Modelling spatial modulation in the visual cortex: An HMM-based approach

Franziska Kaltenberger

Master of Science
Artificial Intelligence
School of Informatics
University of Edinburgh
2023

Abstract

When mice traverse a virtual corridor, neurons in the visual cortex respond differently to the same visual stimulus depending on the location. So far, little is known about the exact mechanism of how spatial information influences visual processing in these navigation tasks. While predictive coding offers a potential explanation, there are currently no computational models to test different mathematical frameworks for generating this spatial modulation of visual activity. This thesis introduces a hierarchical HHM-based model that simulates neural responses in spatio-visual navigation tasks and exhibits stimulus selectivity and spatial modulation in visual activation. Our model can be employed to fit experimental data and explore the impact of different uncertainty-related model parameters on selectivity and modulation. We argue that our approach serves as a baseline for assessing computational theories like predictive coding for their capacity to explain spatial modulation in the visual cortex.

Research Ethics Approval

This project was planned in accordance with the Informatics Research Ethics policy. It did not involve any aspects that required approval from the Informatics Research Ethics committee.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Franziska Kaltenberger*)

Acknowledgements

I want to thank my supervisor, Angus Chadwick. I have greatly appreciated your invaluable support and guidance throughout my dissertation project as well as our insightful discussions. A big thank you also goes to all my lovely and super supportive people here in Edinburgh who have been with me on my journey over this last year. Finally, I would like to thank my amazing friends and family back home. Thank you so much for sending all your love and encouragement up north. I am so glad to have you all in my life.

Table of Contents

Chapter 1

Introduction

On a perceptual level, our brain constantly needs to deal with uncertainty, e.g. when inferring the identity of a three-dimensional object from the two-dimensional retina image. First proposed by Von Helmholtz (1867), behavioural studies have verified that humans act according to approximate (Bayesian) probabilistic inference in various tasks including concept learning (Ellis, 2023; Tenenbaum, 1999), decision making (Stankevicius et al., 2014), and multisensory cue integration (Alais and Burr, 2019; Shams et al., 2000). These studies used computational modelling to fit and predict human performance in the tasks, providing a computational framework that can account for the observed behaviour. Modeling cognition as a probabilistic inference mechanism has also been successfully used to examine psychiatric conditions like schizophrenia (Fletcher and Frith, 2009; Schmack et al., 2015) or autism (Pellicano and Burr, 2012; Powell et al., 2016). Supported by this body of evidence, probabilistic theories seem to provide one of the most plausible universal theories for cognition.

Despite humans' ability to perform close to optimal inference on a cognitive level, the question remains whether the brain also behaves probabilistically on a neural level and how this probabilistic mechanism is implemented (Aitchison and Lengyel, 2017). Is inference only performed by higher cognitive areas or already a tool used in early sensory processing? Are inference processes separated between sensory regions or does the inference process span different brain areas with each of them contributing information? How is sensory evidence information combined with prior expectations about this information? Additionally, how can binary-spiking neurons represent probabilities and compute under uncertainty? Here, too, computational models can explore diverse mathematical frameworks by simulating neural activity to examine how the brain implements inference.

The predictive coding theory (Friston, 2005; Rao and Ballard, 1999), for example, answers these questions by postulating that information between different layers in the hierarchical structure of the cortex is transmitted via predictions and represented by prediction errors. Top-down signals coming from higher cortical levels entail expectations about the bottom-up signals coming from the lower levels in the hierarchical structure. Neural activity in a layer then encodes the prediction error, i.e. the mismatch between these two signals. Although predictive coding was first proposed as a theory to explain receptive field effects in the visual cortex (Rao and Ballard, 1999), it has later been extended to more abstract levels of perception and cognition (Clark, 2013; Friston, 2010) and can thus be considered a prominent candidate for a unifying theory of cognition.

Fiser et al. (2016) used the predictive coding theory to explain modulations in neural activity of the visual cortex in a spatial navigation task. Although animals observed the exact same visual stimulus in different positions throughout their traversal of a corridor, the measured activity of the visual cortex differed significantly between presentations along the corridor. Additionally, visual responses became more informative about the animal's position with experience. This modulation of V1 activity was confirmed by subsequent experiments (Saleem et al., 2018; Diamanti et al., 2021). So far, however, only a few studies attempted to provide a computational modeling analysis of how spatial and visual information interact in navigational tasks (Recanatesi et al., 2021; Ujfalussy and Orbán, 2022).

More importantly, to our knowledge, there exists no published study that provides a biologically inspired modeling framework for simulating spatial navigation experiments in which spatial modulations of visual activity have been observed (Fiser et al., 2016; Saleem et al., 2018). This thesis fills this gap by proposing a computational simulation framework that allows to examine the interaction between spatial and visual information in one-dimensional visual navigation tasks. Starting with a basic Hidden Markov Model (HMM) for inferring the current location from stimuli observations, we introduce an hierarchical HMM version that extends the basic model by latent visual representations. We encode posterior probabilities using three different encoding frameworks: (1) Distributional Distributed Codes, (2) sampling-based encoding, and (3) mean encoding. These encoding methods vary with respect to how they encode the posterior uncertainty.

The proposed simulation approach does not explicitly model selectivity or spatial modulation in the visual neurons. Nevertheless, our model exhibits both stimulus selectivity and spatial modulation in visual neurons. In an initial parameter analysis,

we tested the effect of uncertainty in the model and neural encoding. Both, selectivity and modulation, increased with decreasing model uncertainty and were highest for the mean encoding framework that does not encode a measure of uncertainty. Assuming that learning through experience is reflected in a uncertainty reduction in the system, our model yields results that are in line with empirical evidence (Fiser et al., 2016). This verifies our model as a first modelling approach to examine the computational mechanisms behind the observed spatial modulations in the visual cortex. In future research, it could be used as a baseline for testing the theory that modulations reflect spatially informed top-down predictions about upcoming visual inputs according to predictive coding that was put forward by Fiser et al. (2016).

The following Chapter 2 explains the simulated experiments and their results in more detail and provides an introduction on modelling approaches of how uncertainty might be represented in the brain. In Chapter 3, we present the models that were designed throughout the thesis and explain how we perform inference on them. We first describe the concept of our basic model design, before explaining why and how this was extended into a hierarchical model. Chapter 4 focuses on the encoding schemes that we consider to obtain neural representations of the inference results. Thereafter, we describe the simulation environment and parameter settings that we use to examine our models in Chapter 5. Finally, Chapter 6 examines how selectivity and modulation of visual activity are influenced by uncertainty in our model before pointing out the potential of our modelling approach for explaining neural recordings in visual navigation tasks and how it can be used in future research to test more complex theories such as predictive coding.

Chapter 2

Background

This chapter establishes the context of this thesis in two steps. First, we describe the neuro-scientific foundations of spatio-visual navigation and explain the findings in the literature that reported spatial modulation of V1 activity. Second, we present the computational concepts that we use to model these experiments focussing on the internal model describtion and relevant concepts of neural coding.

2.1 Neural modulation during spatio-visual navigation

Spatio-visual navigation tasks describe experimental setups in which an agent – usually a mouse – has to traverse through an environment using visual cues. These tasks require an information exchange between the brains spatial and visual system. Spatial information is encoded via place cells that form spatial maps in the hippocampus (O'Keefe and Dostrovsky, 1971; Shapiro et al., 1997). The hippocampal area CA1 is especially active in navigational tasks (Hok et al., 2007; Lenck-Santini et al., 2001, 2002; O'Keefe and Speakman, 1987), however representing rather the subjectively estimated than the actual position (Lenck-Santini et al., 2002; O'Keefe and Speakman, 1987; Rosenzweig et al., 2003; Saleem et al., 2018). Visual cues, for example in the form of landmarks, were found to play a crucial role in navigational tasks (Chen et al., 2013; Geiller et al., 2017; Muller and Kubie, 1987; Wiener et al., 1995), with positional representations already evolving in early stages of visual processing (Haggerty and Ji, 2015; Fiser et al., 2016; Saleem et al., 2018). How and to what extend this information is exchanged, still remains to be established.

In order to examine the degree to which early areas of visual processing already entail spatial information, Saleem et al. (2018) recorded from neurons from CA1 and V1

in 4 mice while animals were traversing a 100 cm long corridor in a virtual environment. The corridor comprised two repeated segments of length 40 cm that hold two distinct landmarks (one grating and one plaid). Although the landmarks were identical in the two segments and the animals thus received the same visual input in both positions, Saleem et al. found that the recorded activity for most neurons in V1 differed depending on the position of the mouse in the corridor (see Supplementary Figure C.1a). In particular, neural responses were stronger for the preferred landmark than for the identical landmark shifted by 40 cm along the corridor. Landmark preferences for a single neuron were consistent over trials and neurons would prefer the first or second landmark presentation with equal proportion indicating that activity modulations are not yielded by adaptation.

For their analysis, the authors considered approximately 5000 neurons with sufficiently strong activity along the corridor. They quantified positional activity changes using a spatial modulation ratio MR:

$$MR = \frac{\text{activity in non-preferred position}}{\text{activity in preferred position}} \tag{2.1}$$

The median MR was significantly smaller than 1 (0.61 ± 0.31, $p < 10^{-104}$, Wilcoxon two-sided signed rank test, Saleem et al. (2018, p. 125)). This modulation in V1 activity could not be replicated using a purely visual model that simulated complex cells in the visual cortex, but had no spatial information (see Supplementary Figure C.1b). Thus, the change in response observed in V1 could not be explained by purely visual components. Further analysis using a ridge-regularised General Linear Model with different predictive features to fit recorded neural activity confirmed the necessity of spatial information for the emergence of the modulation in V1.

These findings are supported by experiments performed by Fiser et al. (2016) in which mice moved in a virtual tunnel at any speed (forward or backward motion allowed). The authors did not specify the length of the tunnel, but refered to previous experiments that used a tunnel length of 180 cm (Harvey et al., 2009). The tunnel was divided into equally sized patches. Partitioned by unique landmarks, the patches displayed two sinusoidal gratings, A and B, with equal frequency and contrast, but orthogonal in orientation. The grating presentation alternated along the corridor: grating A was presented in the first an third patch, whereas the second and fourth patch displayed grating B. The grating in the fifth section changed according to one of five conditions, but was set to grating A during learning.

Fiser et al. recorded neural activity in 9 mice over multiple days from approximately

1600 neurons in V1 and 1700 neurons in CA1. Animals received an reward at the end of the tunnel before being reset in position to the beginning of the corridor. The authors analysed the recorded neurons for their preference of grating A or B using a selectivity index *SI* defined as

$$SI = \frac{\overline{r}_A - \overline{r}_B}{\overline{r}_A + \overline{r}_B} \tag{2.2}$$

where $\overline{r}$ denotes the average response to grating A or B. Neurons in V1 were categorised into A- or B-selective if their SI was significantly different from 0. Selectivity in these neurons improved over time and became more stable. Although the visual input in the first/second and third/fourth was identical, the neural response in these sections differed in amplitude in some neurons of V1 depending on the position in the tunnel. Thus, the recorded V1 activity also entailed spatial information, results that match the findings reported by Saleem et al. (2018). The same modulation was found for CA1 neurons where the selectivity pattern might be due to place cells activity that is known to integrate sensory cues for position encoding (Chen et al., 2013; Harvey et al., 2009; Ravassard et al., 2013).

Both, Fiser et al. (2016) and Saleem et al. (2018), excluded experimental factors such as running speed or reward as an explanation for the observed modulation and, thus, verified that the spatial modulation in the hippocampus and the visual cortex were related. Using a decoding analysis (Fiser et al. (2016): Matlab TreeBagger; Saleem et al. (2018): Independent Bayes Decoder), both studies showed that the activity in V1 and CA1 was informative of the location as decoders were able to predict the position given the neural activity with sufficient accuracy. Furthermore, Saleem et al. found that activities in both areas were more informative about the animals subjective position estimate than its actual position. Fiser et al. also showed that experience influences the modulation reflected in the accuracy of the decoder performance. In particular, spatial information in V1 stimulus onset activity increased over time while the mean CA1 activity became less informative of the animal's position. Notably, classification accuracy using CA1 activity onset only was slightly above chance and did not change significantly with experience. The selectivity increase in V1 thus cannot be explained by a selectivity increase in CA1.

As spatial information is present in both spatial and visual activity, these areas may not be processing sensory information independently, but how they interact exactly remains unclear. Saleem et al. (2018) suggested a common influence signal (either feed-forward or feedback) on the processing in CA1 and V1 based on an analysis of simultaneous processing errors in both areas but the authors were indefinite about

the exact nature of this connection. Although the CA1 hippocampal region and the visual system, especially the primary visual cortex (V1), correlate in activity (Ji and Wilson, 2007; Haggerty and Ji, 2015; Niell and Stryker, 2010), they are only connected indirectly through other brain areas such as the anterior cingulate cortex (ACC) (Fiser et al., 2016). The ACC is not only responsible for spatial memory (Maviel et al., 2004; Teixeira et al., 2006), but it is considered to indicate conflicts in sensory information, even if they occur in early stages of processing (Carter et al., 1998). Fiser et al. (2016) examined the signals from the ACC into V1. They postulated the theory that spatial information from the ACC could activate visual representations that would result in prediction-like V1 activity. Indeed, Fiser et al. could distinguish postsynaptic V1 neurons into purely visual and predictive based on the resulting activity onset during grating presentation. By varying the last grating in the corridor in different experimental conditions, they detected a correlation between these two types of neurons in V1 and proposed that visual neurons encode a mismatch signal between the predicted and observed visual input, supporting the predictive coding hypothesis (Friston, 2005, 2010; Rao and Ballard, 1999).

2.2 Modelling uncertainty in the brain

The predictive coding hypothesis is an example for a functional theory of how the brain might implement probabilistic inference. As presented in Chapter 1, evidence from behavioural studies supports the theory that the brain is operating in an inference-like scheme, maybe even performing close to optimal Bayesian inference. The corresponding computations would require a representation of uncertainty in the form of probability distributions. This uncertainty could be reflected in two different, not exclusive ways: (1) in an abstracted computation mechanism following a probabilistic model, or (2) in the neural code representing information about the environment. The following sections will briefly explore these two ways to represent uncertainty, focusing on introducing terminology and concepts relevant to this thesis.

2.2.1 Reasoning under uncertainty

Representations of uncertainty require probabilistic models. In contrast to statistical models that only provide parametric descriptions of relations between variables in observed data, probabilistic models define probability distributions to express or es-

timate the likelihood of different parameters or outcomes. Probabilistic models are especially useful when dealing with situations where there is inherent uncertainty as they can explicitly quantify the uncertainty associated with predictions or inferences. However, this separation is not entirely distinct as statistical models can be considered probabilistic if they define probability distributions over their parameters and variables.

