the cue models, and can hence be expected to contain a relatively large number of false positives. On the other hand, MOD2ON shows smaller differences between precision and recall values, and shows higher F1 performances than MOD1ON in both melodic sets (although the difference between performances is significant only for the instrumental set). This last result highlights the robustness of the optimisation procedure driving MOD2ON. [19]

The large F1 differences between MOD1ON and MOD2ON in respect to COMPLETE suggest that segmentation at the phrase level is a perceptual process which, despite happening in 'real time' (i.e. as music unfolds itself, represented more closely by module 1), might still require repeated exposure and retrospective listening (represented more closely by module 2).

Manual examination COMPLETE reveals that, when segmenting the vocal melody set, the prediction stage of module 1 tends to overestimate the importance of cue models (i.e. it often misclassifies models as relevant when they are not). However, when altering the settings of COMPLETE so that the prediction stage of model 1 is more conservative (i.e. so that it predicts fewer boundaries), there is no significant improvement in performance. Closer analysis of these results points to a trade-off in performance, i.e. while a conservative setting increases precision (predictions have fewer 'false positives'), it also decreases recall (predictions have fewer 'correct positives'). This suggests that the prediction stage of module 1 might require estimation of cue relevance at a local level, i.e. on subsections of the melody rather than on the whole melody.

6. CONCLUSION

In this paper we introduce a multi-strategy system for the segmentation of symbolically encoded melodies. Our system combines the contribution of single strategy models of melody segmentation. The system works in two stages. First, it estimates how relevant the boundaries computed by each selected single strategy model are to the melody being analysed, and then combines boundary predictions using heuristics. Second, it assesses the segmentation produced by combinations of the selected boundary candidates in respect to corpus-learnt priors on segment contour and segment length.

We tested our system on 100 vocal and 100 instrumental folk song melodies. The performance of our system showed a considerable (10% $F1$) improvement upon the state-of-the-art in melody segmentation for instrumental folk music, and showed to perform second best in the case of vocal folk songs.

In future work we will test if the relevance of cue models can be accurately estimated for sections of the melody (and not the whole melody as it is done in this paper). This 'local' account of relevance might play a major role in improving the system's precision. Also, we will incorporate a more advanced model of prior segment knowledge of segment structure in our system. We hypothesise that a model of the characteristics of [2] could constitute a good alternative to model not only segment length and contour, but also to incorporate knowledge of 'template' phrase structure forms. Lastly, we will continue testing our model's generalisation capacity by evaluating on larger sample sizes and genres other than folk (for the latter the authors are currently in the process of annotating a corpus of Jazz melodies).

Acknowledgments: We thank Frans Wiering, Remco Veltkamp, and the anonymous reviewers for the useful comments on earlier drafts of this document. Marcelo Rodríguez-López and Anja Volk (NWO-VIDI grant 276-35-001) and Dimitrios Bountouridis (NWO-CATCH project 640.005.004) are supported by the Netherlands Organization for Scientific Research.

7. REFERENCES

[1] S. Ahlbäck. Melodic similarity as a determinant of melody structure. *Musicae Scientiae*, 11(1):235–280, 2007.

[2] R. Bod. Probabilistic grammars for music. In *Belgian-Dutch Conference on Artificial Intelligence (BNAIC)*, 2001.

[3] M. Bruderer, M. McKinney, and A. Kohlrausch. The perception of structural boundaries in melody lines of western popular music. *Musicae Scientiae*, 13(2):273–313, 2009.

[4] E. Cambouropoulos. The local boundary detection model (LBDM) and its application in the study of expressive timing. In *Proceedings of the International Computer Music Conference (ICMC01)*, pages 232–235, 2001.

[5] E. Cambouropoulos. Musical parallelism and melodic segmentation. *Music Perception*, 23(3):249–268, 2006.

[6] E. Clarke and C. Krumhansl. Perceiving musical time. *Music Perception*, pages 213–251, 1990.

[7] M. Hamanaka, K. Hirata, and S. Tojo. ATTA: Automatic time-span tree analyzer based on extended GTTM. In *ISMIR Proceedings*, pages 358–365, 2005.

[8] M. Pearce, D. Müllensiefen, and G. Wiggins. The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception*, 39(10):1365, 2010.

[9] M. Rodríguez-López and A. Volk. Melodic segmentation using the jensen-shannon divergence. In *International Conference on Machine Learning and Applications (ICMLA12)*, volume 2, pages 351–356, 2012.

[10] G. Sargent, F. Bimbot, E. Vincent, et al. A regularity-constrained Viterbi algorithm and its application to the structural segmentation of songs. In *ISMIR Proceedings*, 2011.

[11] D. Temperley. *The cognition of basic musical structures*. MIT Press, 2004.

[19] If we consider that (with MOD1ON bypassed) the number of candidate boundaries taken as input to MOD2ON often exceeds 'correct' (human annotated) boundaries by a factor 2 or 3, then the number of possible segmentations of the melody shows an exponential increase, leading to local minima issues, and so it would be reasonable to expect a performance equal or worse than that of MOD1ON.