

[NIPS 2016](#)**Neural Information Processing Systems 2016**

December 05 - 10, 2016, Barcelona, Spain

Reviews For Paper**Paper ID** 1016**Title** An Infinite Hidden Markov Model With Local Transitions**Masked Reviewer ID:** Assigned_Reviewer_1**Review:****Question**

Summary: Provide a brief summary of the contents of the paper.	This paper presents a Bayesian nonparametric hidden Markov model where the state transition probabilities can be biased towards states that are similar according to a kernel function. The authors develop an auxiliary Gibbs sampler by recasting the problem as a Markov jump process in continuous time and apply the model and algorithm to the speaker diarization problem where there are groups of individuals. They showed that the proposed model achieves higher accuracy than existing models.
Fatal flaws: Does the paper have a "fatal flaw" making it unfit for publication, regardless of other criteria (may include out of scope, double publication, plagiarism, wrong proofs, flawed experiments)? Use the text box to justify your answer.	No (not as far as I can see)
Technical quality: whether experimental methods are appropriate, proofs are sound, results are well analyzed.	3-Poster level (top 30% submissions)
Novelty/originality: in any aspect of the work, theory, algorithm, applications, experimental.	3-Poster level (some notable novel contributions)

Potential impact or usefulness: could be societal, academic, or practical and should be lasting in time, affecting a large number of people and/or bridge the gap between multiple disciplines.	3-Poster level (looks promising)
Clarity and presentation: explanations, language and grammar, figures, graphs, tables, proper references.	3-Poster level (good enough)
Qualitative assessment: Provide constructive feedback to the authors; justify and complement your ratings above. This is our MOST IMPORTANT QUESTION, we need to get good arguments for our decisions!	<p>This paper presented interesting ideas for extending the modeling power of the HDP-HMM model by inducing higher transition probabilities between states that are "similar". Overall, the paper is well written with nicely executed experiments that demonstrate both the efficacy of the proposed model as well as when it is weaker than the HDP-HMM (i.e. when the assumed local structure doesn't exist) and show that the algorithm can detect this and the user can switch to a simpler HDP-HMM model. I think that this paper will fit into the NIPS proceedings nicely.</p> <p>I do have some comments that I think will improve the quality of the paper. First, there are very few references. I think that including some reference(s) for Markov jump processes would be useful. Second, the authors reference the discrete infinite-logistic normal paper by Paisley, et. al., but recently this work was made more general by the correlated random measure framework of Ranganeth, et. al. The authors should discuss how their work relates to correlated random measures in general. Lastly, the authors are essentially constructing a set of random measures that share atoms via a hierarchy but that depend on the locations of the atoms, this immediately makes me think of dependent random measures and it would be great to see a comment about how the present work relates to dependent random measures.</p> <p>The authors do a bit of work to describe the properties of the local transition measures that they construct, however, a more in depth discussion that the constructed random measures are in fact valid random measures is definitely necessary (even in the supplement). In particular, the authors do need to show that T_j in Eq. (8) is finite and that a countably infinite number of atoms can be generated. See the correlated random measures paper I referenced above for detailed proofs of these properties in their constructions (one of which looks very similar to the π_{jj}'s in Eq. (8)).</p>

	Finally, for ease of reading the authors should carefully proof-read the document as there is some notation in the sampler that is not explicitly defined and in addition the authors should provide all conditional distributions used in the Gibbs sampler.
Reviewer confidence regarding this review.	3-Expert (read the paper in detail, know the area, quite certain of my opinion)

Masked Reviewer ID: Assigned_Reviewer_2

Review:

Question

Summary: Provide a brief summary of the contents of the paper.	<p>This paper builds upon the HDP-HMM (aka IHMM). It allows incorporating prior knowledge, namely preferred transitions between "similar" states. This preference is expressed, at the model design stage, as a symmetric similarity function of two states, with values between 0 and 1.</p> <p>The paper derives a Gibbs sampling algorithm for inference. To achieve this, it introduces auxiliary variables, which arise from considering the hidden state sequence as generated from a Markov jump process in which jumps can fail; an attempted jump is likely to fail if states are dissimilar.</p> <p>Experiments show that the model performs reasonably on ad hoc synthetic data (ie data where we expect state similarity to play a role), and on data generated from an HDP-HMM without the "local transition" feature.</p>
Fatal flaws: Does the paper have a "fatal flaw" making it unfit for publication, regardless of other criteria (may include out of scope, double publication, plagiarism, wrong proofs, flawed experiments)? Use the text box to justify your answer.	No (not as far as I can see)
Technical quality: whether experimental methods are appropriate, proofs are sound, results	3-Poster level (top 30% submissions)