One example for such a model is the Hidden Markov Model (HMM). A HMM defines the relation between visible states that can be observed and measured and their possible hidden causes represented by connected latent states that cannot be observed directly. Each state emits an observation with a certain probability and the transitions between states follow the Markov property, meaning that the probability of transitioning to a particular state depends only on the current state and not on any previous information. Probabilistic inference on these models utilises their Markovian nature to quantify the probability of the underlying latent states and explain the hidden mechanism behind observed quantities.

Probabilistic modeling with HMMs finds application in various fields, including speech recognition (Juang and Rabiner, 1991; Gales et al., 2008), face recognition (Ali et al., 2022; Liu and Cheng, 2003), and more (see Mor et al. (2021) for a systematic review on HHMMs). In neuroscience, HMMs define an abstract internal model of the brain. This model is then used to extract information about possible hidden states and underlying patterns within neural data, such as electroencephalography (EEG) data or local field potentials, which can be complex and challenging to interpret directly. The abstraction level of the hidden states can thereby vary between neural states of different brain areas and different cognitive or behavioural processes. Concrete applications of HMMs in neuroscience include decoding of brain states (Chen et al., 2016; Quinn et al., 2018), spike train analysis (Katahira et al., 2010; Radons et al., 1994), and studying neural diseases or disorders such as epilepsy (Dash et al., 2020) or post-traumatic stress disorder (PTSD) (Ou et al., 2015).

2.2.2 Neural encoding and decoding

The specification of internal models like a HMM, however, is only relevant for certain frameworks of neural coding. Lange et al. (2020) point out the differences between these *encoding* approaches for representing neural activity and similar *decoding* frameworks.

In general, neural *encoding* models focus on internal conceptual structures – represented by an internal model – and how they might yield a neural activity response.

The internal model often comprises unobserved latent variables as well as observable variables – similar to our generative models defined in Chapter 3. A posterior distribution over the latents given a set of observations can be computed or approximated by performing inference on this model using the likelihood of the observations given a certain model status. Notably, the encoding posterior does not take into account the neural representation of the encoded quantities, but encoding models define a mapping from these inference outcomes to the neural representations. The characteristics of the posterior that are used to obtain neural representations vary between frameworks. While distributional codes provide an expectation-based encoding of the whole posterior distribution (Vértes and Sahani, 2018; Zemel et al., 1998), other frameworks only encode samples (Hoyer and Hyvärinen, 2002; Orbán et al., 2016) or predictions and error signals (Friston, 2005).

As one of the main contrasts to encoding, *decoding* frameworks don't postulate any assumptions about an internal model, but focus on the relation between neural activity and the connected an external stimulus – not the internal (noisy) observation of the stimulus. This relation can be characterised using only a purely statistical model from which the posterior over the external stimulus given the neural activity can be computed. While most encoding perspectives separate uncertainty in neural computation from the uncertainty representation in population activity, decoding frameworks like probabilistic population codes (Ma et al., 2006) treat response variability as an explicit way to represent uncertainty about the world, i.e. the sensory stimulus. The precision of the decoding posterior therefore increases if more neural responses are considered for computing the posterior. Notably, the decoding likelihood describes how likely a particular neural response is given a certain external stimulus. How this stimulus is perceived or represented is not considered in these frameworks, which makes the likelihood indirectly dependent on the experimental design. (Lange et al., 2020)

Following Lange et al., both, encoding and decoding frameworks, can be distinguished into those models that encode samples from the posterior, and others that provide a parametric encoding of the posterior distribution. These categories then can be further divided depending on if and how they use the encoded quantities to perform inference. For some frameworks, however, these hard divisions cannot be made (Sahani and Dayan, 2003; Zemel et al., 1998) and some encoding frameworks could be used to perform decoding and vice versa. However, restating the equations and calculating the required quantities is often non-trivial (Vértes, 2020).

Chapter 3

Models

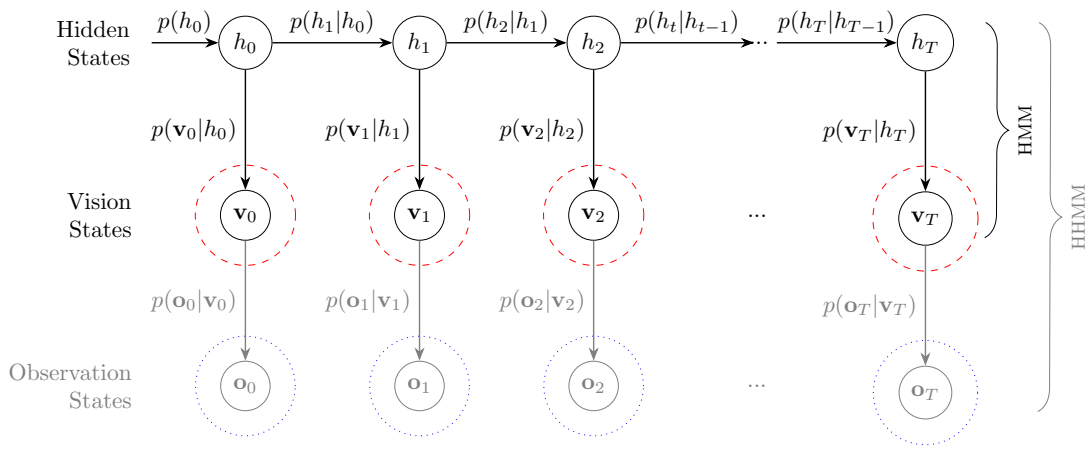

Figure 3.1: **Graphical models for simulating spatio-visual navigation tasks.** The generative model of the basic model is given by a Hidden Markov Model (HMM) with positions modeled by the hidden states h_t and visual observations modeled by the vision states v_t (black lines). Vision stated are directly observed (indicated by red dashed circles). The Hierarchical Hidden Markov Model (HHMM) extends the basic model with explicit observation states (gray lines). In this model, vision states become latent variables and observation states are observed instead (indicated by blue doted lines).

This chapter focuses on the model that we built to simulate experiments discussed previously (see Chapter 2). We start off by explaining a basic Hidden Markov Model (HMM) for performing inference in a spatio-visual task environment. This model, however, lacks an explicit representation of the activity in the visual cortex and was thus extended to a Hierarchical Hidden Markov Model (HHMM).

The subsequent sections focus mainly on the overall structure and parameters of these models, as well as design choices and biological motivations. For detailed

derivations of the equations used for inference and neural encoding please refer to the Appendix A for the HMM and the Appendix B for the HHMM.

3.1 Basic Model

Fiser et al. (2016) as well as Saleem et al. (2018) recorded neural activity from the hippocampus and the primary visual cortex. The hippocampus encodes positions by means of place cells (Harvey et al., 2009; O'Keefe and Dostrovsky, 1971) and could propagate this spatial information to the visual cortex over indirect top-down connections (Fiser et al., 2016; Wang et al., 2011). The primary visual cortex (V1) is the first cortical area in the visual processing path and is often considered to already give rise to simple representations like orientation perceptions (Ben-Yishai et al., 1995; Ferster and Miller, 2000). The presented studies focus on the recordings from these two regions for their selectivity and modulation analysis. As we aim to provide a simulation framework to model these experiments and analyses, our basic model has two main variables representing spatial and visual information, resepectively.

The causal relation of these two variables is given by a Hidden Markov Model (HMM) where the latent variables represent the position and the observable variables represent the visual stimuli. The following section defines this generative model and its parameters, before we explain how we perform inference on this model to obtain probability distributions over the spatial and visual parameters.

3.1.1 Generative model

Although previous experiments used non-colored stimuli like gratings with different orientations (Fiser et al., 2016; Saleem et al., 2018), our basic model represents the visual domain as a three-dimensional vector space. This vector space can, for example, be assumed to represent a color space like the RGB color space (colors defined by outputs of primary lights with long, medium, and short wavelength, e.g. red, green, and blue) or the HSB color space (colors defined by three primary properties, e.g. hue, saturation, and brightness). According to Trichromatic theory, all perceivable colours can be created by mixing three linearly independent primaries and, thus, lie in a three-dimensional linear vector space (Krantz, 1975; Neitz and Jacobs, 1986). However, this numerical definition of the visual domain could, for example, also be used to represent the orientation, phase and contrast of a sinusoidal grating.

In our simulation environment, three-dimensional visual inputs $\mathbf{v}$ are presented at every location h along a one-dimensional corridor. Visual stimuli follow a designed pattern that provides a non-linear mapping $f(h) : h \rightarrow \mathbf{v}$. As the same visual stimulus can be presented in different positions of the corridor, this many-to-one mapping introduces ambiguity into the system. Ambiguity was found to modulate visual processing and perception, on a neural level (Sun et al., 2017) as well as on a perceptual level, e.g. in bistable illusions (see Brascamp and Shevell (2021) for a review). The mapping also provides the mean values for the distribution over the observable variable at a certain location in the corridor. Using Gaussian observation noise $\eta_{\mathbf{v}} \sim \mathcal{N}(0, \sigma_{\mathbf{v}})$, the distribution for a visual $\mathbf{v}_t$ at time step t is given by a multi-variate normal distribution with mean $f(h_t)$ for the location h_t at time step t and variance $\Sigma_{\mathbf{v}} = \sigma_{\mathbf{v}}\mathbb{I}$.

The spatial information is represented by the hidden variables h of the HMM. Due to the defined non-linear mapping $f(h)$, there exists no closed form solution for inferring the latent variables h given the observations. Thus, our model discretises space to enable exact inference. The spatial h_t at time step t then represents the index of the current position in the one-dimensional corridor.

The model simulates the movement of an agent in the corridor using a Poisson Process. The initial position of the agent is Poisson-distributed with parameter $\lambda_{init} \in \mathbb{R}_+$. An agent moves down the corridor by executing an intended action $\lambda_{trans} \in \mathbb{R}_+$ with Poisson-distributed innovation noise, i.e. it can only move forward in the corridor. Thus, the step size in a motion trajectory is given by a Poisson distribution with parameter λ_{trans}. The intended action is currently set to be constant. However, the model could be extended to make the choice of the action probabilistic by sampling it from a normal distribution, for example.

In summary, the basic generative model is given by a Hidden Markov model (HMM) that is defined as follows:

- *Initial probability distribution*

$$p(h_0) \sim Poisson(\lambda_{init}) \tag{3.1}$$

- *Transition probability distribution*

$$p(h_{t+1}|h_t) \sim Poisson(\lambda_{trans}) \tag{3.2}$$

  for a constant intended action λ_{trans}.

- *Emission probability distribution*

$$p(\mathbf{v}_t|h_t) \sim \mathcal{N}(f(h_t), \Sigma_{\mathbf{v}}) \tag{3.3}$$

  where the mean of the normal distribution is given by the predefined mapping between the location and observation at time step t. This mapping is given by the setup of the corridor.

This definition yields the following joint marginal for the HMM:

$$p(\mathbf{v}_0, ..., \mathbf{v}_T, h_0, ..., h_T) = p(h_0) \prod_{t=0}^{T-1} p(h_{t+1}|h_t) \prod_{t=0}^{T} p(\mathbf{v}_t|h_t) \tag{3.4}$$

A graphical representation of this HMM is presented with black arrows in Figure 3.1.

3.1.2 Inference

Our model is designed to provide a probabilistic simulation tool to anaylse the neural modulation that was found in previous experiments (Fiser et al., 2016; Saleem et al., 2018). Given the model setup defined in the previous section, such modulation in the position representation would be reflected in the posterior probabilities over the spatial information, i.e. the hidden variables of the HMM. More precisely, given observations from previous time steps we care to find the probability distribution over the current position: $p(h_t|\mathbf{v}_{1:t})$. For a HMM, this posterior can be found using the forward algorithm also known as filtering.

As the visual information is directly observed, modulation analysis can't be performed for the current time step. Thus, we are also interested in the predictive posterior for both spatial and visual information to get a reflection of uncertainty in the visual domain. In addition, Fiser et al. (2016) attributed the activity modulation to predictive neurons – a theory for which these predictive posteriors provide a intital simple measure for modulation analysis. These predictive posteriors can be computed using the results obtained from the filtering algorithm.

The subsequent sections explain the forward algorithm and how we use it to compute our objective probabilities. Our model is assumed to have perfect knowledge of the generative process. Therefore, all parameters and probability distributions as well as the setup of the corridor (i.e. the mapping between the locations and observations) are known during inference.

3.1.2.1 Filtering - Inferring the current location

Filtering is performed by message passing using the alpha-recursion algorithm. For this, the HMM is considered as a factor graph. Given observations $\{\mathbf{v}_s\}_{s=0}^{t}$, this factor graph reduces to a chain for all time steps $s < t$. The factors of the graph are given by $\phi_0(h_0) = p(\mathbf{v}_0|h_1)\,p(h_0)$ and $\phi_s(h_{s-1}, h_s) = p(\mathbf{v}_s|h_s)\,p(h_s|h_{s-1})$. Messages that are passed from h_s to h_{t+1} are called $\alpha(h_s)$ by convention.

The the alpha-recursion algorithm is then defined as follows[1]:

1. **Init:** $\alpha(h_0) = \phi_0(h_0) = p(\mathbf{v}_0|h_0)p(h_0)$
2. **Update:** For $0 < s \leq t$:

$$\alpha(h_s) = \sum_{h_{s-1}} \phi_s(h_{s-1}, h_s)\alpha(h_{s-1}) = p(v_s|h_s) \sum_{h_{s-1}} p(h_s|h_{s-1})\alpha(h_{s-1})$$

In this notation, the marginals are defined by $\alpha(s_t) = p(h_t, \mathbf{v}_{1:s})$ and $\sum_{h_t} \alpha(h_s) = p(\mathbf{v}_{1:s}) =: Z_s$. Thus, the desired probability distribution over the current location in time step t given the observations up to this time step is given by

$$p(h_t|\mathbf{v}_{1:t}) = \frac{1}{Z_t}\,\alpha(h_t) \tag{3.5}$$

3.1.2.2 Predicting future location and observation

Using the result from the filtering algorithm, the probability of the future location h_{t+1} given the observations up to the current time step t is given by

$$p(h_{t+1}|\mathbf{v}_{1:t}) = \sum_{h_t} p(h_{t+1}|h_t)\,p(h_t|\mathbf{v}_{1:t}) \tag{3.6}$$

Using $p(\mathbf{v}_{t+1}|h_{t+1}) = \mathcal{N}(\mathbf{v}_{t+1}; f(h_{t+1}), \Sigma_{\mathbf{v}})$, the visual predictive posterior is defined as a mixture of Gaussians:

$$p(\mathbf{v}_{t+1}|\mathbf{v}_{1:t}) = \sum_{h_{t+1}} \mathcal{N}(\mathbf{v}_{t+1}; f(h_{t+1}), \Sigma_{\mathbf{v}})\,p(h_{t+1}|\mathbf{v}_{1:t}) \tag{3.7}$$

3.2 Hierarchical model

The basic model already provides a fundamental framework to analyse the selectivity and modulation of spatial neurons. However, we particularly aim to simulate neural

[1] For detailed explanation, see Barber (2012), Section 23.2 Hidden Markov Models (pp. 473-476).

modulation found in V1 activity (Fiser et al., 2016; Saleem et al., 2018). But as the vision states are directly observed, the HMM lacks an explicit representation of visual processing areas such as V1 and thus does not allow for an direct modulation analysis in the visual domain. We therefore extend our basic HMM by an additional emission layer of observation states which are now directly observed. The vision states, on the other hand, become latent states and can now be inferred during the filtering process.

