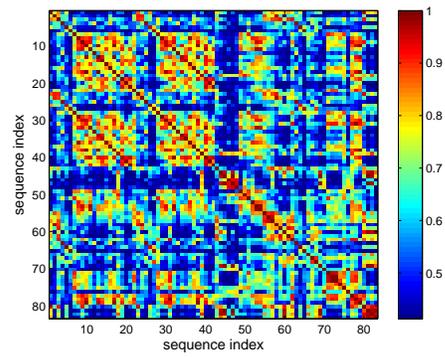


Figure 7: Results of DP-HMM and HDP-HMM mixture modeling for the Sonata No.17. (a) The similarity matrix $E(\mathbf{z}'\mathbf{z})$ from DP-HMM result. (b) The similarity matrix $E(\mathbf{z}'\mathbf{z})$ from HDP-HMM result.



(a)



(b)

Figure 8: Analysis results for the piano music based on the iHMM. (a) Posterior distribution of state number. (b) Approximate similarity matrix by KL-divergence.

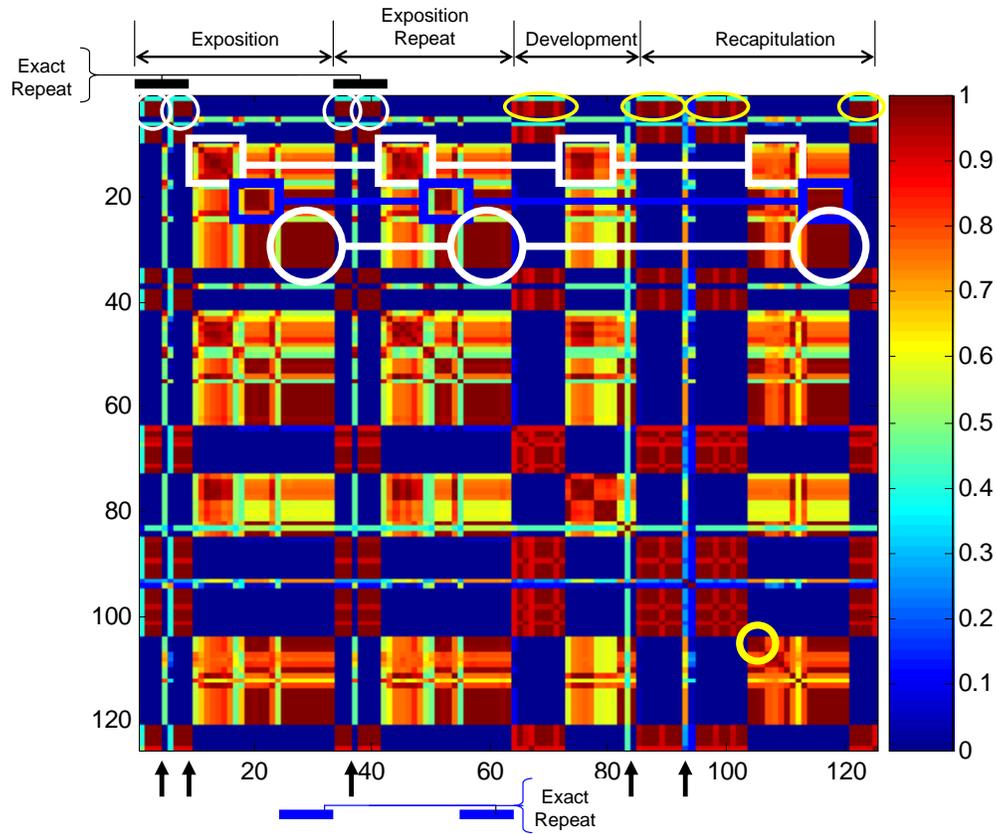


Figure 9: Annotated $E(\mathbf{z}'\mathbf{z})$ for the Beethoven Sonata. The description of the annotations are provided in the text. The numbers along the vertical and horizontal axes correspond to the sequence index, and the color bar quantifies the similarity between any two segments.

Sonata for Piano, K. 333, First Movement

Section	Key Area	Measure
Exposition		
First Theme	Tonic (Bb major)	1-10
Transition	Tonic modulates to Dominant (F major). Cadences on V/V.	10-22
Second Theme	Dominant (F major)	23-30
Second Theme restated	Dominant	31-38
Closing Theme	Dominant	38-50
Coda	Dominant	50-63
64-93		
Development		
First Theme variation	Dominant	64-71
Improvisatory section	Dominant minor (F minor) ending on V/vi	71-86
Retransition	Extends V	87-93
94-165		
Recapitulation		
First Theme	Tonic (Bb major)	94-103
Transition (extended)	Tonic	103-118
Second Theme	Tonic	119-126
Second Theme restated	Tonic	127-134
Closing Theme	Tonic	134-152
Coda	Tonic	152-165

Figure 10: Summary of the traditional musical analysis of Sonata for Piano, K. 333, First Movement.

