

# gazeHMM: Parameter recovery simulation

Malte Lüken, Simon Kucharsky, & Ingmar Visser

5 5 2020

In this document, we are preregistering a simulation study to estimate the parameter recovery of a hidden Markov model (HMM) which is part of our recently developed algorithm gazeHMM. The purpose of this algorithm is to classify eye movements into distinct events (e.g., fixations). The full algorithm will be available on Github upon publication.

In general, gazeHMM consists of three steps: First, the raw gaze data is preprocessed. Missing samples are identified, the data are filtered, and for each sample the velocity, acceleration, and difference in angle to the next sample is calculated. Second, the HMM estimates the parameters of response, transition, and initial state models for the specified eye movement events. Moreover, the HMM computes the posterior probability of each sample for belonging to the events. Each sample is labeled as the event with the highest posterior probability. Third, a postprocessing routine relabels samples that are violating theoretical and practical constraints. Since the HMM might not model the data perfectly, this step improves classification performance. This simulation study will only consider the parameter recovery of the HMM but not the performance of the preprocessing and postprocessing routines of the algorithm.

## Model description

The HMM estimates unobservable states that have generated the data. In the context of eye-tracking, each state corresponds to an eye movement event. In gazeHMM, the model can have between two and four states. State one always corresponds to fixations, state two always to saccades, state three to postsaccadic oscillations (PSOs), and state four to smooth pursuits. Moreover, the HMM is multivariate but conditionally independent on the states. Thus, it has three independent response models for every state describing velocity, acceleration, and angle. The velocity and acceleration response models for all four states use two-parameter gamma distributions (shape and scale parametrization). The angle response model for the fixation state uses a uniform distribution, while the other three states use a two-parameter von-Mises distribution (mean and concentration parametrization) to model angle. Both the transition model and the initial state model use a multinomial distribution (with the logit and the identity link function, respectively). Note that no covariates are included in the model, only intercepts for the parameters are estimated.

Depending on the number of states, the HMM has a different number of parameters. The response models of the HMM contain six parameters per state. The transition models have  $k^k$  parameters in total with  $k$  being the number of states. The initial state models have one parameter per state. In sum, HMMs with two, three, and four states have X, X, and X parameters. The denotation of each parameter is shown in Table 1.

### *Optimization*

Table 1

Parameter	Description
$\pi$	Initial probabilities for state $i$
$a_{ij}$	Transition probability from state $i$ to $j$
$\$$	

## Model parameter recovery

The parameter recovery study for the HMM is divided into two parts: In the first part, we will vary the parameters of the model and estimate the recovery. In the second part, the parameters of the model will be fixed. Instead, we will manipulate the size of the sample generated by the model level, the noise added to the generated data, the starting values for the HMM, and the amount of missing data.

### Variable parameters

First, we will repeatedly generate data with the HMM and the parameters displayed in Table 2. Second, the HMM estimates the parameters from the generated data. Third, we will compare the estimated parameters with the true parameters. We will treat the parameter recovery as satisfactory when the 95% confidence bands include the true parameter values.

Table 2

State	Parameter	Min.	Max.	Fixed
1-4	$\pi$	-	-	$1/k$
1-4	$a_{i=j}$	0.01	0.99	0.9
1-4	$a_{i \neq j}$	0.01	0.99	0.1
1	$\alpha_{vel}$	1	5	3
1	$\beta_{vel}$	0.5	1	0.75
1	$\alpha_{acc}$	1	5	3
1	$\beta_{acc}$	0.05	0.5	0.15
2	$\alpha_{vel}$	1	5	3
2	$\beta_{vel}$	5	15	10
2	$\alpha_{acc}$	1	5	3
2	$\beta_{acc}$	1	5	3
2	$\mu$	-	-	0
2	$\kappa$	0.1	10	1
3	$\alpha_{vel}$	1	5	3
3	$\beta_{vel}$	1	5	3
3	$\alpha_{acc}$	1	3	2
3	$\beta_{acc}$	1	3	2
3	$\mu$	-	-	$\pi$
3	$\kappa$	0.1	10	1
4	$\alpha_{vel}$	1	5	3
4	$\beta_{vel}$	1	2	1.5
4	$\alpha_{acc}$	1	5	3
4	$\beta_{acc}$	0.05	0.25	0.15
4	$\mu$	-	-	0
4	$\kappa$	0.1	10	1

### Scenario 2: Sample size and noise

In this scenario, the parameters of the HMM are fixed to the values shown in Table 2. For HMMs with two, three, and four states, we will repeatedly generate data with the model for 500, 2500, and 10.000 samples. In addition, gaussian white noise with parameters  $\sigma_{vel;acc} \in \{1, 5, 25\}$  and  $\kappa_{angle} \in \{0.1, 1, 10\}$  will be added to the data. This will results in 27 ( $5 \times 3 \times 3$ ) conditions with  $D$  repetitions each.