On a highly abstracted level, this additional observation layer can be assumed to represent the activity on the retina, whereas the visual layer represents primary cortical areas of visual processing (such as V1) that also receive information from cortical areas of spatial processing (such as the hippocampus). Although we omit intermediate layers of visual processing like the thalamus, this model already provides a simple hierarchical structure of spatio-visual processing. We thus call it a Hierarchical Hidden Markov Model (HHMM).

3.2.1 Generative model

The hierarchical generative model is given by a Hidden Markov model (HMM) with two emission layers. The generative process largely follows the process defined in Section 3.1.1. In addition to the visual layer $\mathbf{v}_t$, we introduce an explicit observation variable $\mathbf{o}_t = g(\mathbf{v}_t) + \eta_\mathbf{o}$ with $\eta_\mathbf{o} \sim \mathcal{N}(0, \sigma_\mathbf{o}^2 \mathbb{I})$. For now, we assume $g(\mathbf{v}_t) = \mathbf{A}\,\mathbf{v}_t$ with $\mathbf{A}$ assumed to be linear and invertible ($\exists \mathbf{A}^{-1} : \mathbf{A}\,\mathbf{A}^{-1} = \mathbb{I}$). Notably, this implies that the observations and visuals both lie in a three-dimensional space.

Thus, the generative process gets expanded by the following emission function:

- *Observation emission probability distribution*

$$p(\mathbf{o}_t|\mathbf{v}_t) \sim \mathcal{N}\big(\mathbf{A}\mathbf{v}_t, \sigma_\mathbf{o}^2 \mathbb{I}\big) \tag{3.8}$$

This extends the joint marginal of the HMM (see Equation (3.4)) to the following joint model for the HHMM:

$$p(\mathbf{o}_0, ..., \mathbf{o}_T, \mathbf{v}_0, ..., \mathbf{v}_T, h_0, ..., h_T) = p(h_0) \prod_{t=0}^{T-1} p(h_{t+1}|h_t) \prod_{t=0}^{T} p(\mathbf{v}_t|h_t) p(\mathbf{o}_t|h_t) \tag{3.9}$$

In Figure 3.1, this model extension is denoted in the graphical representation of the HHMM by gray arrows.

3.2.2 Inference

According to the inference in the HMM, the inference process in the HHMM has two main objectives: (1) perform filtering to obtain posterior distributions over the latent variables, and (2) perform prediction to get predictive posterior distributions. In contrast to the HMM, however, the hierarchical structure of the HHMM allows to infer posterior distributions over both spatial and visual information, i.e. $p(h_t|\mathbf{o}_{1:t})$ and $p(\mathbf{v}_t|\mathbf{o}_{1:t})$.

These distributions can be computed by adapting the alpha-recursion algorithm presented above (see Section 3.1.2.1). The subsequent section presents the HHMM version of the filtering algorithm that was derived in the context of this thesis, before the equations for all target probability distributions are presented. Detailed derivations for the following equations are provided in Appendix B.

3.2.2.1 Filtering in the HHMM

For calculating the posterior of the latents in time step t given all observations up to the current time step, we first need to compute the respective marginals. These are given by the following alpha-omega-recursion algorithm:

1. **Init:**

$$\begin{aligned}\alpha(h_1,\mathbf{v}_1) &= p(h_1,\mathbf{v}_1,\mathbf{o}_1)\\ &= a\,\mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_1; f(h_1),\Sigma'_{\mathbf{o}_t}\big)\,\mathcal{N}\big(\mathbf{v}_1;\mu'_{\mathbf{v}_1}(h_1),\Sigma'_{\mathbf{v}_1}\big)\,p(h_1)\end{aligned} \tag{3.10}$$

2. **Update:** For $0 < s \leq t$:

$$\begin{aligned}\alpha(h_s,\mathbf{v}_s) &= p(h_s,\mathbf{v}_s,\mathbf{o}_{1:s})\\ &= a\,\mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_s; f(h_s),\Sigma'_{\mathbf{o}_s}\big)\,\mathcal{N}\big(\mathbf{v}_s;\mu'_{\mathbf{v}_s}(h_s),\Sigma'_{\mathbf{v}_s}\big)\,\omega(h_s)\end{aligned} \tag{3.11}$$

   using

$$\omega(h_s) = a\sum_{h_{s-1}} p(h_s|h_{s-1})\,p(h_{s-1},\mathbf{o}_{1:s-2})\,\mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_{s-1}; f(h_{s-1}),\Sigma'_{\mathbf{o}_s}\big) \tag{3.12}$$

with variances $\Sigma'_{\mathbf{o}_s} := \sigma^2_{\mathbf{o}}\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top + \Sigma_{\mathbf{v}}$ and $\Sigma'_{\mathbf{v}_s} = \Big(\big(\sigma^2_{\mathbf{o}}\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top\big)^{-1} + \big(\Sigma_{\mathbf{v}}\big)^{-1}\Big)^{-1}$ and mean $\mu'_{\mathbf{v}_s}(h_s) := \Sigma'_{\mathbf{v}_s}\Big(\big(\sigma^2_{\mathbf{o}}\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top\big)^{-1}\mathbf{A}^{-1}\mathbf{o}_s + \big(\sigma^2_{\mathbf{v}}\mathbb{I}\big)^{-1} f(h_s)\Big)$, as well as constant $a := |\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top|^{\frac{1}{2}}$ (see Appendices B.3 and B.4).

3.2.2.2 Posterior distributions in the HHMM

The results obtained through the filtering algorithm can then be used to define the posterior distributions over the latent variables. For the current time step t, the joint posterior over spatial and visual latent variables, as well as the respective spatial and visual posterior distributions are defined as follows:

$$p(h_t, \mathbf{v}_t|\mathbf{o}_{1:t}) = \frac{1}{\mathcal{Z}} \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big) \mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t, \mathbf{o}_t), \Sigma'_{\mathbf{v}_t}\big)\, \omega(h_t) \tag{3.13}$$

$$p(h_t|\mathbf{o}_{1:t}) = \frac{1}{\mathcal{Z}} \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big)\, \omega(h_t) \tag{3.14}$$

$$p(\mathbf{v}_t|\mathbf{o}_{1:t}) = \frac{1}{\mathcal{Z}} \sum_{h_t} \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big) \mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t), \Sigma'_{\mathbf{v}_t}\big)\, \omega(h_t) \tag{3.15}$$

Accordingly, the predictive posteriors for the spatial and visual latent variables are given by

$$p(h_{t+1}|\mathbf{o}_{1:t}) = \frac{1}{\mathcal{Z}} \sum_{h_t} \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big)\, p(h_{t+1}|h_t)\, \omega(h_t) \tag{3.16}$$

$$p(\mathbf{v}_{t+1}|\mathbf{o}_{1:t}) = \frac{1}{\mathcal{Z}} \sum_{h_t, h_{t+1}} \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \mathbb{I}\big) \mathcal{N}\big(\mathbf{v}_{t+1}; f(h_{t+1}), \sigma^2_{\mathbf{v}}\mathbb{I}\big)\, p(h_{t+1}|h_t)\, \omega(h_t) \tag{3.17}$$

Derivations of these equations as well as for the normalisation constant $\mathcal{Z}$ can be found in Appendices B.6 and B.5, respectively.

Chapter 4

Neural Encoding

To compare the probabilities resulting from filtering and prediction to neural recordings from the literature reported by Fiser et al. (2016) and Saleem et al. (2018), the inferred probability distributions are encoded into firing rates. We compare firing rates obtained using three different encoding methods: (1) *mean encoding* (Ujfalussy and Orbán, 2022), (2) *distributed distributional code*s (Vértes and Sahani, 2018; Zemel et al., 1998), or (3) *sampling-based encoding* (Hoyer and Hyvärinen, 2002; Orbán et al., 2016). These encoding frameworks differ with respect to the quantity they encode, as well as their relation to representing uncertainty, as we discuss in subsequent sections. However, all three methods can be used to encode the inferred posterior probability distributions obtained in time step t into firing rates for N neurons.

The following section defines how we set up the neural populations and basis functions used in all encoding mehtods. Thereafter, we introduce all three encoding methods on a conceptual level and explain how these methods encode the HMM and HHMM posterior distributions obtained during inference.

4.1 Neural basis functions

The visual and spatial information are encoded by different populations of neurons. Our neural population setup is inspired by the place-cell strucutre in the hippocampus and follows Ujfalussy and Orbán (2022), adapted from Rich et al. (2014). For both populations, the receptive fields of the neurons are defined by respective basis functions that are given by a mixture of Gaussians. These basis functions include the tuning curves ψ_{ik} of K different subfields for every neuron i.

In the spatial domain, these tuning curves are given by univariate Gaussians:

$$\psi_{ik}(x) = \exp\Big(-\frac{(x-\mu_{ik})^\top (x-\mu_{ik})}{2\sigma_{ik}^2}\Big) \tag{4.1}$$

Since observations are multidimensional, the subfield tuning curves in the visual domain are given by multivariate Gaussians:

$$\psi_{ik}(\mathbf{x}) = \exp\Big(-\frac{1}{2}(\mathbf{x}-\mu_{ik})^\top \Sigma^{-1}(\mathbf{x}-\mu_{ik})\Big) \tag{4.2}$$

The basis function ϕ_i for neuron i sums up the activity given by the subfield tuning curves ψ_{ik}, weighted by activation weights ρ_{ik}. This weighted activity is then added to the neurons base rate ρ_{i0}. Following this definition, spatial basis functions are given by

$$\phi_i(x) = \rho_{i0} + \sum_{k=1}^{K_i} \rho_{ik}\psi_{ik}(x) \tag{4.3}$$

Accordingly, the basis functions in the visual domain are defined as follows:

$$\phi_i(\mathbf{x}) = \rho_{i0} + \sum_{k=1}^{K_i} \rho_{ik}\psi_{ik}(\mathbf{x}) \tag{4.4}$$

All parameters are defined and sampled following Ujfalussy and Orbán (2022) (see Chapter 5).

4.2 Distributed distributional codes (DDC)

Distributional codes – or convolutional codes (Pouget et al., 2003) – are an example for parametric neural encoding. The underlying concept was first introduced by Zemel et al. (1998) as *distributional population codes* (DPC). DPCs encode the expected response rate $\mathbb{E}[r_i]$ for neuron i under a certain probability distribution $p(x)$ over a variable X as

$$\mathbb{E}[r_i] = \mathbb{E}_{p(x)}[\phi_i(x)] \tag{4.5}$$

(corresponding to Equation (4.7) in Vértes (2020)). Variability in neural responses is then induced via a Poisson firing model that takes the expected response rate as parameter. Notably, if the variable X is observed, the probability $p(x)$ collapses to a Dirac-delta function and the encoding of the observed value x_o is given by the value of the basis functions at location x_o:

$$\mathbb{E}_{\delta(x-x_o)}[\phi_i(x)] = \phi_i(x_o) \tag{4.6}$$

(corresponding to Equation (4.8) in Vértes (2020)).

More recent encoding models have built on the distributional concept of DPCs. Sahani and Dayan (2003), for example, expanded the framework into *doubly distributional population codes* (DDPC) to encode multiple stimuli that are presented simultaneously. Vértes and Sahani (2018) omit the Poission firing model specified by the DPC framework, but keep the rate-based encoding model to define *distributed distributional codes* (DDC). Using these DDC encodings Vértes and Sahani propose a expectation-based computational framework that allows to perform inference and learning in generative models with a hierarchical latent variable structure - similar to the HHMM presented in Section 3.2.

Using the DDC encoding framework (Vértes and Sahani, 2018; Zemel et al., 1998), we define the firing rates encoding a probability distribution $\gamma(x)$ for a random variable X are determined as follows:

$$r_i(\gamma) = \int_X \phi_i(x)\,\gamma(x)\,dx \tag{4.7}$$

This yields the following definitions of the firing rate encoding of the inferred current and predicted location:

$$r_i(p(h_t|\mathbf{x}_{1:t})) = \sum_{h_t} \phi_i(h_t)\,p(h_t|\mathbf{x}_{1:t}) \tag{4.8}$$

$$r_i(p(h_{t+1}|\mathbf{x}_{1:t})) = \sum_{h_{t+1}} \phi_i(h_{t+1})\,p(h_{t+1}|\mathbf{x}_{1:t}) \tag{4.9}$$

with $\mathbf{x} = \mathbf{v}$ for the HMM and $\mathbf{x} = \mathbf{o}$ for the HHMM.

As the visual posteriors are given as mixture of Gaussians, calculating the DDC encoded firing rate in the visual domain is not trivial, although closed-form solutions can be derived. The DDC encoding of the current visual posterior inferred in the HHMM is given by

$$r_i(p(\mathbf{v}_t|\mathbf{o}_{1:t})) = \int_{-\infty}^{\infty} \phi_i(\mathbf{v}_t) p(\mathbf{v}_t|\mathbf{o}_{1:t})\,d\mathbf{v}_t \tag{4.10a}$$

$$\overset{B.7}{=} \frac{1}{Z}\sum_{h_t}\Bigg[\mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big)\,\omega(h_t) \left(\rho_{i0} + \sum_{k=1}^{K_i} \rho_{ik}\sqrt{(2\pi)^D|\Gamma_{ik}|}\,\mathcal{N}\Big(\mu_{ik}; \mu'_{\mathbf{v}_t}(h_t), \Gamma_{ik} + \Sigma'_{\mathbf{v}}\Big)\right)\Bigg] \tag{4.10b}$$

The predictive visual posterior for both generative models is encoded in the DDC framework as follows:

$$r_i(p(\mathbf{v}_{t+1}|\mathbf{x}_{1:t})) = \int_{-\infty}^{\infty} \phi_i(\mathbf{v}_{t+1})\,p(\mathbf{v}_{t+1}|\mathbf{x}_{1:t})\,d\mathbf{v}_{t+1} \tag{4.11a}$$

$$= \sum_{h_{t+1}} \left[p(h_{t+1}|\mathbf{x}_{1:t}) \left(\rho_{i0} + \sum_{k=1}^{K_i} \rho_{ik} \sqrt{(2\pi)^D |\Gamma_{ik}|}\, \mathcal{N}\left(\mu_{ik}; f(h_{t+1}), \Gamma_{ik} + \sigma_{\mathbf{v}}^2 \mathbb{I}\right) \right) \right] \tag{4.11b}$$

with $\mathbf{x} = \mathbf{v}$ for the HMM and $\mathbf{x} = \mathbf{o}$ for the HHMM.