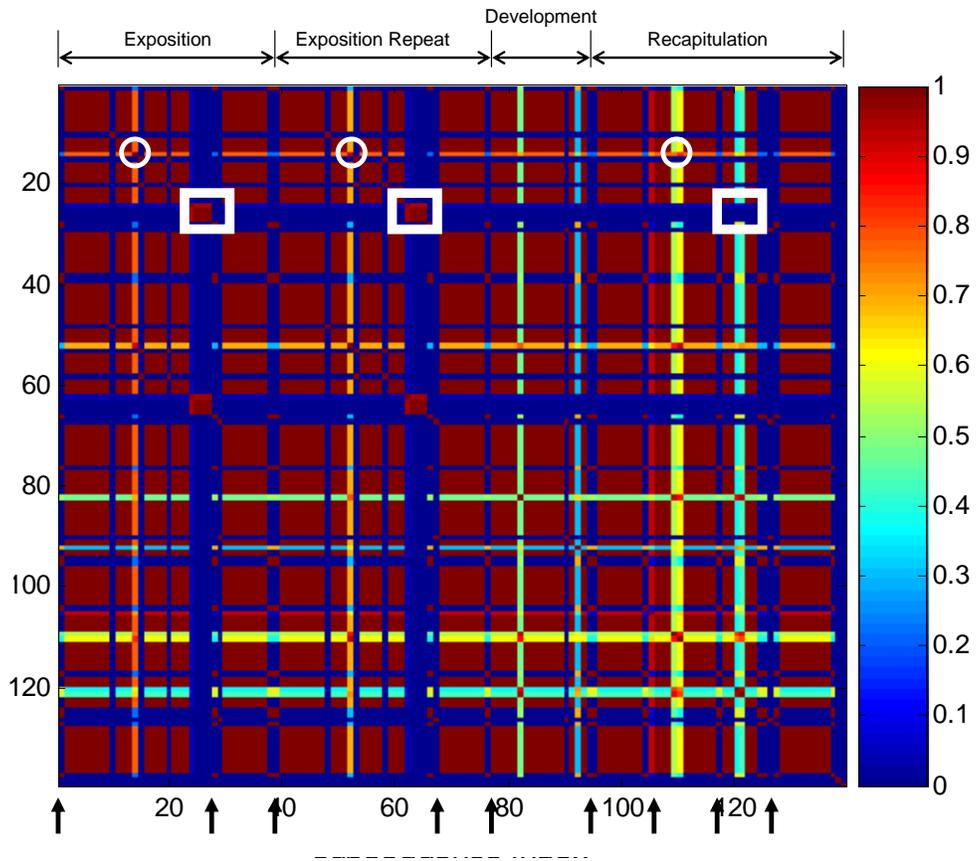


Figure 11: Annotated $E(\mathbf{z}'/\mathbf{z})$ for the Mozart. The description of the annotations are provided in the text. The numbers along the vertical and horizontal axes correspond to the sequence index, and the color bar quantifies the similarity between any two segments.

dHDP Segmentation of K. 333

Conventional Analysis	dHDP Analysis	Measure Numbers
Exposition		
First Theme	Segment	1
Transition	No segment	10
(Texture change in Transition)	Segment	14, beat 4-16
(Dissimilarity of unaccompanied R.H.)	Segment	22-23
Second Theme	Segment	23
Second Theme restatement	Segment	31
Closing Theme	Segment	38
(Texture change in Closing Theme)		46
Coda	Segment	50
	Final cadence	63
Development		
Variation of First Theme	Segment	64
Improvisatory section in Fm	Segment?	71
Retransition	Several small segments	87-93
Recapitulation		
First Theme	Segment	94
Transition	No segment	103
(Texture change in Transition)	Segment	110, beat 4 -112
(Dissimilarity of unaccompanied R.H.)	Segment	118-119
Second Theme	Segment	119
Second Theme restatement	No segment	127
Closing Theme	Segment	134
(Texture change in Closing Theme)	Segment	142
Coda	Segment	152
(Final cadence)	Segment	165

Figure 12: Summary of the dHDP analysis of Sonata for Piano, K. 333, First Movement.