are well analyzed.	
Novelty/originality: in any aspect of the work, theory, algorithm, applications, experimental.	3-Poster level (some notable novel contributions)
Potential impact or usefulness: could be societal, academic, or practical and should be lasting in time, affecting a large number of people and/or bridge the gap between multiple disciplines.	3-Poster level (looks promising)
Clarity and presentation: explanations, language and grammar, figures, graphs, tables, proper references.	4-Oral level (excellent in every respect)
Qualitative assessment: Provide constructive feedback to the authors; justify and complement your ratings above. This is our MOST IMPORTANT QUESTION, we need to get good arguments for our decisions!	<p>Update: I have read author feedback. I leave my assessment unchanged.</p> <p># Experiments The speaker diarization task is standard for HDP-HMMs and related models. It would obviously be nice to have an experiment on a real dataset. It is good to have the error bars on the plots in fig 1 and 2. It is an acceptable restriction and a useful twist to constrain the weight matrix to the ground truth line 229 in order to make latent dimensions identifiable. The experiment described sec 4.3 is a good sanity check (data generated from HDP-HMM to see whether the present model will cancel its LT feature).</p> <p># Further concerns * No mention is made of runtimes, or at least computational complexity; this would be interesting. * A limitation of the HDP-HMM-LT model is that similarity is symmetric (Φ is symmetric), ie I cannot model that transition from j to j' is likely, but not the reverse transition from j' to j. This could arise in an NLP setting if using, for instance, an asymmetric Levenshtein distance (made asymmetric because different weights are used for different modifications). Maybe mention this? * It would be nice to have the code available, esp because it is rather complex.</p>

	<p>* It would be useful to have the derivations from the Markov jump process with failed transitions, and the Gibbs sampling steps, laid out in supplementary material.</p> <p>* The final remark line 266 is interesting.</p> <p># Presentation suggestions</p> <p>Overall, I would recommend to watch out for confused semantics in some sentences. What may sound like pedantry for the reader familiar with the topic, turns out to improve chances of in-depth understanding for the reader coming from another field, by removing stumbling blocks and extra sources of perplexity.</p> <p>* line 9 are no longer -> is</p> <p>* HDP-HMM presentation lines 25 sqq gets a bit ahead of itself: the HDP-HMM is not just a "prior on the transition matrix": suggest reword</p> <p>* line 31: I parse the sentence as: *Each of (a countably infinite set of states), indexed by j, has parameters* -> reword like *Each state, indexed j, stems from a countably infinite set, and has parameters θ_j ...*</p> <p>* eq 3 and others: you shortcut the categorical distribution, and leave it to the reader to work out that z_t is implicitly variable here -- maybe fix to $\sim \text{Cat}(\pi_{z_{t-1}})$?</p> <p>* sec 2.1 The normalized gamma process representation of the DP is not often used in the machine learning literature; it would be worth pointing this out explicitly, before taking the reader by the hand and saying something to the effect of *The normalized gamma process construction in eq 6 is equivalent to the DP. It stems from [Ferguson I think], who also proves the equivalence, and is discussed further in [8, sec 2.1]*</p> <p>* line 79: I strongly suggest sticking to one notation for 'from' and 'to' states (eg j resp j'), and making this clear</p> <p>* line 81: you (implicitly) identify the state with its θ_j here; maybe I want to use features of j which are not contained in θ_j for the similarity computation? I suggest refraining from this. Also, it seems to contradict your remarks line 186 and 264.</p> <p>* line 87: Ω shows up for the first time without introduction here, but cf remark for line 81.</p> <p>* line 88: can we not unify the notations Φ (line 81, function, but uppercase seems to hint at matrix) and ϕ ?</p> <p>* line 101: I like that you take a step back and describe where we stand.</p> <p>* line 114: isn't that τ_t ?</p> <p>* line 117: suggest just complete Poisson, enhances readability for a cost of 3 letters.</p> <p>* eq 10: is $L(\dots) = p(z, u, Q \pi, \phi)$? Otherwise I'm confused by the $=$ sign in the first line. Maybe point this out if so?</p> <p>* line 208: introduce the notation for conversations/ clusters c (saying clusters helps I think)</p>
Reviewer confidence regarding this review.	2-Confident (read it all; understood it all reasonably well)