The detailed derivation of equations (4.10) and (4.11) are presented in Appendices B.7 and A.1, respectively. All DDC equations for the HHMM are listed in Appendix B.8.

4.3 Mean encoding

In contrast to DDCs, the mean encoding framework (Ujfalussy and Orbán, 2022) only encodes a single value instead of whole a distribution, i.e. the distribution mean:

$$r_i(\gamma) = \phi_i(\bar{x}_\gamma) \tag{4.12}$$

This parametric encoding model encodes a point estimate of the posterior distributions and disregards all measures of model uncertainty. Uncertainty in the neural representation is then only induced by the basis functions used for population encoding. It can thus be used as a baseline model to analyse whether the observed neural modulation in spatio-visual navigation tasks is influenced by uncertainty representation through neural encoding.

For both models presented in Chapter 3, the encoded spatial mean is given by the weighted average of indices according to the respective spatial posterior distribution. As the visual posteriors are mixtures of Gaussians, the respective means are defined as the weighted averages of the means of the normal distributions inside the sum. For the predictive visual posterior in the HMM, the mean is given by

$$\bar{\mathbf{v}}_{t+1} = \sum_{h_{t+1}} f(h_{t+1})\, p(h_{t+1}|\mathbf{v}_{1:t}). \tag{4.13}$$

Accordingly, the encoded means in the HHMM are defined as follows:

$$\bar{\mathbf{v}}_t = \frac{1}{Z} \sum_{h_t} \mu'_{\mathbf{v}_t}(h_t)\, \mathcal{N}\left(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\right) \omega(h_t) \tag{4.14}$$

$$\bar{\mathbf{v}}_{t+1} = \frac{1}{Z} \sum_{h_t, h_{t+1}} f(h_{t+1}) \mathcal{N}\left(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \mathbb{I}\right) p(h_{t+1}|h_t)\, \omega(h_t) \tag{4.15}$$

4.4 Sampling encoding

In contrast to the previous parametric encoding frameworks, sampling-based methods (Berkes et al., 2011; Hoyer and Hyvärinen, 2002; Orbán et al., 2016) establish a direct connection between neural variability and model-induced uncertainty by only encoding samples drawn from the encoded probability distributions. In the sampling framework, the encoding of any probability distribution $\gamma(x)$ is realised by sampling from the distribution: $\hat{x}_\gamma \sim \gamma$. This sample is then used to determine the firing rates of neuron i:

$$r_i(\gamma) = \phi_i(\hat{x}_\gamma) \tag{4.16}$$

For the spatial probability distributions, $\hat{h}$ is obtained by sampling from the corridor indices with the provided posterior probability distributions. As the visual predictive posterior is given by a mixture of Gaussians, it cannot be sampled from directly. Instead, a normal distribution with mean $f(h_{t+1})$ is chosen with probability $p(h_{t+1}|\mathbf{v}_{1:t})$ or $p(h_{t+1}|\mathbf{o}_{1:t})$ for the HMM and HHMM, respectively. $\hat{\mathbf{v}}_{t+1}$ is then sampled from this chosen normal distribution.

Chapter 5

Experimental setup

The main goal of this project was to provide a computational method to simulate and analyse the spatial modulation in the virtual cortex reported in the literature (Fiser et al., 2016; Saleem et al., 2013). We therefore tested our models in an experimental setup that followed the virtual environment used in these studies. Experiments simulated an agent moving forward through an one-dimensional corridor and observing different visual stimuli along the way. Figure 5.1a shows the schematic setup of this corridor. It is divided into equally sized patches in which a certain stimulus is displayed. As we discretise space in our model, these patches are divided into P positional bins with $P = 20$ for all presented experiments. The length of the corridor with X separate stimuli patches is consequently given as $L = P \cdot X$. For every positional bin, the mapping $f(h_t)$ defined in Section 3.1.1 determines what visual stimulus is observed in this location.

Following the virtual environment structure of Fiser et al. (2016) (see Supplementary Figure C.2a), presentations of testing stimuli A and B in the corridor are intercepted by neutral stimuli N and landmark stimuli L1-L4. According to the model description (see Chapter 3), stimuli are represented in a three-dimensional space and could, for example, define colours. For the HMM, stimuli values range within $[0,255]^3$ with the neutral stimulus N marking the center of this space. Stimuli A and B were selected manually such that they have equal distance to the center and therefore are equally likely to be covered by random neural basis functions with uniformly distributed mean and variance. Landmark stimuli were sampled uniformly and are fixed over all presented experiments and trials. Table 5.1 gives the values for all HMM stimuli in its left column.

As the HHMM extends the basic model, the stimulus values of the HMM become the values of the visual representations in the HHMM. This maintains the properties in the neural encoding described for the HMM and thereby ensures comparability

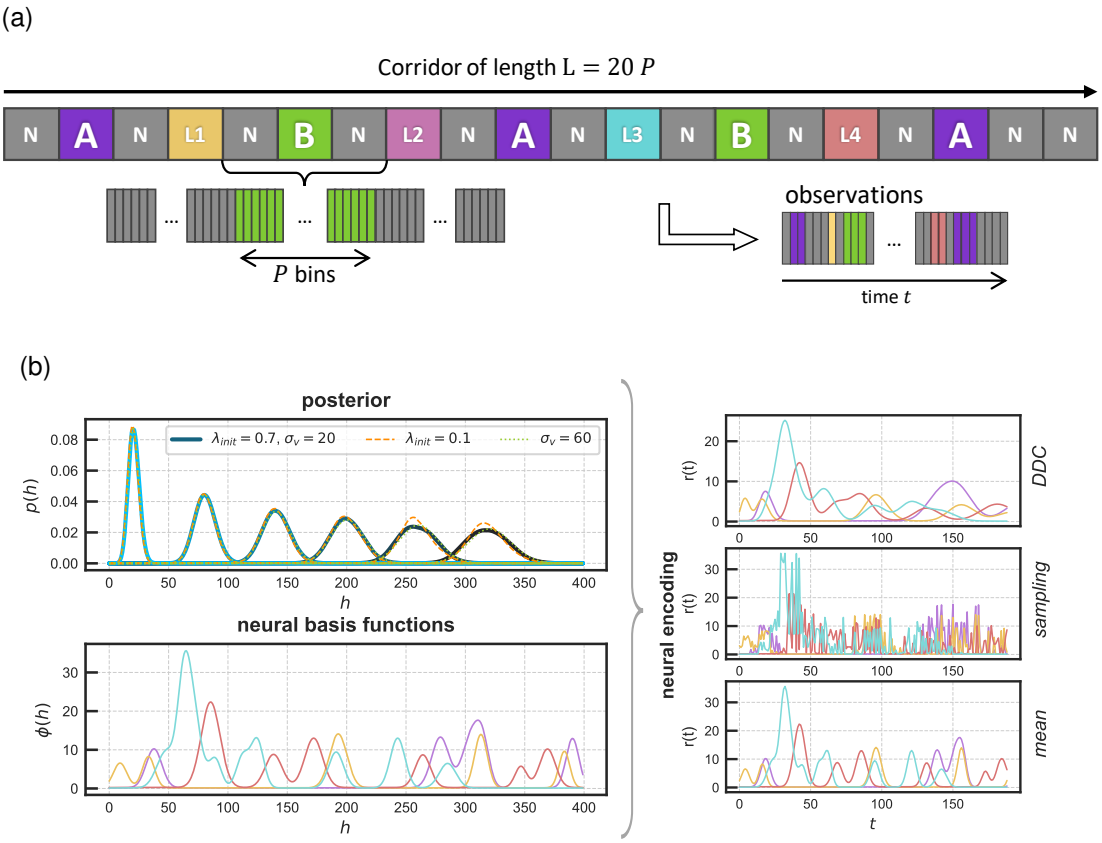

Figure 5.1: **Observations from one traversal through the corridor are used to infer posteriors that are encoded using different methods. (a)** The corridor setup follows the virtual environment structure of Fiser et al. (2016). Space is discretised with every stimulus patch comprising P discrete position bins. **(b)** For every time step t of a single trial traversal, spatial and visual posteriors are inferred and encoded using different neural encoding methods. All encoding methods use the same neural populations with possibly multimodal basis functions.

between the two models in terms of neural activity. Visual stimuli in the HHMM are then obtained by applying the linear transformation $\mathbf{A}$ defined by the second emission function to the visual representations (see Equation (3.8)). Transformation matrix $\mathbf{A}$ was sampled uniformly within $[0, 1]^3$ and accepted if $|\mathbf{A}| \neq 0$ to ensure that $\mathbf{A}$ is invertable. It was keep constant through all presented experiments and was sampled as follows[1]:

$$\mathbf{A} = \begin{pmatrix} 0.16 & 0.00031 & 0.22 \\ 0.37 & 0.0020 & 0.19 \\ 0.99 & 0.79 & 0.12 \end{pmatrix} \tag{5.1}$$

[1] Values are rounded to the first two significant decimal digits.

stimulus	stimulus values	
	HMM	*HHMM*
N	(127.5, 127.5, 127.5)	(48.1, 71.8, 242.1)
A	(127.5, 52.5, 202.5)	(64.3, 86.1, 192.2)
B	(127.5, 202.5, 52.5)	(79.2, 115.1, 310.2)
L1	(197.4, 111.9, 218.9)	(31.9, 57.5, 292.1)
L2	(177.8, 24.0, 248.8)	(82.5, 113.5, 225.3)
L3	(194.1, 200.4, 32.7)	(38.3, 78.1, 354.0)
L4	(114.8, 94.6, 236.3)	(69.7, 88.1, 216.9)

Table 5.1: **Stimulus values in three-dimensional visual space.** Stimulus values for the HMM (left) become visual representations in the HHMM and stimulus values of the HHMM (right) are a linear transformation of these representations given by $\mathbf{A}$ (Eq. (5.1)). Values are rounded to the first significant decimal digit.

Resulting HHMM stimulus values are defined in the right column of Table 5.1.

The simulation of an agents movement through the corridor follows the generative model presented in Section 3.1.1. The intended action represented by the parameter of the transition distribution was fixed to $\lambda_{trans} = 2$. In a small grid search analysis, the parameter of the initial probability distribution λ_{init} and the emission noise parameter $\sigma_{\mathbf{v}}$ – also called mapping noise standard deviation subsequently – were varied over a grid of values to test the effect of model uncertainty onto the encoded neural activity ($\lambda_{init} \in [0.1, 0.3, 0.5, 0.7, 0.9]$; $\sigma_{\mathbf{v}} \in [20, 40, 60]$). For the HHMM, the variance of the second emission distribution was kept constant at $\sigma_{\mathbf{o}} = 40$. For every parameter configuration, we run 50 trials that differed in motion trajectory and observation noise. Trial trajectories ended with the last sampled position within the defined corridor, i.e. $h_T < L$. Filtering was performed in log-space to avoid numerical underflow. Inferred posterior probabilities were then transformed back into the probability space for neural encoding.

The generation of neural populations for neural encoding followed the parameter choices of Ujfalussy and Orbán (2022). All three encoding methods used the same spatial and visual neural populations that comprised 1000 neurons each. The distribution over number of different subfields K in the basis function of a neuron i was given by a gamma distribution with parameters $\alpha = 0.57$ and $\beta = 1/0.14$ (Rich et al., 2014; Ujfalussy and Orbán, 2022). Sampled values for K were rejected for $K < 1$. Means μ_{ik} of subfield gaussians were sampled uniformly within the respective domain. For

the spatial neurons, this yields subfield centers that are uniformly distributed along the corridor, whereas visual subfield means are uniformly distributed in $[0, 255]^3$. Subfield variances were sampled uniformly within a fraction of the respective space considered in the experiments. In particular, spatial subfields spanning a corridor of length L had variances σ_{ik} uniformly distributed in range $[0.05 \cdot L, 0.15 \cdot L]$. As visual information is presented in a higher dimensional space, visual variances σ_{ik} were uniformly distributed within a larger fraction of the space and were sampled from $[0.05 \cdot 255, 0.45 \cdot 255]$. The maximal firing rate ρ_{ik} of a subfield is given by the gaussians' amplitude which was sampled uniformly between 5 Hz and 15 Hz. For every neuron i, subfield activities are added to a baseline firing rates ρ_i that were sampled uniformly between 0.1 Hz and 0.25 Hz.

It should be noted that we don't limit the range of subfield basis functions. As a result they cover values that are not represented in the current simulation setup. Figure 5.1b indeed shows that spatial basis functions extend beyond the edges of the corridor which may result in inaccurate representations at the end of the traversal.

All code used for the generative model, the filtering and encoding process, as well as the analysis was implemented over the course of this thesis.

Chapter 6

Discussion

Our HHM based model was designed to infer the unobserved spatial positions during a traversal through a corridor by observing the repeating visual stimuli. We did not account explicitly for any properties that could induce stimulus selectivity or position modulation. Furthermore, we ensured numerical balance between stimuli through the choice of the stimulus values and uniform sampling of all neural encoding parameters. In this chapter, we provide a brief analysis of our model's performance showing that it can indeed simulate neural selectivity and modulation. An initial parameter grid search also indicates that selectivity and modulation in visual responses foster with decreasing uncertainty in the model. Our results are in line with experimental findings in the literature and, thus, validate our approach as a potential candidate for future computational modelling studies of neural modulation in spatio-visual navigation tasks.

6.1 Analysis of selectivity and modulation

Figure 5.1b shows the components of the neural encoding for an example trial in the HHMM ($\sigma_{\mathbf{v}} = 20$, $\lambda_{init} = 0.7$). The inferred posterior over current position is shown for different time steps throughout the traversal ($t = [10, 40, 70, 100, 130, 160]$). As the agent moves down the corridor, the model is able to keep track of the position indicated by the mode of the posterior that shifts accordingly. However, the uncertainty about the most probable position increases constantly over time. The agents starts off with a fairly certain initial probability distribution as $\lambda_{init} < 1$ only yields a small variance over the initial position. By the end of the corridor, however, a larger proportion of the probability mass is distributed over multiple visual stimuli considering the patch size $P = 20$.

Changing the mapping noise standard deviation ($\sigma_{\mathbf{v}} = 60$) yields the same phenomenon as Figure 5.1b shows. Small shifts in the posteriors of later time steps can be explained by innovation noise. On the other hand, reducing the initial position uncertainty to $\lambda_{init} = 0.1$ increases the amplitude and decreases the width of the posterior distributions along the track. This indicates that the increase of posterior uncertainty is due to uncertainty accumulation by innovation noise and is influenced by initial uncertainty conditions. These observations are in line with the theory of path integration in mamals (McNaughton et al., 2006).

