Word networks for BCI decoding purposes

T. Pfeiffer[1*], R. T. Knight[2], G. Rose[1]

[1]*Institute for Medical Engineering, Univ. of Magdeburg;* [2]*Helen Wills Neuroscience Institute, UC Berkeley, USA*

*Universitaetsplatz 2, 39106 Magdeburg, Germany. E-mail: tim.pfeiffer@ovgu.de

Introduction: In a previous work [1] we demonstrated that the beneficial properties of hidden Markov models (HMM) can be used to extract additional information without further classifier training efforts. For data from a complex visual paradigm, HMMs could be used to decode the type of the stimulus (picture category) and additionally predict its duration with high accuracy when trained on category information only. There is no analogous approach for the widely used static classifiers (e.g. SVMs). However, due to the small amount of training data available and low signal-to-noise ratios, parameter estimation for HMMs is difficult and therefore, the achievable decoding accuracies are slightly lower than for static classifiers. Here, we show how the use of word networks (Fig. 1a) - a technique from the field of speech decoding (where HMMs represent the gold-standard) - can be used to conveniently incorporate prior knowledge into the decoding. This information boosts the decoding accuracies reached with an HMM approach up to the level of SVMs. Considering the additional information that can be extracted due to the dynamic nature of HMMs, the routine turns out superior overall.

Material and Method: The analysis is performed on two electrocorticography (ECoG) datasets as described in [1]. High gamma features are computed using Matlab implementations as described in [1,2]. The paradigm design (Fig. 1b) causes some trials to contain multiple stimuli, thus providing a challenging dataset for HMMs. We compare the accuracies for a standard HMM decoder as used in [1] ('Standard'; this represents a single-word recognizer) and three setups using WNs (semi-continuous decoding) with increasing levels of prior knowledge:

(1) 'WN default': No prior knowledge. Here, the longest consecutive interval determines the decoded category.

(2) 'WN first word': For trials modeled by multiple HMMs, we assign the label of the first model.
    This setup reflects that all trials start with the original stimulus (i.e. on which the trial label is based).

(3) 'WN FW + no short': Use the first model's label, but ignore sequences shorter than 20 samples (≙ 375 ms).
    This adds the information that no such stimuli (i.e. shorter than 300 ms) appeared in the paradigm.
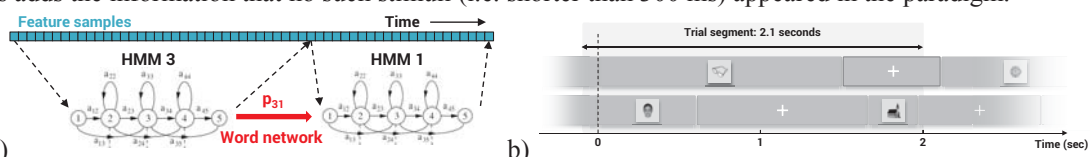

Figure 1.   *a) Principle of HMM decoding using word networks [3]. Contrary to the 'Standard' case, in which each trial must be modeled by a single HMM, the time series of feature values of a trial can be modeled as a combination of multiple HMMs using WNs. b) Visual paradigm: pictures from three categories (objects, faces, watches) were presented for 5 different durations (300ms-1500 ms). Note, that some trials (esp. with short first stimulus) contain an additional stimulus at the end of the time segment.*
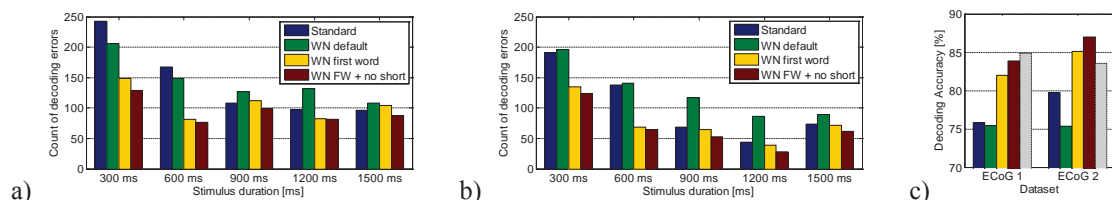

Figure 2.   *Count of category decoding errors using different decoding strategies for both datasets (a) ECoG 1, b) ECoG 2). c) Decoding accuracies for both datasets and all decoding strategies (chance level: 33.3%; dashed gray bar: SVM results).*

Results and Discussion: In the 'Standard' case, most decoding errors occur for shorter stimuli (Fig. 2a,b). This is likely caused by additional stimuli appearing within these trial segments. A single HMM cannot compensate for this and misclassifies the trial eventually. WNs allow combinations of HMMs within a single trial but increase the complexity of the decoding routine. For the 'WN default' case this results in decreasing accuracy, especially for longer stimulus durations. However, with increasing level of prior knowledge, decoding accuracies rise significantly. With the combination of both prior knowledge aspects, the 'FW + no shorts' case consistently leads to the best results. The gain in decoding accuracy is 8.1 % and 7.2 % (ECoG 1 and 2) with respect to the 'Standard' case. Thus, the routine can catch up (ECoG 1) or even outperform (ECoG 2) respective SVM results while preserving the benefits of dynamic classifiers (see introduction).

Significance: We show that the use of WNs facilitates the incorporation of prior knowledge. This can significantly improve decoding accuracies and offers huge potential for more sophisticated analysis.

References

[1] Pfeiffer T, Heinze N, Frysch R, Deouell L Y, Schoenfeld M A, Knight R T, Rose G. Extracting duration information in a picture category decoding task using Hidden Markov Models. In *Journal of Neural Engineering*, 13(2): 026010, 2016.
[2] Wissel T, Pfeiffer T, Frysch R, Knight R T, Chang E F, Hinrichs H, Rieger J W, Rose G. Hidden markov model and support vector machine based decoding of finger movements using electrocorticography. In *Journal of Neural Engineering*, 10(5):056020, 2013.
[3] HTK version 3.4.1, Cambridge University Engineering Department, available online at: http://htk.eng.cam.ac.uk