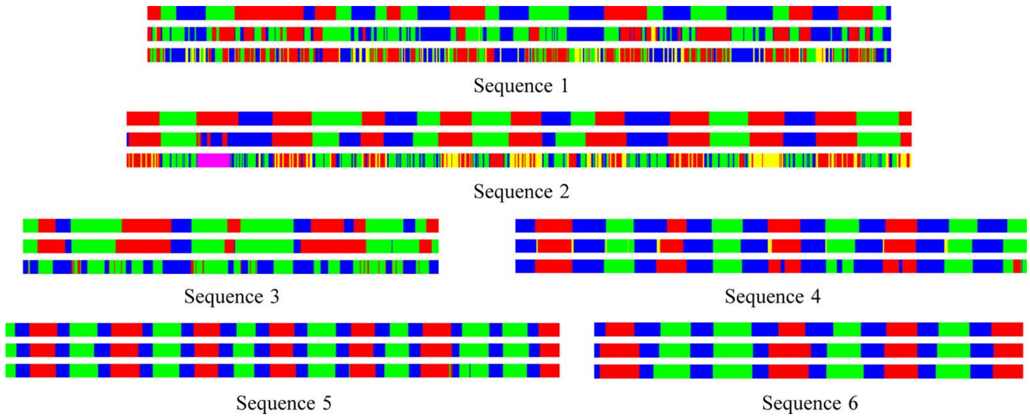

Fig. 6. Estimated mode sequences at the 1000th Gibbs iteration representing the median error over 100 trials. For each sequence we plot the true labels (top), labels from the partially supervised HDP-VAR-HMM (middle) and unsupervised HDP-VAR-HMM (bottom). Colors are as in Fig. 4.

TABLE III
MEDIAN LABEL ACCURACY OF THE HDP-VAR(1)-HMM USING UNSUPERVISED AND PARTIALLY SUPERVISED GIBBS SAMPLING, COMPARED TO ACCURACY OF THE SUPERVISED SLDS DATA-DRIVEN MCMC (DD-MCMC) *MAP* SEGMENTATION AND PS-SLDS APPROXIMATE *VITERBI* SEGMENTATION PROCEDURES OF OH *et al.* [10]

Sequence	1	2	3	4	5	6
HDP-VAR(1)-HMM unsupervised	45.0	42.7	47.3	88.1	92.5	88.2
HDP-VAR(1)-HMM partially supervised	55.0	86.3	81.7	89.0	92.4	89.6
SLDS DD-MCMC	74.0	86.1	81.3	93.4	90.4	90.2
PS-SLDS approx. Viterbi	75.9	92.4	83.1	93.4	91.0	90.4

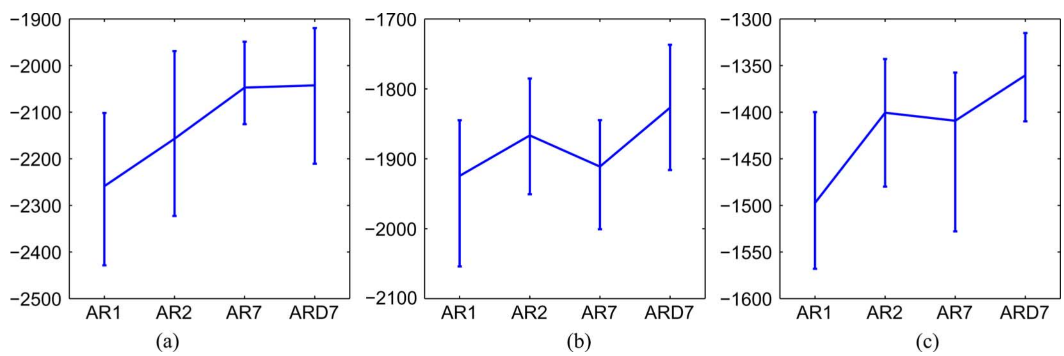

Fig. 7. For an order 1, 2, and 7 HDP-AR-HMM with a MNIW prior and an order 7 HDP-AR-HMM with an ARD prior, we plot the shortest intervals containing 95% of the held-out log-likelihoods calculated based on a set of Gibbs samples taken at iteration 1000 from 100 chains. (a) Log-likelihood of the second half of honey bee dance sequence 4 based on model parameters inferred from the first half of the sequence. (b)-(c) Similarly for sequences 5 and 6, respectively.

switching between dance modes is more irregular. This dramatically affects our performance since we do not use domain-specific information. For sequence 2 in particular, our learned segmentations often create new, sequence-specific waggle dance modes contributing to our calculated Hamming distance errors on this sequence. Overall, however, we are able to achieve reasonably good segmentations without having to manually input domain-specific knowledge.

2) MNIW Prior—Partially Supervised: The discrepancy in performance between our results and the supervised approach of Oh *et al.* [10] motivated us to also consider a partially supervised variant of the HDP-VAR(1)-HMM in which we fix the ground truth mode sequences for five out of six of the sequences and jointly infer both a combined set of dynamic parameters and the left-out mode sequence. This is equivalent to informing the prior distributions with the data from the five fixed sequences and using these updated posterior distributions as the prior distributions for the held-out sequence. As we see in Table III and the segmentations of Fig. 6, this partially supervised approach considerably improves performance for these three sequences, especially sequences 2 and 3. In this analysis, we hand-aligned sequences so that the waggle dances tended to have head angle measurements centered about $\pi/2$ radians. Aligning the waggle dances is possible by looking at the high frequency portions of the head angle measurements. Additionally, the pre-processing of the unsupervised approach is not appropriate here as the scalings and shiftings are dance-specific and such transformations modify the associated switching VAR(1) model. Instead, to account for the varying frames of reference (i.e., point of origin for each bee body) we allowed for a mean $\boldsymbol{\mu}^{(k)}$ on the process noise and placed an independent $\mathcal{N}(0, \Sigma_0)$ prior on this parameter. See the Appendix for details on how the hyperparameters of these prior distributions are set.

3) ARD Prior: Using the cleaner sequences 4 to 6, we investigate the affects of the sparsity-inducing ARD prior by assuming a higher order switching VAR model and computing the likelihood of the second half of each dance sequence based on parameters inferred from Gibbs sampling using the data from