The change in uncertainty is reflected in the neural encoding of this spatial posterior. Figure 5.1b (left) shows the corresponding rates for all three encoding methods – DDC, sampling, and mean encoding – using the same set of neural basis functions from four selected neurons. Rates are represented as a function of time step t in the trial. We defined the mean encoding as baseline encoding framework that does not encode uncertainty. The presented example indeed shows that the change in uncertainty is not reflected in the mean encoding. Instead, the mean rates mimic the neural basis functions, as the mode of the posterior moves close to linearly along the corridor. The rates obtained by the DDC and sampling encoding, however, also encode the posterior variance and, consequently, reflect its change over time. The change in uncertainty representation can be observed, for example, by comparing the first two subfield activities of the orange neuron ($T < 30$) with the last two subfield activities of the red neuron ($t > 170$). Both subfields have similar modes and amplitudes in their basis functions. For the orange neuron, these are well reflected in the DDC encoding. For the red neuron, on the other hand, increased uncertainty at the end of the traversal blurs neural activity and the two subfield activities cannot be distinguished anymore. The sampling encoding reflects similar effects although they can not be observed as clearly due to the strong variability in response rates within one trial. These observations validate our employed neural encoding representations of the spatial posterior information.

To compare the overall simulated spatial response to results from Fiser et al. (2016), we plotted activities of spatial neurons that showed the most selectivity for either stimulus A or B ($\sigma_{\mathbf{v}} = 20, \lambda_{init} = 0.5$) and sorted them by peak position according to Figure 1g in Fiser et al. (see Supplementary Figure C.2b). Stimulus selectivity was determined using the selectivity index (SI) defined by Fiser et al. (2016) (see Equation (2.2)). Mean response rates for stimulus A and B where averaged over trials. Notably, this definitions yields a negative SI-value for B-selective neurons, and positive SI-value

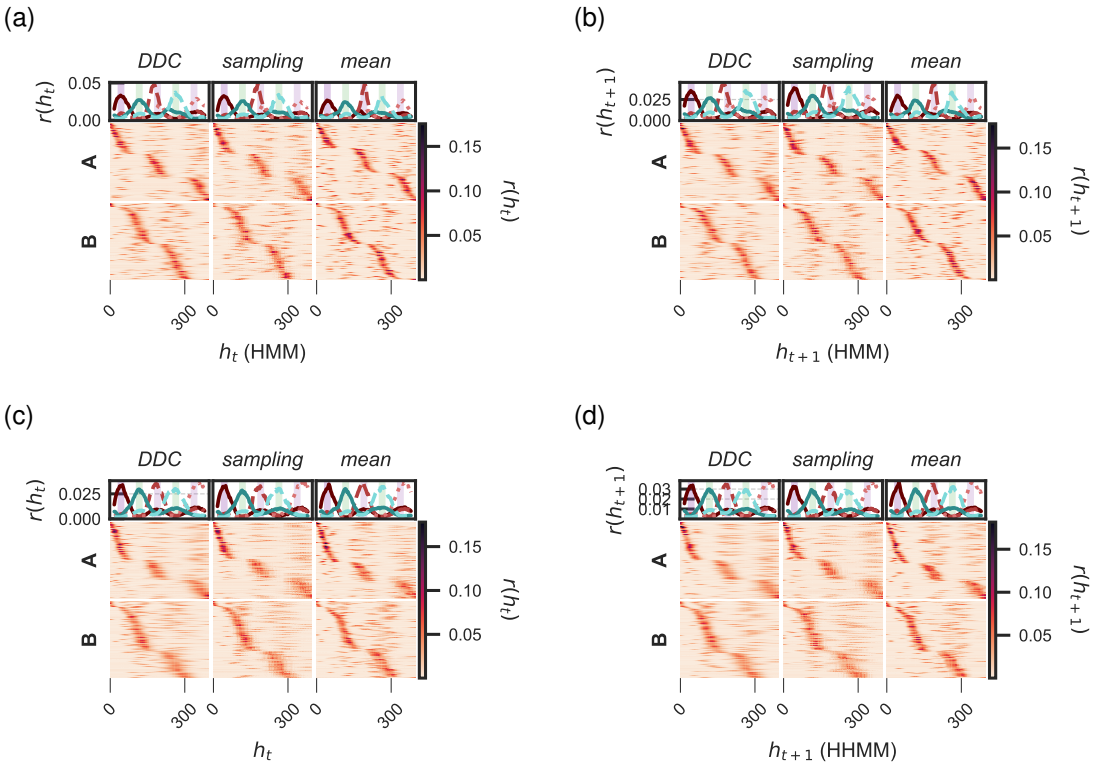

Figure 6.1: **Activities of 200 most stimulus selective spatial neurons sorted by peak position.** Similar to (Fiser et al., 2016) (see Supplementary Figure C.2b), spatial neurons respond differently for different stimulus positions in the corridor. This can be explained by the subfield based basis function in neural encoding that coincide with the stimulus position for these neurons. Uncertainty about the spatial position accumulates across motion trajectories and is reflected in blurred activities towards the end of the corridor.

for A-selective neurons. Figure 6.1 shows the neural activities for the 200 most A-selective and B-selective neurons, i.e. the 200 neurons with the highest/lowest SI-values. Activities were averaged over all 50 trials, binned in position intervals of size 4, and sorted by peak position.

The resulting plots allow for two main observations. First, the change of posterior uncertainty described above for a single example trial is consistent over trials and is reflected more in the encoding frameworks that account for uncertainty. Furthermore, the blurring of neural activity is increased in the HHMM over all encoding methods reflecting the additional observation uncertainty in the position estimation induced by the second emission probability distribution. Second, according to Fiser et al., the neurons exhibiting stimulus selectivity also change in response activity depending on

Figure 6.2: **Selectivity grid search results.** For all three encoding frameworks – (a) DDC, (b) sampling-based, and (c) mean encoding – the influence of changes in uncertainty parameters onto the visual population selectivity are evaluated by the mean absolute selectivity index SI. Stimulus selectivity in visual neurons seems to foster with decreasing uncertainty about the expected stimulus in the inferred position.

stimulus position and don't have a distinct onset response. This characteristic was mirrored by a reduction in the accuracy of their position classification analysis on spatial neurons. Our models successfully emulate these findings (see Figure 6.1). Since we haven't incorporated any specific model design to accommodate this selectivity, it emerges from a neural population with randomly sampled neural parameters. The subfield basis functions span an area that fortuitously coincides with the stimulus position. As a result, we anticipate a more pronounced onset responses for narrower subfield widths, and vice versa. In general, the ability of our model to replicate these plots serves as an indicator of its validity as a simulation tool for modelling spatial information spatio-visual navigation tasks.

To asses the extend to which our model can replicate selectivity in the visual neurons,

we conduct a gird search over the mapping uncertainty and initial position uncertainty. These two parameters are potential influences for the spatial top-down spatial information received by visual representations. As presented earlier, the initial position distribution modulates the posterior during corridor traversal and might therefore manifest in the posterior over visual representations, as it is defined using the corresponding spatial posterior (see Chapter 3). The information exchange between spatial and visual neurons is governed by a mapping that represents the corridor setup assuming full environmental knowledge. In an ideal scenario without any noise parameters, the model would have perfect information about what visual stimulus to expect in a certain positions. Consequently, increasing mapping noise results in the blurring of spatial stimulus information. We evaluate the influence of these parameters by calculating the mean of the absolute selectivity index $|SI|$ for all 1000 visual neurons over the grid. Grid search results for selectivity are presented in Figure 6.2, sorted by method and inferred parameter.

The clearest effect of parameter changes can be observed for the DDC encoding framework. While the change in the initial position uncertainty does not effect the selectivity of visual neurons, mean absolute SI increases exponentially with a linear decrease in mapping uncertainty. Further, selectivity is also increased for the visual information in the current time step compared to predictive posteriors that incorporate additional top-down uncertainty induced by the innovation noise distribution. This implies that visual neurons would have maximum selectivity for a optimal emission mapping with no uncertainty about the visual information in a location. The mean encoding supports this theory of the relation between uncertainty representations and selectivity as it shows no exponential relation between mapping uncertainty and visual selectivity. As it only encodes a point estimate of the visual posterior given by its mean, changes in variance are only slightly reflected in the amplitude of the neural responses since the mode amplitude changes with increasing posterior variance.

Selectivity results of the sampling encoding hints towards the same conclusion. For low emission mapping uncertainty, the selectivity of neurons is consistently high and similar in magnitude compared to the mean results. However, as the model uncertainty increases and visual posterior distributions get wider, the encoded point estimate is more likely to be sampled further away from the distribution mean. As a result, SI values significantly vary in magnitude for higher mapping noise standard deviation. Notably, the small inconsistencies in the mean encoding selectivity as well as these drastically increased variability of the grid search results for the sampling encoding

are explained by an analysis of the sample size (see Supplementary Figure C.3). To better approximate the effects of changes in uncertainty related parameters in these encoding methods, future analysis should increase the number of trials per experiment significantly.

In general, selectivity values are quite small indicating that most visual neurons actually don't exhibit stimulus selectivity. This can be explained by the uniform sampling of subfield means that more likely to cover the center of the defined visual space compared to the stimulus positions that are closer to the edge of this space. In future experiments, the mean sampling for visual basis functions could be changed to test account for more equally distributed coverage of all possible stimuli in the visual space.

Having established the existence of an effect in change of uncertainty on the selectivity of visual responses, we next analyse whether our model is also able to exhibit spatial modulation in stimulus related activity and how they reflect changes in model representations of uncertainty. As we don't classify neurons based on their stimulus selectivity, we cannot use the modulation ratio *MR* employed by Saleem et al. (2018). Instead, we introduce another quantification metric for spatial modulation that adapts the selectivity index *SI* used by Fiser et al. (2016), i.e. the *modulation index (MI)*:

$$MI = \frac{\overline{r}_{P1} - \overline{r}_{P2}}{\overline{r}_{P1} + \overline{r}_{P2}} \tag{6.1}$$

We denote $\overline{r}$ as the average response of a neuron in the first ($P1$) and second ($P2$) position of a stimulus in the corridor. Negative MI-values indicate a preference of the first position, and vice versa.

Figure 6.3 shows the cumulative distribution functions (CDF) of the distribution of MI over all visual neurons for stimulus A (red) and stimulus B (green) over the grid of tested paramters. Steeper CDFs indicate that less neurons exhibit spatial modulation of visual responses, with a step function shape indicating close to no visual modulation. However, as some curves mimic a sigmoidal function, these results show that our model is able to simulate spatial modulation in visual neurons. Similar to the selectivity analysis, initial position uncertainty did not yield any observable difference in neural modulation. For the DDC encoding, a linear decrease in mapping noise standard deviation also resulted in a exponential increase in MI magnitude, similar to the selectivity index and spatial modulation was, again, largest in the mean encoding method, for which it did not change for different emission uncertainty parameters. Additionally, the shape of the sampling CDFs only approximate a sigmoid for the smallest tested mapping

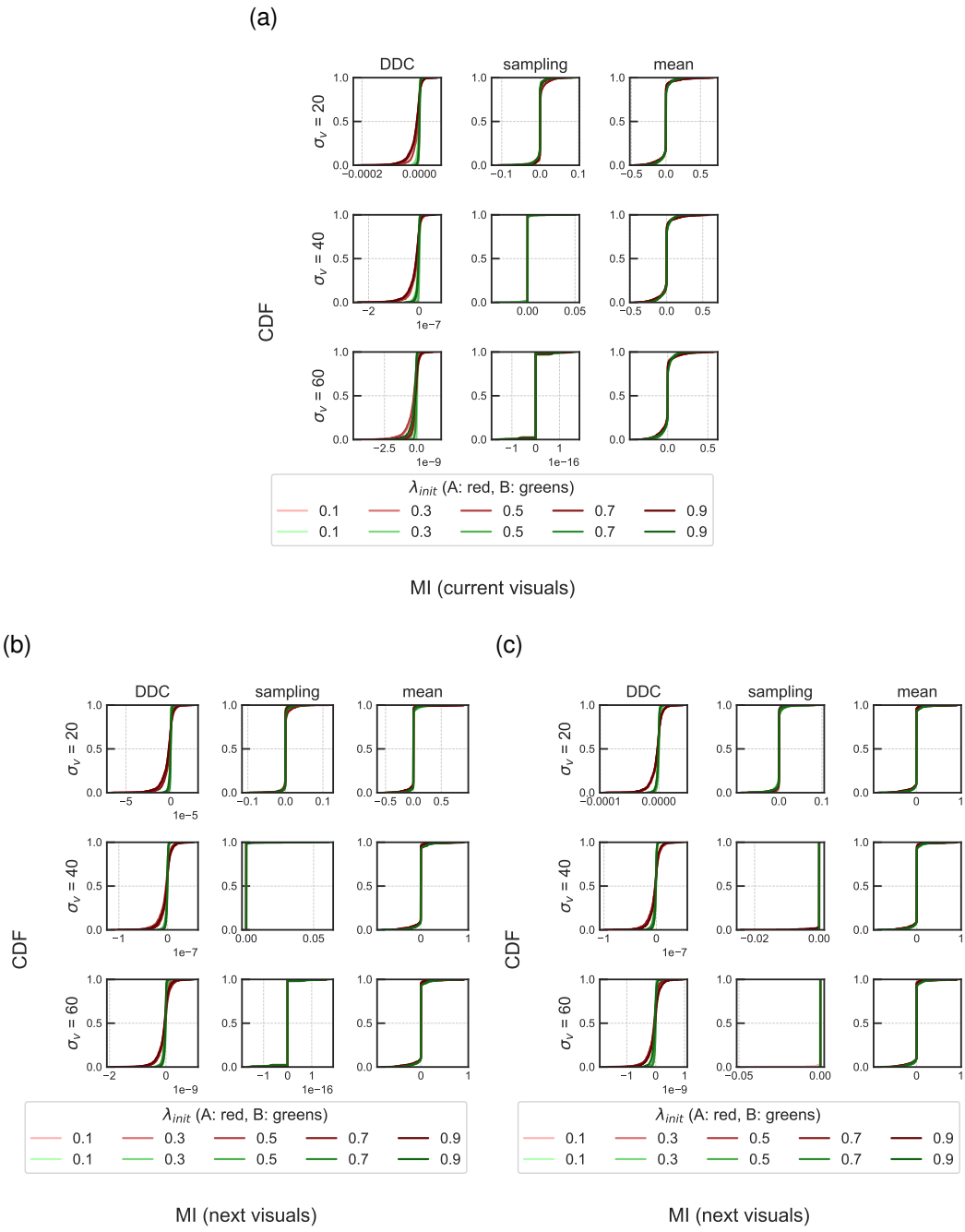

Figure 6.3: **Modulation grid search results.** For all three encoding frameworks, changes in uncertainty parameters onto the modulation of visual responses between the first and second position of stimulus representation are evaluated using the cumulative density function (CDF) of the modulation index MI. Activity of visual neurons becomes more distinct between presentation locations if the model uncertainty decreases. The difference between stimulus A (red) and stimulus B (green) reflects the evolution of the spatial posterior during traversals.

noise standard derivation, consistently representing that encoded samples are closer to the mean of the posterior distribution for smaller emission noise variance. These observations complement the conclusions drawn from the selectivity analysis: Spatial modulation fosters as uncertainty about the expected visual information decreases.