Masked Reviewer ID: Assigned_Reviewer_3**Review:****Question**

Summary: Provide a brief summary of the contents of the paper.	This paper introduces a new prior distribution for infinite hidden Markov models by allowing the user to specify a similarity function over the emission parameters of the model. The prior over the transition matrix will then make transitions between states with similar parameters more likely.
Fatal flaws: Does the paper have a "fatal flaw" making it unfit for publication, regardless of other criteria (may include out of scope, double publication, plagiarism, wrong proofs, flawed experiments)? Use the text box to justify your answer.	No (not as far as I can see)
Technical quality: whether experimental methods are appropriate, proofs are sound, results are well analyzed.	3-Poster level (top 30% submissions)
Novelty/originality: in any aspect of the work, theory, algorithm, applications, experimental.	3-Poster level (some notable novel contributions)
Potential impact or usefulness: could be societal, academic, or practical and should be lasting in time, affecting a large number of people and/or bridge the gap between multiple disciplines.	2-Sub-standard for NIPS

Clarity and presentation: explanations, language and grammar, figures, graphs, tables, proper references.	3-Poster level (good enough)
Qualitative assessment: Provide constructive feedback to the authors; justify and complement your ratings above. This is our MOST IMPORTANT QUESTION, we need to get good arguments for our decisions!	<p>My major concern with this paper is that it doesn't introduce any major novel techniques/theory/proofs. The new prior distribution that was introduced could nonetheless be useful. However, in the experiments the authors generate data from the model (and a standard iHMM) and then show that their methods works. This is an absolutely fine quality check but isn't very convincing that this paper will have high impact in practice.</p> <p>In addition, the authors don't mention the release of any source code that could prove me wrong about how frequently it gets used in practice.</p> <p>A third point is that the paper feels a bit rushed: although I am quite familiar with Bayesian non-parametrics, from the description of the paper I cannot intuitively see whether the new prior distribution is well defined and if so, under what conditions.</p> <p>Minor comments</p> <p>eq 2.: i.i.d -> i.i.d. l 40: one shortcomming l 74: should μ not be defined with the $\tilde{\pi}$'s rather than the π's? l 88: shouldn't $\phi(\theta_j, \theta_j')$ not use the capital Φ as defined in line 87? l 88: Can you make explicit any conditions on the similarity function?</p> <p>section 2.2: Your construction seems intuitive bu I feel like this section is missing quite a bit of rigor. Surely the similarity function must have other constraints (e.g. it cannot be the all 0 function). How do I know that T_j is well defined and has a non-infinite value since you've now introduced a second parameter in the gamma distribution. Generally, I would like to know whether the distribution you've defined is OK?</p> <p>section 3.2: Can you describe in a bit more detail how the HMM can visit new states? You only have a finite representation of π so I assume there is some on the fly sampling happening?</p>
Reviewer confidence regarding this review.	3-Expert (read the paper in detail, know the area, quite certain of my opinion)

Masked Reviewer ID: Assigned_Reviewer_4

Review:

Question

Summary: Provide a brief summary of the contents of the paper.	<p>This paper presents the HDP-HMM-LT model, a Bayesian non-parametric model based on the HDP-HMM that introduces the concept of “similarity” among latent states. The transition probabilities are higher for states that are similar. Similarity is measured as a function of the emission parameters.</p> <p>The paper also derives an alternative representation of the HDP-HMM-LT model as a Markov jump process with “failed” jumps, in which the jumps between two states fail with a probability that increases with the dissimilarity between those states. This allows for a Gibbs-base inference scheme with conjugate updates.</p>
Fatal flaws: Does the paper have a "fatal flaw" making it unfit for publication, regardless of other criteria (may include out of scope, double publication, plagiarism, wrong proofs, flawed experiments)? Use the text box to justify your answer.	No (not as far as I can see)
Technical quality: whether experimental methods are appropriate, proofs are sound, results are well analyzed.	2-Sub-standard for NIPS
Novelty/originality: in any aspect of the work, theory, algorithm, applications, experimental.	3-Poster level (some notable novel contributions)
Potential impact or usefulness: could be societal, academic, or practical and should be lasting in time, affecting a large number of people and/or	3-Poster level (looks promising)

