**General Comments:**

Linear dynamic models were first introduced to speech recognition applications by Digilakis et al. (1993). Unfortunately, the techniques described in those papers simply did not scale to the problems addressed here. The main contribution of this paper is to extend this work to a very challenging large-scale task – speech in noise task (Aurora 4) – that features a mismatched training condition. This paradigm is still beyond most state of the art machine learning systems.

Below we address the specific concerns of the reviewers:

**Reviewer #3:**

**The paper "Continuous speech recognition using linear dynamic models" submitted by you to International Journal of Speech Technology with the number IJST-D-13-00013 is THE SAME PAPER AS: "Continuous Speech Recognition Using Linear Dynamic Models" of the authors T. Ma, S. Srinivasan, G. Lazarou and J. Picone submitted to be published in IEEE Signal Processing Letters on November 8, 2010 with the number SPL-06538-2010. The paper was web-published in:** [**http://www.isip.piconepress.com/publications/unpublished/journals/2010/ieee\_spl/ldm/paper\_double\_v11.docx**](http://www.isip.piconepress.com/publications/unpublished/journals/2010/ieee_spl/ldm/paper_double_v11.docx) **As I don't have a conclusion about the revision process of IEEE SPL , lamentably I must reject it.**

Perhaps there was a misunderstanding due to the blind review process. This above web site that is referenced is, in fact, the web site of the authors. An early version of this paper was submitted to the IEEE Signal Processing Letters and rejected, with no chance for resubmission. Since then, the graduate student doing the work, Tao Ma, finished the work and completed his PhD. We have made substantial revisions to the paper and resubmitted to this journal.

We have not previously published this work in a journal or conference paper. The experimental results are much more complete and overall the paper does a much better job of explaining the contributions. Due to various circumstances, the co-authors have not had time to complete the necessary revisions until recently. We see no reason why it does not qualify for publication in your journal since it is not currently under consideration in any other journal and has not been previously published.

**Reviewer #2:**

**1. Two pass ASR system with different acoustic models in different pass has been proved and used widely to improve upon one pass ASR system. In the proposed framework LDM is used in the second pass rescoring. However, during second pass rescoring, I am not sure why the authors need to combine the LDM likelihood score with HMM likelihood score in the 2nd pass. Is it neccessary?**

**What if only LDM likelihood score is used for 2nd pass rescoring? Will the results be very different? If so, could the authors explain the reason?**

This is an excellent observation. This problem was studied extensively in Dr. Ma’s dissertation: Ma, T. (2010). *Linear dynamic model for continuous speech recognition*. Mississippi State University. This work was cited in the paper. Due to space limitations we could not explore all the details presented in the dissertation.

The continuous speech recognition experiments with various hybrid systems are described more fully in Chapter V. Table 5.2 directly addresses the concern of the reviewer. Mixing of these parameters does not dramatically change the results but does provide some benefit.

In our previous work with similar hybrid systems, we have observed that mixing scores is desirable. The basis for our hybrid system work dates back to this publication:

Ganapathiraju, A., Hamaker, J., & Picone, J. (2004). Applications of Support Vector Machines to Speech Recognition. *IEEE Transactions on Signal Processing*, 52(8), 2348-2355.

Our understanding of this phenomenon is that it relates to the interdependency of segmentation and scoring. We refer to this at several points in the paper. Since the LDM model is not integrated inside the training loop, relying on LDM solely creates a mismatch during the search process between the HMM scores/segmentation process and the LDM model. It is very hard to decouple these in practice. Relying solely on LDM scores produces marginal improvements in performance, because the LDM model is not allowed to reach into the search pass and produce better hypotheses. Combining the scores seems to give LDM a better chance to access and rescore hypotheses that can make a difference in the WER.

Is this optimal? Of course not. We would ultimately like to integrate this approach inside the training loop, and then we believe there would be no need to decouple the scores. These experiences provided the motivation to adopt an integration of the scores.

We also explored this issue in the context of a basic phone classification task for TIDigits. The LDM model is described in Chapter IV (p. 66). In the summary, we state:

Results for each phonetic class are presented individually in Figure 4.5. The relative differences in classification accuracy are not consistent among the phonemes. It can be seen that the classification results for fricatives and stops are high, while classification results for glides are lower (~85%). Vowels and nasals result in mediocre accuracy (89% and 93% respectively). Overall, LDMs provide a reasonably good classification performance for TIDigits.

We referenced these results in the paper, but chose not to explore them fully due to space limitations. We have modified the paper to explain this issue in more detail.

**2. The experiments demonstrate LDM will be a good complementary for traditional HMM/GMM score because LDM is used in 2nd pass recoring. To directly compare LDM likelihood score with HMM/GMM likelihood score, perhaps the authors could consider putting LDM in the 1st pass recognition. My understanding of the difficulty of such experiment is that it's not straightforward to use LDM to do segmentation. What if HMM/GMM is used do segmentation, then LDM is used to do scoring on the provided segments?**

Again, Dr. Ma’s dissertation explores this issue in more detail (and this work is referenced). We have not been able, due to time, funding and other limitations, to fully explore integrating LDM into the first pass of the recognition system. This is not easy to do from a theoretical or implementation standpoint (the software is extremely complex). The hybrid systems approach we have taken is fairly standard and meant to give simply an indication of the promise of this method. The improvements we see are typical of other published work on such N-best rescoring approaches and demonstrate there would be benefit to integrating this into the first pass. But the effort to do this is beyond our resources at this time.

**3. Several places in the manuscript mentioned that "HMMs typically assume a diagonal covariance matrix...". To be more accurate, it is HMMs/GMMs(HMMs with GMMs observation distribution) that assumes a diagonal covariance matrix for GMMs.**

Excellent point. The need for brevity influenced our decision to use the term HMM to refer to both the Markov structure and the observation distributions. We have modified the manuscript appropriately to make this distinction more clear.

Note also that both diagonal and full covariance models were explored in Dr. Ma’s dissertation work for a phone classification problem. Again, due to space limitations, we chose not to discuss these experiments in this paper.

**4. Apprently there are some format issues with all the equations in this manuscript. Please fix those.**

We apologize for this inconvenience. When we submitted the paper, we pointed this out. There are some MS Word problems encountered when we upload the paper. We have provided a pdf and Word document via email to the editor so hopefully the reviewers can access a clean copy of the paper. We have also tried to upload a different version to see if we can resolve these problems.

**Summary Comments:**

We would like to thank the reviewers for their thoughtful and insightful feedback. We have done our best to address each of their concerns. We hope our responses have adequately addressed your concerns and we can proceed with publication.

We want to emphasize that this is the first time this particular model has been shown to provide a statistically significant improvement on a challenging task. Recognition results on the Aurora 4 Corpus are well calibrated. These improvements are significant compared to the types of improvements typically published in human language technology papers.

The work constitutes the dissertation of a graduate student and represents four years of work on these problems, and is extremely comprehensive in nature. Again, this work has not been previously published and has been struggling through the review process for a long time. We selected ISTJ, a relatively new journal for us, because of its timely review process. The IEEE journal that we previously submitted to took an excessive amount of time. We hope we can bring this to a speedy conclusion.