Notably, in the DDC framework, the *MI* CDFs for stimulus B are consistently steeper than for stimulus A, indicating more spatial modulation in the latter. This most likely reflects the uncertainty evolution of the spatial posterior that was described above, which is most prominent in the DDC encoding. At the beginning of the corridor, the agent is still quite certain about its position and variance in spatial posterior distributions is low, but then increases as the agent moves along the corridor and has to integrate innovation noise over the movement trajectory to estimate its actual position. As the position estimation of the spatial posterior distribution is reflected in the visual posterior, the response difference between the the first and second stimulus presentation is larger for stimulus A that is presented before stimulus B. However, to test this theory, experiments should be repeated using a different stimulus order.

In summary, we showed that our model can exhibit both selectivity and modulation in visual responses. Our results are in line with reported findings in experimental studies as spatial modulation in visual neurons reflects subjective position estimation along the corridor traversal (Saleem et al., 2018) and fosters with decreasing uncertainty about the environmental structure that can be considered to simulate learning through experience (Fiser et al., 2016).

6.2 Conclusion and future research directions

The previous section showed that our proposed HHM-based modeling approach of spatio-visual navigation tasks can simulate stimulus selectivity and spatial modulation in visual neurons. However, we only provide an initial analysis of how these effects are influenced by different parameter settings, focussing on uncertainty-related model parameters. Further analysis is required to establish a better understanding of these influences. Besides extending the parameter space of the grid search for the initial position uncertainty and mapping noise standard derivation, these analyses should also test the influence of the observation noise in the second emission probability distribution for the HHMM on the selectivity and modulation of the visual neurons. The metrics for analysing these quantities should also be extended by a position classification analysis that was performed by both, Fiser et al. (2016) and Saleem et al. (2018), to quantify

the spatial information entailed in the visual responses. For all analysis purposes, the number of trials per experiment should be increased to obtain a better approximation of true *SI* and *MI* values, especially for the sampling encoding framework.

To further examine the spatial information entailed in the visual representations of our model, the HHMM should be compared to a non-spatial baseline model. Such a baseline could be provided by a purely visual model of receptive field activity in the visual cortex that was employed by Saleem et al. (2018) as a null-hypothesis model for spatial modulation analysis. To allow for better comparison between this model and our spatially informed approach, visual representations of the HHMM also could be changed to follow the receptive field structure of V1. Although such a more biological plausible representation of V1 representations would be desirable, the mapping from spatial to visual observations as well as the process of inverting the observation generation in such a modified generative model may not be trivial.

In terms of biological plausibility, the simplified representation of visual information in V1 is not the only flaw our model has. Up to now, it also does not represent memory within and between trials. Such a memory representation would optimally include a representation of time that passes within a motion trajectory. For example, loosing memory about previous observations could be modeled using a decay parameter and/or a memory cache that can extend beyond trials. Observations stored in the memory could then be used to model learning through experience, providing a more detailed approach to simulate findings and fit data in experiments similar to Fiser et al. (2016). If we interpret learning as gaining better knowledge about the environment and, thus, reducing uncertainty, the optimal fit for experience-dependent modulation reported by (Fiser et al., 2016) should show a decrease of the fitted mapping uncertainty over time.

Changing the uncertainty parameters of the model is the only way our approach allows to simulate learning so far. In theory, future research could implement learning the generative model of the HHMM by fitting a recognition model using variational learning algorithms. Given the non-linearity introduced by the mapping of spatial onto visual information, approximation of the latent posterior distributions may not be trivial, though. As an alternative approach, learning could be models by extending the hierarchical probabilistic structure of the HHMM. Instead of fitting parameters of uncertainty representation, one could fit distributions over these parameters as a function of time. Inference in such a model may build on the equations provided in this thesis and would not require the learning in the probabilistic sense. The model would still have full knowledge of the optimal parameters, but becomes increasingly certain

(one might say ”aware”) of that knowledge.

Finally, we want to point out another possible research direction related to how learning may be realised in our HHMM. By model design via the definition of conditional dependency structures, the visual posteriors of both our models have spatial information provided by the emission mapping. In the hierarchical model, additional bottom-up information about the stimulus is provided by the observation. Referring to definition of the posterior over the current visual representation in the HHMM in Equation (3.15), the term $\mathcal{N}\left(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\right)$ could be interpreted as bottom-up signals that invert the linear transformation in the generative model. Accordingly, the term $\mathcal{N}\left(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t), \Sigma'_{\mathbf{v}_t}\right)\omega(h_t)$ could be interpreted as the spatially informed top-down signal that holds expectations about visual representations in this time step. Summing over the product of these two signals for all positions h_t, mimics a convolution of the two Gaussians and, thus, corresponds to calculating a mismatch signal between spatially informed expectations and stimulus informed observations. Interpreting this mismatch signal as a form of prediction error, we believe that our model could actually be used to implement learning and inference in the context of predictive coding by minimising the prediction errors in the visual representations of the hierarchical structure.

In fact, the HHMM might actually already implement a weak version of predictive inference, although it only defines conditional independences and performs inference by adapting the commonly used filtering algorithm for HMMs. According to the testing conditions in Fiser et al. (2016), future experiments should examine this by varying the observations in the final stimulus position which is fixed to stimulus A in our experiment, without changing the corresponding mapping for the inference process, and evaluate whether the resulting posterior over the visual representations can already be interpreted as prediction errors. However, probabilistic inference through predictions does not necessarily imply the implementation of predictive coding (Aitchison and Lengyel, 2017). Therefore, examining whether our inference equations could represent a exact correspondence of the predictive coding framework also remains subject to future research. However, if this correspondence can be established, our hierarchical HHM-based model provides a computational framework to test predictive coding as a potential explanation of observed spatial modulation in visual activity, as proposed by Fiser et al. (2016).

Bibliography

Aitchison, L. and Lengyel, M. (2017). With or without you: predictive coding and bayesian inference in the brain. *Current opinion in neurobiology*, 46:219–227.

Alais, D. and Burr, D. (2019). Cue combination within a bayesian framework. *Multisensory processes: The auditory perspective*, pages 9–31.

Ali, D., Touqir, I., Siddiqui, A. M., Malik, J., and Imran, M. (2022). Face recognition system based on four state hidden markov model. *IEEE Access*, 10:74436–74448.

Barber, D. (2012). *Bayesian reasoning and machine learning*. Cambridge University Press.

Ben-Yishai, R., Bar-Or, R. L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences*, 92(9):3844–3848.

Berkes, P., Orbán, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, 331(6013):83–87.

Brascamp, J. W. and Shevell, S. K. (2021). The certainty of ambiguity in visual neural representations. *Annual Review of Vision Science*, 7:465–486.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., and Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, 280(5364):747–749.

Chen, G., King, J. A., Burgess, N., and O'Keefe, J. (2013). How vision and movement combine in the hippocampal place code. *Proceedings of the National Academy of Sciences*, 110(1):378–383.

Chen, S., Langley, J., Chen, X., and Hu, X. (2016). Spatiotemporal modeling of brain dynamics using resting-state functional magnetic resonance imaging with gaussian hidden markov model. *Brain connectivity*, 6(4):326–334.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, 36(3):181–204.

Dash, D. P., Kolekar, M. H., and Jha, K. (2020). Multi-channel eeg based automatic epileptic seizure detection using iterative filtering decomposition and hidden markov model. *Computers in biology and medicine*, 116:103571.

Diamanti, E. M., Reddy, C. B., Schröder, S., Muzzu, T., Harris, K. D., Saleem, A. B., and Carandini, M. (2021). Spatial modulation of visual responses arises in cortex with active navigation. *elife*, 10:e63705.

Ellis, K. (2023). Modeling human-like concept learning with bayesian inference over natural language. *arXiv preprint arXiv:2306.02797*.

Ferster, D. and Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annual review of neuroscience*, 23(1):441–471.

Fiser, A., Mahringer, D., Oyibo, H. K., Petersen, A. V., Leinweber, M., and Keller, G. B. (2016). Experience-dependent spatial expectations in mouse visual cortex. *Nature neuroscience*, 19(12):1658–1664.

Fletcher, P. C. and Frith, C. D. (2009). Perceiving is believing: a bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, 10(1):48–58.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, 360(1456):815–836.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2):127–138.

Gales, M., Young, S., et al. (2008). The application of hidden markov models in speech recognition. *Foundations and Trends® in Signal Processing*, 1(3):195–304.

Geiller, T., Fattahi, M., Choi, J.-S., and Royer, S. (2017). Place cells are more strongly tied to landmarks in deep than in superficial ca1. *Nature communications*, 8(1):14531.

Haggerty, D. C. and Ji, D. (2015). Activities of visual cortical and hippocampal neurons co-fluctuate in freely moving rats during spatial behavior. *elife*, 4:e08902.

Harvey, C. D., Collman, F., Dombeck, D. A., and Tank, D. W. (2009). Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature*, 461(7266):941–946.

Hok, V., Lenck-Santini, P.-P., Roux, S., Save, E., Muller, R. U., and Poucet, B. (2007). Goal-related activity in hippocampal place cells. *Journal of Neuroscience*, 27(3):472–482.

Hoyer, P. and Hyvärinen, A. (2002). Interpreting neural response variability as monte carlo sampling of the posterior. *Advances in Neural Information Processing Systems*, 15:277–284.

Ji, D. and Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature neuroscience*, 10(1):100–107.

Juang, B. H. and Rabiner, L. R. (1991). Hidden markov models for speech recognition. *Technometrics*, 33(3):251–272.

Katahira, K., Nishikawa, J., Okanoya, K., and Okada, M. (2010). Extracting state transition dynamics from multiple spike trains using hidden markov models with correlated poisson distribution. *Neural Computation*, 22(9):2369–2389.

Krantz, D. H. (1975). Color measurement and color theory: I. representation theorem for grassmann structures. *Journal of Mathematical Psychology*, 12(3):283–303.

Lange, R. D., Shivkumar, S., Chattoraj, A., and Haefner, R. M. (2020). Bayesian encoding and decoding as distinct perspectives on neural coding. *BioRxiv*, pages 2020–10.

Lenck-Santini, P.-P., Muller, R. U., Save, E., and Poucet, B. (2002). Relationships between place cell firing fields and navigational decisions by rats. *Journal of Neuroscience*, 22(20):9035–9047.

Lenck-Santini, P.-P., Save, E., and Poucet, B. (2001). Evidence for a relationship between place-cell spatial firing and spatial memory performance. *Hippocampus*, 11(4):377–390.

Liu, X. and Cheng, T. (2003). Video-based face recognition using adaptive hidden markov models. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 1, pages I–I. IEEE.

Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature neuroscience*, 9(11):1432–1438.

Maviel, T., Durkin, T. P., Menzaghi, F., and Bontempi, B. (2004). Sites of neocortical reorganization critical for remote spatial memory. *Science*, 305(5680):96–99.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., and Moser, M.-B. (2006). Path integration and the neural basis of the'cognitive map'. *Nature Reviews Neuroscience*, 7(8):663–678.

Mor, B., Garhwal, S., and Kumar, A. (2021). A systematic review of hidden markov models and their applications. *Archives of computational methods in engineering*, 28:1429–1448.

Muller, R. U. and Kubie, J. L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *Journal of Neuroscience*, 7(7):1951–1968.

Neitz, J. and Jacobs, G. H. (1986). Polymorphism of the long-wavelength cone in normal human colour vision. *Nature*, 323(6089):623–625.

Niell, C. M. and Stryker, M. P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*, 65(4):472–479.

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain research*.

O'Keefe, J. and Speakman, A. . (1987). Single unit activity in the rat hippocampus during a spatial memory task. *Experimental brain research*, 68:1–27.

Orbán, G., Berkes, P., Fiser, J., and Lengyel, M. (2016). Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron*, 92(2):530–543.

Ou, J., Xie, L., Jin, C., Li, X., Zhu, D., Jiang, R., Chen, Y., Zhang, J., Li, L., and Liu, T. (2015). Characterizing and differentiating brain state dynamics via hidden markov models. *Brain topography*, 28:666–679.

Pellicano, E. and Burr, D. (2012). When the world becomes 'too real': a bayesian explanation of autistic perception. *Trends in cognitive sciences*, 16(10):504–510.

Pouget, A., Dayan, P., and Zemel, R. S. (2003). Inference and computation with population codes. *Annual review of neuroscience*, 26(1):381–410.

Powell, G., Meredith, Z., McMillin, R., and Freeman, T. C. (2016). Bayesian models of individual differences: Combining autistic traits and sensory thresholds to predict motion perception. *Psychological science*, 27(12):1562–1572.

Quinn, A. J., Vidaurre, D., Abeysuriya, R., Becker, R., Nobre, A. C., and Woolrich, M. W. (2018). Task-evoked dynamic network analysis through hidden markov modeling. *Frontiers in neuroscience*, 12:603.

Radons, G., Becker, J., Dülfer, B., and Krüger, J. (1994). Analysis, classification, and coding of multielectrode spike trains with hidden markov models. *Biological cybernetics*, 71(4):359–373.

Rao, R. P. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1):79–87.

Ravassard, P., Kees, A., Willers, B., Ho, D., Aharoni, D., Cushman, J., Aghajan, Z. M., and Mehta, M. R. (2013). Multisensory control of hippocampal spatiotemporal selectivity. *Science*, 340(6138):1342–1346.

Recanatesi, S., Farrell, M., Lajoie, G., Deneve, S., Rigotti, M., and Shea-Brown, E. (2021). Predictive learning as a network mechanism for extracting low-dimensional latent space representations. *Nature communications*, 12(1):1417.

Rich, P. D., Liaw, H.-P., and Lee, A. K. (2014). Large environments reveal the statistical structure governing hippocampal representations. *Science*, 345(6198):814–817.

Rosenzweig, E. S., Redish, A. D., McNaughton, B. L., and Barnes, C. A. (2003). Hippocampal map realignment and spatial learning. *Nature neuroscience*, 6(6):609–615.

Sahani, M. and Dayan, P. (2003). Doubly distributional population codes: simultaneous representation of uncertainty and multiplicity. *Neural Computation*, 15(10):2255–2279.

Saleem, A. B., Ayaz, A., Jeffery, K. J., Harris, K. D., and Carandini, M. (2013). Integration of visual motion and locomotion in mouse visual cortex. *Nature neuroscience*, 16(12):1864–1869.

Saleem, A. B., Diamanti, E. M., Fournier, J., Harris, K. D., and Carandini, M. (2018). Coherent encoding of subjective spatial position in visual cortex and hippocampus. *Nature*, 562(7725):124–127.

