











has been labelled as 'movement' and the other as 'no movement', but both may have large velocity/acceleration. Our 'Neighbours' condition is one attempt to deal with that, and we think that the fact that the improvement over the baseline is larger in this condition indicates that it is useful to look at possibilities beyond the word. We intend to pursue other strategies for this in the future.

Another, less problematic, case is that more than one head movement can co-occur with the same word. Our feature extraction deals with that as it is not dependant upon the number of peaks within a word, just the max, the average etc.

## 5 Conclusion

We have developed a system for the detection of word-related head movements in audiovisual recordings of read news. The task seems feasible; our data seems to have predictive power. The results show no effects from using individual vs groups of labellers. Furthermore, they show that it is possible to generalize over several different news readers. Labelling at word boundaries causes some issues when head movements occur across boundaries.

## Acknowledgements

This work was supported by an infrastructure grant from the Swedish Research Council (Swe-Clarín, 2014–2018; grant number 821-2013-2003), a grant from the Marcus and Amalia Wallenberg Foundation (grant number 2012.0103) and also partially funded by the Bank of Sweden Tercentenary Foundation (grant number P12-0634:1). We also thank our two additional annotators Anneliese Kelterer and Otto Ewald.

## Reference

- Ambrazaitis, G., Svensson Lundmark, M. & House, D. (2015). Multimodal levels of prominence : a preliminary analysis of head and eyebrow movements in Swedish news broadcasts. In Svensson Lundmark, M., Ambrazaitis, G. & van de Weijer, J. (Eds.) Working Papers in General Linguistics and Phonetics (Proceedings from Fonetik 2015) (pp. 11-16), 55. Centre for Languages and Literature, Lund University.
- Boersma, P., Weenink, D. 2014. Praat: doing phonetics by computer [Computer program]. <http://www.praat.org/>
- Bruce, G. 1977. Swedish Word Accents in Sentence Perspective. *Travaux de l'institut de linguistique de Lund* 12. Malmö: Gleerup.
- Bruce, G., B. Granström (1993). Prosodic modelling in Swedish speech synthesis. *Speech Communication* 13, 63–73.
- Chen, T. & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In 22nd SIGKDD Conference on Knowledge Discovery and Data Mining.
- Fleiss, J. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378-382.
- Nystrom, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, 42, 188-204. doi:10.3758/BRM.42.1.188
- Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36, 1627-1639.
- Viola, P., & Jones, M. J. (2001) Rapid Object Detection using a Boosted Cascade of Simple Features, Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. Volume: 1, pp.511–518.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. 2006. ELAN: a professional framework for multimodality research. Proc. of LREC 2006, Fifth International Conference on Language Resources and Evaluation. See also: <http://tla.mpi.nl/tools/tla-tools/elan/>
- Zhang, S., Wu, Z., Meng, H., Cai, L. (2007) Head Movement Synthesis based on Semantic and Prosodic Features for a Chinese Expressive Avatar In: ICASSP 2007, Vol. 4, pp.837-840, 2007.4