bridge the gap between multiple disciplines.	
Clarity and presentation: explanations, language and grammar, figures, graphs, tables, proper references.	2-Sub-standard for NIPS
Qualitative assessment: Provide constructive feedback to the authors; justify and complement your ratings above. This is our MOST IMPORTANT QUESTION, we need to get good arguments for our decisions!	<p>Strengths -----</p> <p>+The paper presents an interesting extension of the HDP-HMM, and (up to my knowledge) the concept of a Bayesian non-parametric HMM with similarities among states is novel.</p> <p>+The connection to the Markov jump process (with failed jumps) is also interesting and novel, and it allows for more efficient inference.</p> <p>Weaknesses -----</p> <p>+The major weakness of the paper is the lack of a motivating application. All the experiments presented are on synthetic data. The authors should consider (i) describe a motivating application in which the HDP-HMM-LT model is more appropriate than previous models; and (ii) including experiments on this real-world application (maybe moving some inference details to the Supplement to gain space).</p> <p>+The authors did not analyze the computational complexity of the resulting Gibbs sampler, either empirically or theoretically (with the big O notation). The authors should consider (i) providing the computational cost as a function of the parameters (length of the sequence, etc.), and (ii) comparing the time per iteration of the experiments in Section 4 for all models.</p> <p>Minor comments -----</p> <p>+In general, abstract+section 1 are clear model-wise, but they lack the motivation of why the model with local transitions is important. Consider discussing motivating applications in which this model is appropriate and previous models (HDP-HMM) would fail.</p> <p>+Eq. (10): How can you define a posterior distribution over ϕ, when it has not been given a prior?</p> <p>+Eq. (10): Why doesn't $(1-\phi_{jj'})$ appear as part of the $\exp()$ function? Does $\pi_{jj'} * u_j * \phi_{jj'}$ cancel out with some other term? Maybe you should consider detailing this in the Supplement.</p>

+Section 2.4 was not clear to me (too summarized).

+Section 3.3: What do you mean by "based directly on the emission distributions"? Does this mean that $\phi_{jj'}$ is a deterministic function of θ_j and $\theta_{j'}$? Similarly, what does it mean "based on a separate set of variables"? Consider giving concrete examples of both cases.

+Section 4.2 lacks some intuition/motivation about why you choose such a model and generative process. It also lacks details about the generative process itself (e.g., number of time steps, how you choose the ground truth transition probabilities, etc.). The same applies to Section 4.3.

+Line 229: How did you fix the weight matrix to the true values? Do you mean all methods were given the true matrix (including the true number of states)?

+Section 4.3: Why doesn't λ go to 0 in this case? It remains with a value of approximately 0.1 or 0.2, and the confidence interval doesn't include 0. Could the authors provide more intuition about why this is happening?

+Will the code be available?

Typos/Writing

+My (personal) preference is to lower-case everything that is not a proper noun (e.g., "hierarchical Dirichlet process hidden Markov model with local transitions", "normalized gamma process", etc.). In my opinion, this is more reader-friendly.

+Lines 22-23: "when the data is generated directly from the comparison models." This sentence was not clear to me when I first read it.

+Line 24: Change "Background and Related Work" to "Introduction and Background".

+Line 59: What is "context" here?

+Lines 70-71: "where nearness of j and j' as a function of θ_j and $\theta_{j'}$ ". Bring this idea also to Section 1.

+Line 76: Why do you use \mathbf{w} instead of just w for a base measure? Please be consistent (as in line 74).

+Section 2.1: I think it would be more clear if you replaced Eq. (6) with the HDP-HMM formulation directly, instead of the DP formulation, similarly to Eq. (8). In this way, the generalization in Eq. (8) would be more clear as well.

+Eq. (7) is similar to Eq. (3), you may want to write it inline to save

	<p>some space.</p> <p>+Section 2: State at some point that you parameterize the gamma distribution in terms of shape and rate.</p> <p>+Line 115: Typo (remove the comma before "is then").</p> <p>+Eq. (9): What is $n_{j \cdot}$? Did you mean n_j, which was defined in line 115?</p> <p>+Line 117: Typo ("distributed AS $\text{Pois}(\dots)$").</p> <p>+Eq. (10): Please define $n_{jj'}$ as the number of jumps from j to j'.</p> <p>+Section 3: Consider moving some of the details to a Supplement, and provide more details there, specially about Sections 3.2 and 3.3.</p> <p>+Line 177: Typo (remove comma before "is application-specific").</p> <p>+Lines 178+: Why do you use η_j instead of θ_j here?</p> <p>+Section 4.1 can be summarized here and the details moved to the Supplement.</p> <p>+The labels and legend in Fig. 1 are hard to read. Please replace them with something more reader-friendly. Also, the caption refers to (a)-(d), but no subcaptions are present in the figure.</p>
Reviewer confidence regarding this review.	2-Confident (read it all; understood it all reasonably well)