Schmack, K., Schnack, A., Priller, J., and Sterzer, P. (2015). Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. *Schizophrenia Research: Cognition*, 2(2):72–77.

Shams, L., Kamitani, Y., and Shimojo, S. (2000). What you see is what you hear. *Nature*, 408(6814):788–788.

Shapiro, M. L., Tanila, H., and Eichenbaum, H. (1997). Cues that hippocampal place cells encode: dynamic and hierarchical representation of local and distal stimuli. *Hippocampus*, 7(6):624–642.

Stankevicius, A., Huys, Q. J., Kalra, A., and Seriès, P. (2014). Optimism as a prior belief about the probability of future reward. *PLoS computational biology*, 10(5):e1003605.

Sun, S., Yu, R., and Wang, S. (2017). A neural signature encoding decisions under perceptual ambiguity. *eneuro*, 4(6).

Teixeira, C. M., Pomedli, S. R., Maei, H. R., Kee, N., and Frankland, P. W. (2006). Involvement of the anterior cingulate cortex in the expression of remote spatial memory. *Journal of Neuroscience*, 26(29):7555–7564.

Tenenbaum, J. (1999). Rules and similarity in concept learning. *Advances in neural information processing systems*, 12.

Ujfalussy, B. B. and Orbán, G. (2022). Sampling motion trajectories during hippocampal theta sequences. *Elife*, 11:e74058.

Vértes, E. (2020). *Probabilistic learning and computation in brains and machines*. PhD thesis, UCL (University College London). UCL Discovery. `https://discovery.ucl.ac.uk/id/eprint/10103090/`.

Vértes, E. and Sahani, M. (2018). Flexible and accurate inference and learning for deep generative models. *Advances in Neural Information Processing Systems*, 31.

Von Helmholtz, H. (1867). *Handbuch der physiologischen Optik: mit 213 in den Text eingedruckten Holzschnitten und 11 Tafeln*, volume 9. Voss.

Wang, Q., Gao, E., and Burkhalter, A. (2011). Gateways of ventral and dorsal streams in mouse visual cortex. *Journal of Neuroscience*, 31(5):1905–1918.

Wiener, S. I., Korshunov, V. A., Garcia, R., and Berthoz, A. (1995). Inertial, substratal and landmark cue control of hippocampal ca1 place cell activity. *European Journal of Neuroscience*, 7(11):2206–2219.

Zemel, R. S., Dayan, P., and Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural computation*, 10(2):403–430.

Appendix A

Detailed derivations - basic model

A.1 Derivation of future observation DDC equation

$$r_i(\mathbf{v}_{t+1}) = \int_{-\infty}^{\infty} \phi_i(\mathbf{v}_{t+1})\, p(\mathbf{v}_{t+1}|\mathbf{v}_{1:t})\, d\mathbf{v}_{t+1} \tag{A.1a}$$

$$= \int_{-\infty}^{\infty} \phi_i(\mathbf{v}_{t+1}) \Big(\sum_{h_t, h_{t+1}} p(v_{t+1}|h_{t+1})\, p(h_{t+1}|h_t)\, p(h_t|\mathbf{v}_{1:t}) \Big)\, d\mathbf{v}_{t+1} \tag{A.1b}$$

$$= \int_{-\infty}^{\infty} \Big(\sum_{h_t, h_{t+1}} \phi_i(\mathbf{v}_{t+1})\, p(v_{t+1}|h_{t+1})\, p(h_{t+1}|h_t)\, p(h_t|\mathbf{v}_{1:t}) \Big)\, d\mathbf{v}_{t+1} \tag{A.1c}$$

$$\overset{1}{=} \sum_{h_t, h_{t+1}} \left[\int_{-\infty}^{\infty} \phi_i(\mathbf{v}_{t+1})\, p(v_{t+1}|h_{t+1})\, p(h_{t+1}|h_t)\, p(h_t|\mathbf{v}_{1:t})\, d\mathbf{v}_{t+1} \right] \tag{A.1d}$$

$$\overset{2}{=} \sum_{h_t, h_{t+1}} \left[\int_{-\infty}^{\infty} \left(\rho_{i0} + \sum_{k=1}^{K_i} \rho_{ik}\, \psi_{ik}(\mathbf{v}_{t+1}) \right) p(v_{t+1}|h_{t+1})\, p(h_{t+1}|h_t)\, p(h_t|\mathbf{v}_{1:t})\, d\mathbf{v}_{t+1} \right] \tag{A.1e}$$

$$= \sum_{h_t, h_{t+1}} \left[p(h_{t+1}|h_t)\, p(h_t|\mathbf{v}_{1:t}) \int_{-\infty}^{\infty} \left(\rho_{i0} + \sum_{k=1}^{K_i} \rho_{ik}\, \psi_{ik}(\mathbf{v}_{t+1}) \right) p(v_{t+1}|h_{t+1})\, d\mathbf{v}_{t+1} \right] \tag{A.1f}$$

$$= \sum_{h_t, h_{t+1}} \left[p(h_{t+1}|h_t)\, p(h_t|\mathbf{v}_{1:t}) \left(\int_{-\infty}^{\infty} \rho_{i0}\, p(v_{t+1}|h_{t+1})\, d\mathbf{v}_{t+1} + \int_{-\infty}^{\infty} \sum_{k=1}^{K_i} \rho_{ik} \psi_{ik}(\mathbf{v}_{t+1})\, p(v_{t+1}|h_{t+1})\, d\mathbf{v}_{t+1} \right) \right] \tag{A.1g}$$

$$\stackrel{3}{=} \sum_{h_t,h_{t+1}} \Bigg[p(h_{t+1}|h_t)\, p(h_t|\mathbf{v}_{1:t}) \left(\rho_{i0} + \int_{-\infty}^{\infty} \sum_{k=1}^{K_i} \rho_{ik} \psi_{ik}(\mathbf{v}_{t+1})\, p(v_{t+1}|h_{t+1})\, d\mathbf{v}_{t+1} \right) \Bigg] \tag{A.1h}$$

$$\stackrel{1}{=} \sum_{h_t,h_{t+1}} \Bigg[p(h_{t+1}|h_t)\, p(h_t|\mathbf{v}_{1:t}) \left(\rho_{i0} + \sum_{k=1}^{K_i} \int_{-\infty}^{\infty} \rho_{ik} \psi_{ik}(\mathbf{v}_{t+1}) p(v_{t+1}|h_{t+1})\, d\mathbf{v}_{t+1} \right) \Bigg] \tag{A.1i}$$

$$\stackrel{4}{=} \sum_{h_t,h_{t+1}} \Bigg[p(h_{t+1}|h_t)\, p(h_t|\mathbf{v}_{1:t}) \left(\rho_{i0} + \sum_{k=1}^{K_i} \frac{\rho_{ik} \sqrt{|\Gamma_{ik}|}}{\sqrt{|\Gamma_{ik} + \Sigma_{\mathbf{v}}|}} \exp\Big(-\frac{1}{2} \big(\mu_{ik} - f(h_{t+1})\big)^{\top} \big(\Gamma_{ik} + \Sigma_{\mathbf{v}}\big)^{-1} \big(\mu_{ik} - f(h_{t+1})\big) \Big) \right) \Bigg] \tag{A.1j}$$

$$= \sum_{h_{t+1}} \Bigg[p(h_{t+1}|\mathbf{v}_{1:t}) \left(\rho_{i0} + \sum_{k=1}^{K_i} \rho_{ik} \sqrt{(2\pi)^D |\Gamma_{ik}|}\, \mathcal{N}\Big(\mu_{ik}; f(h_{t+1}), \Gamma_{ik} + \Sigma_{\mathbf{v}}\Big) \right) \Bigg] \tag{A.1k}$$

1. (TONELLI'S THEOREM) IF $\forall n,x : f_n(x) \geq 0 \Leftrightarrow \int \sum_n f_n(x)\, dx = \sum_n \int f_n(x)\, dx$
2. Definition Tuning Curves, see Equation 4.4
3. $\forall$ PROBABILITY DISTRIBUTIONS $p(x) : \int_{\mathcal{R}} p(x) dx = 1$
4. Derivation of subfield integral:

$$\int_{-\infty}^{\infty} \rho_{ik} \psi_{ik}(\mathbf{v}_{t+1}) p(v_{t+1}|h_{t+1})\, d\mathbf{v}_{t+1} \tag{A.2a}$$

$$\stackrel{4a}{=} \rho_{ik} \int_{-\infty}^{\infty} \sqrt{(2\pi)^D |\Gamma_{ik}|}\, \frac{\psi_{ik}(\mathbf{v}_{t+1})}{\sqrt{(2\pi)^D |\Gamma_{ik}|}} p(v_{t+1}|h_{t+1})\, d\mathbf{v}_{t+1} \tag{A.2b}$$

$$\stackrel{4b}{=} \rho_{ik} \sqrt{(2\pi)^D |\Gamma_{ik}|} \int_{-\infty}^{\infty} \mathcal{N}\Big(\mathbf{v}_{t+1}; \mu_{ik}, \Gamma_{ik}\Big)\, \mathcal{N}\Big(\mathbf{v}_{t+1}; f(h_{t+1}), \Sigma_{\mathbf{v}}\Big)\, d\mathbf{v}_{t+1} \tag{A.2c}$$

$$\stackrel{4c}{=} \rho_{ik} \sqrt{(2\pi)^D |\Gamma_{ik}|} \int_{-\infty}^{\infty} \mathcal{N}\Big(\mathbf{v}_{t+1}; \mu', \Sigma'\Big)\, \mathcal{N}\Big(\mu_{ik}; f(h_{t+1}), \Gamma_{ik} + \Sigma_{\mathbf{v}}\Big)\, d\mathbf{v}_{t+1} \tag{A.2d}$$

$$= \rho_{ik} \sqrt{(2\pi)^D |\Gamma_{ik}|}\, \mathcal{N}\Big(\mu_{ik}; f(h_{t+1}), \Gamma_{ik} + \Sigma_{\mathbf{v}}\Big) \int_{-\infty}^{\infty} \mathcal{N}\Big(\mathbf{v}_{t+1}; \mu', \Sigma'\Big)\, d\mathbf{v}_{t+1} \tag{A.2e}$$

$$\stackrel{3}{=} \rho_{ik} \sqrt{(2\pi)^D |\Gamma_{ik}|}\, \mathcal{N}\Big(\mu_{ik}; f(h_{t+1}), \Gamma_{ik} + \Sigma_{\mathbf{v}}\Big) \tag{A.2f}$$

$$\overset{(4.2)}{=} \frac{\rho_{ik}\sqrt{(2\pi)^D|\Gamma_{ik}|}}{\sqrt{(2\pi)^D|\Gamma_{ik}+\Sigma_{\mathbf{v}}|}} \exp\Big(-\frac{1}{2}\big(\mu_{ik}-f(h_{t+1})\big)^\top\big(\Gamma_{ik}+\Sigma_{\mathbf{v}}\big)^{-1}\big(\mu_{ik}-f(h_{t+1})\big)\Big) \tag{A.2g}$$

$$= \frac{\rho_{ik}\sqrt{|\Gamma_{ik}|}}{\sqrt{|\Gamma_{ik}+\Sigma_{\mathbf{v}}|}} \exp\Big(-\frac{1}{2}\big(\mu_{ik}-f(h_{t+1})\big)^\top\big(\Gamma_{ik}+\Sigma_{\mathbf{v}}\big)^{-1}\big(\mu_{ik}-f(h_{t+1})\big)\Big) \tag{A.2h}$$

(a) $\psi_{ik}(\mathbf{v}_{t+1})$ is Gaussian with mean μ_{ik} and cov. mat. $\Gamma_{ik} \Rightarrow$ adding normalisation constant yields a probability distribution (normal distribution)

(b) Define normal distributions:

$$\frac{\psi_{ik}(\mathbf{v}_{t+1})}{\sqrt{(2\pi)^D|\Gamma_{ik}|}} =: \mathcal{N}\Big(\mathbf{v}_{t+1};\mu_{ik},\Gamma_{ik}\Big) \tag{A.3}$$

$$p(v_{t+1}|h_{t+1}) =: \mathcal{N}\Big(\mathbf{v}_{t+1};f(h_{t+1}),\Sigma_{\mathbf{v}}\Big) \tag{A.4}$$

(c) GAUSSIANS ARE CLOSED UNDER MULTIPLICATION

$$\mathcal{N}\big(x;a,A\big)\,\mathcal{N}\big(x;b,B\big) = \mathcal{N}\big(x;c,C\big)\,Z$$

WITH $C = \big(A^{-1}+B^{-1}\big)^{-1}$, $c = C(A^{-1}a+B^{-1}b)$ AND $Z = \mathcal{N}\big(a;b,A+B\big)$

Note: $\sqrt{|\Gamma_{ik}|} = \prod_j^D \sqrt{\gamma_{jj}}$ and $\sqrt{|\Gamma_{ik}+\Sigma_{\mathbf{v}}|} = \prod_j^D \sqrt{\gamma_{jj}+\sigma_{jj}}$.
Special case: Γ_{ik} and Σ_{ik} are diagonal (no correlations between stimulus dimensions)

$$\forall\gamma_{jj} : \gamma = \gamma_{jj} \Rightarrow \prod_j^D \sqrt{\gamma_{jj}} = \big(\sqrt{\gamma}\big)^D$$

$$\forall\gamma_{jj} : \gamma = \gamma_{jj}, \forall\sigma_{jj} : \sigma = \sigma_{jj} \Rightarrow \prod_j^D \sqrt{\gamma_{jj}+\sigma_{jj}} = \big(\sqrt{\gamma+\sigma}\big)^D$$

Appendix B

Detailed derivations - hierarchical model

B.1 Change Gaussian over $\mathbf{o}_t$ to Gaussian over $\mathbf{v}_t$

To derive the recursive forwarding algorithm, we need to express the second emission distribution as a function the visuals, i.e. a distribution over $\mathbf{v}_t$.

