predictive performance. These results are specific for the features used, the complexity, and potentially the model choice might differ if other features were utilized. Future work will reveal if other structures can be found in features that describe different aspects of music; structures that are relevant for describing and predicting aspects regarding emotions expressed in music.

Another consideration when using the generative models is that the entire feature time series is replaced as such by the model, since the distances between tracks are now between the models trained on each of the tracks and not directly on the features [5]. These models still have to be estimated, which takes time, but this can be done offline and provide a substantial compression of the features used.

7. CONCLUSION

In this work we presented a general approach for evaluating various track-level representations for music emotion prediction, focusing on the benefit of modeling temporal aspects of music. Specifically, we considered datasets based on robust, pairwise paradigms for which we extended a particular kernel-based model forming a common ground for comparing different track-level representations of music using the probability product kernel. A wide range of generative models for track-level representations was considered on two datasets, focusing on evaluating both using continuous and discretized observations. Modeling both the valence and arousal dimensions of expressed emotion showed a clear gain in applying temporal modeling on both the datasets included in this work. In conclusion, we have found evidence for the hypothesis that a statistically significant gain is obtained in predictive performance by representing the temporal aspect of music for emotion prediction using MFCC's.

8. REFERENCES

[1] J-J. Aucouturier and F. Pachet. Music similarity measures: What's the use? In *3rd International Conference on Music Information Retrieval (ISMIR)*, pages 157–163, 2002.

[2] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[3] R. D. Bock and J. V. Jones. *The measurement and prediction of judgment and choice*. Holden-day, 1968.

[4] F. Huszar. A GP classification approach to preference learning. In *NIPS Workshop on Choice Models and Preference Learning*, pages 1–4, 2011.

[5] V. Imbrasaite, T. Baltrusaitis, and P. Robinson. Emotion tracking in music using continuous conditional random fields and relative feature representation. In *ICME AAM Workshop*, 2013.

[6] T. Jebara and A. Howard. Probability Product Kernels. *Journal of Machine Learning Research*, 5:819–844, 2004.

[7] J. H. Jensen, D. P. W. Ellis, M. G. Christensen, and S. Holdt Jensen. Evaluation of distance measures between gaussian mixture models of mfccs. In *8th International Conference on Music Information Retrieval (ISMIR)*, 2007.

[8] J. Madsen, B. S. Jensen, and J. Larsen. Predictive modeling of expressed emotions in music using pairwise comparisons. *From Sounds to Music and Emotions, Springer Berlin Heidelberg*, pages 253–277, Jan 2013.

[9] J. Madsen, B. S. Jensen, J. Larsen, and J. B. Nielsen. Towards predicting expressed emotion in music from pairwise comparisons. In *9th Sound and Music Computing Conference (SMC) Illusions*, July 2012.

[10] A. Meng, P. Ahrendt, J. Larsen, and L. K. Hansen. Temporal feature integration for music genre classification. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(5):1654–1664, 2007.

[11] A. Meng and J. Shawe-Taylor. An investigation of feature models for music genre classification using the support vector classifier. In *International Conference on Music Information Retrieval*, pages 604–609, 2005.

[12] E. M. Schmidt and Y. E. Kim. Modeling musical emotion dynamics with conditional random fields. In *12th International Conference on Music Information Retrieval (ISMIR)*, 2011.

[13] E. M. Schmidt, J. Scott, and Y. E. Kim. Feature learning in dynamic environments: Modeling the acoustic structure of musical emotion. In *13th International Conference on Music Information Retrieval (ISMIR)*, 2012.

[14] B. Schölkopf, R. Herbrich, and A. J. Smola. A generalized representer theorem. *Computational Learning Theory*, 2111:416–426, 2001.

[15] D. Sculley. Web-scale k-means clustering. *International World Wide Web Conference*, pages 1177–1178, 2010.

[16] K. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, 2009.

[17] Y. Vaizman, R. Y. Granot, and G. Lanckriet. Modeling dynamic patterns for emotional content in music. In *12th International Conference on Music Information Retrieval (ISMIR)*, pages 747–752, 2011.

[18] Y-H. Yang and H.H. Chen. Ranking-Based Emotion Recognition for Music Organization and Retrieval. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(4):762–774, May 2011.

[19] J. Zhu and T. Hastie. Kernel logistic regression and the import vector machine. In *Journal of Computational and Graphical Statistics*, pages 1081–1088. MIT Press, 2001.

[5] We do note that using a single model across an entire musical track could potentially be over simplifying the representation, in our case only small 15-30-second excerpts were used and for entire tracks some segmentation would be appropriate.