Masked Reviewer ID: Assigned_Reviewer_6

Review:

Question

Summary: Provide a brief summary of the contents of the paper.	The authors propose an extension to HDP-HMM model that capture the local transition dynamic. This is a useful attempt in generalizing the self-transitioning property of sticky-HDP-HMM. In fact, as the authors mention, the proposed model closely relates to the infinite factorial topic modeling and correlated topic modeling literature in general. Authors propose a Gibbs sampler algorithm for the resulting non-conjugate model. The algorithm is tested on synthetic data sets and compared against the vanilla HDP-HMM and binary factorial HMM.
Fatal flaws: Does the paper have a "fatal flaw" making it unfit for publication, regardless of other criteria (may	No (not as far as I can see)

include out of scope, double publication, plagiarism, wrong proofs, flawed experiments)? Use the text box to justify your answer.	
Technical quality: whether experimental methods are appropriate, proofs are sound, results are well analyzed.	1-Low or very low
Novelty/originality: in any aspect of the work, theory, algorithm, applications, experimental.	2-Sub-standard for NIPS
Potential impact or usefulness: could be societal, academic, or practical and should be lasting in time, affecting a large number of people and/or bridge the gap between multiple disciplines.	2-Sub-standard for NIPS
Clarity and presentation: explanations, language and grammar, figures, graphs, tables, proper references.	3-Poster level (good enough)
Qualitative assessment: Provide constructive feedback to the authors; justify and complement your ratings above. This is our	<p>Authors do a good job of deriving a Gibbs sampler for the non-conjugate model that I believe can have other applications. I certainly acknowledge that it is a tedious work and the paper has a novel approach to the inference problem. However, I have two main criticism about the paper which leads me to my decision.</p> <p>First is the lack of results on real-life data sets. As far as I understand, the 'cocktail party' data set is based on a real-life scenario but the data is synthetic. Finding a suitable data set for the model can be</p>

<p>MOST IMPORTANT QUESTION, we need to get good arguments for our decisions!</p>	<p>overwhelming but without a real-life application, it is hard to conclude that the proposed model is useful.</p> <p>Second is about the lack of comparison with variational inference. The authors mention that re-scaled HDP models already exist in the literature and their model differs in two aspects. Modeling the temporal dependency and deriving Gibbs sampler (as opposed to variational inference previously used). I think it is necessary to explain why Gibbs sampler is preferred over variational inference in this specific setting. If it is possible to modify existing variational inference algorithms to handle temporal dependency, then the results should include a comparison of Gibbs and variational inference. I am not convinced that Gibbs sampler is the appropriate choice here, especially given that variational inference tends to be faster.</p> <p>Overall, the paper is promising but not complete.</p>
<p>Reviewer confidence regarding this review.</p>	<p>2-Confident (read it all; understood it all reasonably well)</p>

Masked Reviewer ID: Assigned_Reviewer_7

Review:

Question

<p>Summary: Provide a brief summary of the contents of the paper.</p>	<p>This paper extends the well-known HDP-HMM model by providing an alternative prior for state transitions that encourages *local* transitions between similar states. The classic HDP-HMM using the normalized-Gamma representation of the HDP samples the (unnormalized) transition probability from state j to state k from $\text{Gamma}(\alpha \beta_k, 1)$, where α is the concentration parameter and β_k is the root-level probability of state k. Instead, this paper samples from $\text{Gamma}(\alpha \beta_k, \phi_{jk}^{-1})$, where ϕ_{jk} is a positive value representing the similarity of states j and k. This idea directly comes from the DILN topic model [8], but is not widely known and somewhat new in the context of HDP-HMMs.</p> <p>While the DILN paper pursued variational inference, this paper develops a thorough Gibbs sampling approach with an auxiliary variables inspired by a new interpretation as a Markov jump process with "failed state transitions". That is, we interpret ϕ_{jk} as the probability of successfully jumping from j to k. Sensibly, under this interpretation, the classic HDP-HMM where $\phi_{jk} = 1$ for all pairs never fails. Overall, the sampler tracks the usual sequence of state assignments at each timestep as well as counts for how often transitions from j to k succeeded or failed. The sampler uses blocked updates when possible and samples all hyperparameters.</p> <p>Experiments consider a particular emission model where each state has a binary vector parameter, which is fed into a linear-Gaussian likelihood. This is designed for a synthetic "cocktail party" speaker diarization task,</p>
---	---

	where several conversations happen in one "audio" track, each with one speaker at a time. When the data is generated in a way that uses local transitions the proposed model does much better than competitors (factorial HMM and standard HDP-HMM), while alternative data without local transitions the proposed model performs on-par with the HDP-HMM, as expected.
Fatal flaws: Does the paper have a "fatal flaw" making it unfit for publication, regardless of other criteria (may include out of scope, double publication, plagiarism, wrong proofs, flawed experiments)? Use the text box to justify your answer.	No (not as far as I can see)
Technical quality: whether experimental methods are appropriate, proofs are sound, results are well analyzed.	3-Poster level (top 30% submissions)
Novelty/originality: in any aspect of the work, theory, algorithm, applications, experimental.	3-Poster level (some notable novel contributions)
Potential impact or usefulness: could be societal, academic, or practical and should be lasting in time, affecting a large number of people and/or bridge the gap between multiple disciplines.	3-Poster level (looks promising)
Clarity and presentation: explanations,	4-Oral level (excellent in every respect)

language and grammar, figures, graphs, tables, proper references.	
<p>Qualitative assessment: Provide constructive feedback to the authors; justify and complement your ratings above. This is our MOST IMPORTANT QUESTION, we need to get good arguments for our decisions!</p>	<p>Review summary =====</p> <p>Overall the paper is well-written, the experiments well-executed, and deliberately encouraging transitions between similar/related states is a nice idea. I recommend (barely) accepting as a poster presentation. The weaknesses I see are a small lack of novelty in the main modeling idea (in some ways its a straightforward extension of the DILN model to the HDP-HMM context) and a lack of real-world experiments that will limit the excitement of practitioners. However, I think the work is technically solid and with some revision to better address practical concerns (runtime, scalability) would bear fruit during poster discussions.</p> <p>Technical feedback =====</p> <p>The experiments neatly show the proposed model performs sensibly on a variety of reasonably complicated synthetic tasks, which reasonably convinced me that the model will usually perform well when its assumptions are appropriate and not do terribly when its local transtions are not needed. I am quite disappointed that no "real-world" datasets were discussed. This is probably the biggest weakness of the paper, because it might suggest to practioners this method isn't ready for bigger/messier applications.</p> <p>I'd like to see more careful discussion of the runtime complexity of the algorithm, both in terms of big-oh scaling and practical wallclock measurements (eg how long does it take to process a sequence of length T with K states)? I'd also like to see some attention to scalability... how big a corpus could this method handle?</p> <p>For the synthetic data, I'd like to see some kind of not-a-line-plot that helps the reader appreciate the qualitative structure discovered by the method. Perhaps some visualization of the transition matrix, which should have a block-structure due to the similarity-favoring local transitions?</p> <p>For the synthetic dataset, I'm not too surprised that the LT model discovered more states than the noLT model, but I am surprised at how large the gap was: ~100 more states in LT model! Can the authors say more about this? One quick test would be to see how the noLT version of the sampler performs when initialized with the final set of many states discovered by the LT sampler... does its performance get much worse? Do these superfluous states go away? It may be plausible that somehow the added flexibility of the LT sampler helps it discover useful states more often than the noLT version, which might suffer from mixing problems.</p> <p>Presentation issues =====</p> <p>Throughout most of the paper the symbol θ represents emission parameters of each state. However, in Sec. 4, the symbol η is</p>

	<p>confusingly used instead to represent the binary vectory associated with each state, which is plugged into the linear-Gaussian likelihood. I'd recommend sticking with one specific symbol.</p> <p>I'm not sure I like the phrase "local transitions". The word local means all kinds of things to folks in machine learning, and I'd hate to see more overloading of an already stressed term. Perhaps you could use "similarity-favoring transitions"?</p> <p>You could improve the description of the cocktail toy dataset by making the exact parameters (one single sequence of length _____, 12 dimensions, 4 speakers per conversation, 4 conversations) clear from the beginning. Also, including some visualization of the dataset in the supplement would be nice.</p> <p>In Sec. 4.2 shouldn't # of states $K = 16$, and # of dimensions $D = 12$?</p> <p>PS: To the authors, I'm sorry for the late review submission. There was an unexpected death in my extended family that required my attention this past weekend.</p>
Reviewer confidence regarding this review.	3-Expert (read the paper in detail, know the area, quite certain of my opinion)