$$\mathcal{N}(\mathbf{o}_t; \mathbf{A}\mathbf{v}_t, \sigma_o^2 \mathbb{I}) \stackrel{1}{=} \mathcal{N}(\mathbf{A}\mathbf{v}_t; \mathbf{o}_t, \sigma_o^2 \mathbb{I}) \tag{B.1a}$$

$$= \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_o|^{\frac{1}{2}}} \exp\Big[-\frac{1}{2} (\mathbf{A}\mathbf{v}_t - \mathbf{o}_t)^\top \Sigma_o^{-1} (\mathbf{A}\mathbf{v}_t - \mathbf{o}_t) \Big] \tag{B.1b}$$

$$= \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_o|^{\frac{1}{2}}} \exp\Big[-\frac{1}{2} (\mathbf{A}\mathbf{v}_t - \mathbf{A}\mathbf{A}^{-1}\mathbf{o}_t)^\top \Sigma_o^{-1} (\mathbf{A}\mathbf{v}_t - \mathbf{A}\mathbf{A}^{-1}\mathbf{o}_t) \Big] \tag{B.1c}$$

$$= \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_o|^{\frac{1}{2}}} \exp\Big[-\frac{1}{2} \big(\mathbf{A}(\mathbf{v}_t - \mathbf{A}^{-1}\mathbf{o}_t)\big)^\top \Sigma_o^{-1} \big(\mathbf{A}(\mathbf{v}_t - \mathbf{A}^{-1}\mathbf{o}_t)\big) \Big] \tag{B.1d}$$

$$= \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_o|^{\frac{1}{2}}} \exp\Big[-\frac{1}{2} (\mathbf{v}_t - \mathbf{A}^{-1}\mathbf{o}_t)^\top \mathbf{A}^\top \Sigma_o^{-1} \mathbf{A} (\mathbf{v}_t - \mathbf{A}^{-1}\mathbf{o}_t) \Big] \tag{B.1e}$$

$$\stackrel{2}{=} \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_o|^{\frac{1}{2}}} \exp\Big[-\frac{1}{2} (\mathbf{v}_t - \mathbf{A}^{-1}\mathbf{o}_t)^\top (\mathbf{A}^{-1}\Sigma(\mathbf{A}^{-1})^\top)^{-1} (\mathbf{v}_t - \mathbf{A}^{-1}\mathbf{o}_t) \Big] \tag{B.1f}$$

$$= \frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma_o|^{\frac{1}{2}}} \cdot \frac{(2\pi)^{\frac{D'}{2}}|\mathbf{A}^{-1}\Sigma(\mathbf{A}^{-1})^\top|^{\frac{1}{2}}}{(2\pi)^{\frac{D'}{2}}|\mathbf{A}^{-1}\Sigma(\mathbf{A}^{-1})^\top|^{\frac{1}{2}}} \tag{B.1g}$$
$$\exp\Big[-\frac{1}{2}(\mathbf{v}_t - \mathbf{A}^{-1}\mathbf{o}_t)^\top(\mathbf{A}^{-1}\Sigma(\mathbf{A}^{-1})^\top)^{-1}(\mathbf{v}_t - \mathbf{A}^{-1}\mathbf{o}_t)\Big]$$

$$= \frac{(2\pi)^{\frac{D'}{2}}|\mathbf{A}^{-1}\Sigma(\mathbf{A}^{-1})^\top|^{\frac{1}{2}}}{(2\pi)^{\frac{D}{2}}|\Sigma_o|^{\frac{1}{2}}} \cdot \frac{1}{(2\pi)^{\frac{D'}{2}}|\mathbf{A}^{-1}\Sigma(\mathbf{A}^{-1})^\top|^{\frac{1}{2}}} \tag{B.1h}$$
$$\exp\Big[-\frac{1}{2}(\mathbf{v}_t - \mathbf{A}^{-1}\mathbf{o}_t)^\top(\mathbf{A}^{-1}\Sigma(\mathbf{A}^{-1})^\top)^{-1}(\mathbf{v}_t - \mathbf{A}^{-1}\mathbf{o}_t)\Big]$$

$$= \frac{(2\pi)^{\frac{D'}{2}}|\mathbf{A}^{-1}\Sigma(\mathbf{A}^{-1})^\top|^{\frac{1}{2}}}{(2\pi)^{\frac{D}{2}}|\Sigma_o|^{\frac{1}{2}}} \mathcal{N}\big(\mathbf{v}_t; \mathbf{A}^{-1}\mathbf{o}_t, \mathbf{A}^{-1}\Sigma(\mathbf{A}^{-1})^\top\big) \tag{B.1i}$$

$$\overset{3}{=} a \cdot \mathcal{N}\big(\mathbf{v}_t; \mathbf{A}^{-1}\mathbf{o}_t, \sigma_o \mathbf{A}^{-1}(\mathbf{A}^{-1})^\top\big) \tag{B.1j}$$

1. GAUSSIANS ARE SYMMETRICAL IN $\mathbf{x}$ AND μ:

$$\mathcal{N}\big(\mathbf{x}; \mu, \Sigma\big) = \mathcal{N}\big(\mu; \mathbf{x}, \Sigma\big)$$

2. INVERSE OF MATRIX PRODUCT:

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$
$$\Rightarrow (\mathbf{A}\Sigma^{-1})\mathbf{A} = (\mathbf{A}^{-1}(\mathbf{A}^\top\Sigma^{-1})^{-1}) = (\mathbf{A}^{-1}\Sigma(\mathbf{A}^{-1})^\top)^{-1}$$

3. Given that $\mathbf{A}$ needs to be squared for $\mathbf{A}^{-1}$ to exist, it must hold that $D = D'$. Thus,

$$\begin{aligned}
\frac{(2\pi)^{\frac{D'}{2}}|\mathbf{A}^{-1}\Sigma(\mathbf{A}^{-1})^\top|^{\frac{1}{2}}}{(2\pi)^{\frac{D}{2}}|\Sigma_o|^{\frac{1}{2}}} &= \frac{|\mathbf{A}^{-1}\Sigma(\mathbf{A}^{-1})^\top|^{\frac{1}{2}}}{|\Sigma_o|^{\frac{1}{2}}} = \frac{|\mathbf{A}^{-1}(\sigma_o^2\mathbb{I})(\mathbf{A}^{-1})^\top|^{\frac{1}{2}}}{|\Sigma_o|^{\frac{1}{2}}} \\
&= \frac{|\sigma_o^2\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top|^{\frac{1}{2}}}{|\sigma_o^2\mathbb{I}|^{\frac{1}{2}}} = \frac{\sigma_o^{2D}|\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top|^{\frac{1}{2}}}{\sigma_o^{2D}|\mathbb{I}|^{\frac{1}{2}}} \\
&= |\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top|^{\frac{1}{2}} \\
&=: a
\end{aligned}$$

B.2 Joint marginal for HHMM

Using the independencies that are defined through the HHMM definitions, the joint marginal for HHMM, $p(h_t, \mathbf{v}_t, \mathbf{o}_{1:t})$, can be derived as follows:

$$p(h_t, \mathbf{v}_t, \mathbf{o}_{1:t}) = \sum_{h_{t-1}} p(h_t, h_{t-1}, \mathbf{v}_t, \mathbf{o}_{1:t}) \tag{B.2a}$$

$$= \sum_{h_{t-1}} p(\mathbf{o}_t|h_t, h_{t-1}, \mathbf{v}_t, \mathbf{o}_{1:t-1}) p(h_t, h_{t-1}, \mathbf{v}_t, \mathbf{o}_{1:t-1}) \tag{B.2b}$$

$$= \sum_{h_{t-1}} p(\mathbf{o}_t|\mathbf{v}_t) p(\mathbf{v}_t|h_t, h_{t-1}, \mathbf{o}_{1:t-1}) p(h_t, h_{t-1}, \mathbf{o}_{1:t-1}) \tag{B.2c}$$

$$= \sum_{h_{t-1}} p(\mathbf{o}_t|\mathbf{v}_t) p(\mathbf{v}_t|h_t) p(h_t|h_{t-1}, \mathbf{o}_{1:t-1}) p(h_{t-1}, \mathbf{o}_{1:t-1}) \tag{B.2d}$$

$$= p(\mathbf{o}_t|\mathbf{v}_t) p(\mathbf{v}_t|h_t) \sum_{h_{t-1}} p(h_t|h_{t-1}, \mathbf{o}_{1:t-1}) \int_{-\infty}^{\infty} p(h_{t-1}, \mathbf{v}_{t-1}, \mathbf{o}_{1:t-1}) d\mathbf{v}_{t-1} \tag{B.2e}$$

B.3 Proof initial marginal

Using results from previous sections, the initialisation for the filtering algorithm in the HHMM can be derived as follows:

$$\alpha(h_1, \mathbf{v}_1) = p(h_1, \mathbf{v}_1, \mathbf{o}_1) \tag{B.3a}$$

$$= p(\mathbf{o}_1|\mathbf{v}_1) p(\mathbf{v}_1|h_1) p(h_1) \tag{B.3b}$$

$$= \mathcal{N}\big(\mathbf{o}_1; \mathbf{A}\mathbf{v}_1, \sigma_o^2 \mathbb{I}\big)\, \mathcal{N}\big(\mathbf{v}_1; f(h_1), \sigma_v^2 \mathbb{I}\big)\, p(h_1) \tag{B.3c}$$

$$\overset{B.1}{=} a \cdot \mathcal{N}\big(\mathbf{v}_1; \mathbf{A}^{-1}\mathbf{o}_1, \sigma_o \mathbf{A}^{-1}(\mathbf{A}^{-1})^\top\big)\, \mathcal{N}\big(\mathbf{v}_1; f(h_1), \sigma_v^2 \mathbb{I}\big)\, p(h_1) \tag{B.3d}$$

$$\overset{1}{=} a \cdot p(h_1)\, \mathcal{N}\big(\mathbf{v}_1; \mu'_{\mathbf{v}_1}(h_1, \mathbf{o}_1), \Sigma'_{\mathbf{v}_1}\big)\, \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_1; f(h_1), \Sigma'_{\mathbf{o}_1}\big) \tag{B.3e}$$

with

$$\Sigma'_{\mathbf{o}_1} := \sigma_{\mathbf{o}}^2 \mathbf{A}^{-1}(\mathbf{A}^{-1})^\top + \sigma_{\mathbf{v}}^2 \mathbb{I} \tag{B.4}$$

and

$$\Sigma'_{\mathbf{v}_1} := \Big(\big(\sigma_o^2 \mathbf{A}^{-1}(\mathbf{A}^{-1})^\top\big)^{-1} + \big(\sigma_v^2 \mathbb{I}\big)^{-1} \Big)^{-1} \tag{B.5}$$

$$\mu'_{\mathbf{v}_1}(h_1) := \Sigma'_{\mathbf{v}_1} \Big(\big(\sigma_o^2 \mathbf{A}^{-1}(\mathbf{A}^{-1})^\top\big)^{-1} \mathbf{A}^{-1}\mathbf{o}_1 + \big(\sigma_v^2 \mathbb{I}\big)^{-1} f(h_1) \Big) \tag{B.6}$$

1. GAUSSIANS ARE CLOSED UNDER MULTIPLICATION

$$\mathcal{N}\big(x; a, A\big)\, \mathcal{N}\big(x; b, B\big) = \mathcal{N}\big(x; c, C\big)\, Z$$

WITH $C = \big(A^{-1} + B^{-1}\big)^{-1}$, $c = C(A^{-1}a + B^{-1}b)$ AND $Z = \mathcal{N}\big(a; b, A + B\big)$

B.4 Proof recursive marginal

Starting from the joint definition derived in B.2, the recursive marginal definition is derived as follows:

$$\alpha(h_t, \mathbf{v}_t) = p(h_t, \mathbf{v}_t, \mathbf{o}_{1:t}) \tag{B.7a}$$

$$= p(\mathbf{o}_t|\mathbf{v}_t)\, p(\mathbf{v}_t|h_t) \sum_{h_{t-1}} p(h_t|h_{t-1}) \int_{-\infty}^{\infty} p(h_{t-1}, \mathbf{v}_{t-1}, \mathbf{o}_{1:t-1}) d\mathbf{v}_{t-1} \tag{B.7b}$$

$$\overset{B.3}{=} \mathcal{N}\big(\mathbf{o}_t; \mathbf{A}\mathbf{v}_t, \sigma_o^2\mathbb{I}\big)\, \mathcal{N}\big(\mathbf{v}_t; f(h_t), \sigma_v^2\mathbb{I}\big) \sum_{h_{t-1}} p(h_t|h_{t-1}) \int_{-\infty}^{\infty} a \cdot p(h_{t-1}) \mathcal{N}\big(\mathbf{v}_{t-1}; \mu'_{\mathbf{v}_{t-1}}(h_{t-1}, \mathbf{o}_{t-1}), \Sigma'_{\mathbf{v}_{t-1}}\big) \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_{t-1}; f(h_{t-1}), \sigma_o^2\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top + \sigma_v^2\mathbb{I}\big)\, d\mathbf{v}_{t-1} \tag{B.7c}$$

$$= \mathcal{N}\big(\mathbf{o}_t; \mathbf{A}\mathbf{v}_t, \sigma_o^2\mathbb{I}\big)\, \mathcal{N}\big(\mathbf{v}_t; f(h_t), \sigma_v^2\mathbb{I}\big) \sum_{h_{t-1}} p(h_t|h_{t-1}) \Big(a\, p(h_{t-1}, \mathbf{o}_{1:t-2}) \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_{t-1}; f(h_{t-1}), \sigma_o^2\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top + \sigma_v^2\mathbb{I}\big) \int_{-\infty}^{\infty} \mathcal{N}\big(\mathbf{v}_{t-1}; \mu'_{\mathbf{v}_{t-1}}(h_{t-1}, \mathbf{o}_{t-1}), \Sigma'_{\mathbf{v}_{t-1}}\big)\, d\mathbf{v}_{t-1} \Big) \tag{B.7d}$$

$$\overset{1}{=} \mathcal{N}\big(\mathbf{o}_t; \mathbf{A}\mathbf{v}_t, \sigma_o^2\mathbb{I}\big)\, \mathcal{N}\big(\mathbf{v}_t; f(h_t), \sigma_v^2\mathbb{I}\big) \Big(a \sum_{h_{t-1}} p(h_t|h_{t-1})\, p(h_{t-1}, \mathbf{o}_{1:t-2}) \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_{t-1}; f(h_{t-1}), \sigma_o^2\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top + \sigma_v^2\mathbb{I}\big) \Big) \tag{B.7e}$$

$$\overset{2}{=} \mathcal{N}\big(\mathbf{o}_t; \mathbf{A}\mathbf{v}_t, \sigma_o^2\mathbb{I}\big)\, \mathcal{N}\big(\mathbf{v}_t; f(h_t), \sigma_v^2\mathbb{I}\big)\, \omega(h_t) \tag{B.7f}$$

$$\overset{B.1}{=} a\, \mathcal{N}\big(\mathbf{v}_t; \mathbf{A}^{-1}\mathbf{o}_1, \sigma_o\, \mathbf{A}^{-1}(\mathbf{A}^{-1})^\top\big)\, \mathcal{N}\big(\mathbf{v}_t; f(h_t), \sigma_v^2\mathbb{I}\big)\, \omega(h_t) \tag{B.7g}$$

$$\overset{3}{=} a\, \omega(h_t)\, \mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t), \Sigma'_{\mathbf{v}_t}\big)\, \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big) \tag{B.7h}$$

with

$$\Sigma'_{\mathbf{o}_t} := \sigma_{\mathbf{o}}^2\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top + \sigma_{\mathbf{v}}^2\mathbb{I} \tag{B.8}$$

and

$$\Sigma'_{\mathbf{v}_t} := \Big(\big(\sigma_o^2\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top\big)^{-1} + \big(\sigma_v^2\mathbb{I}\big)^{-1} \Big)^{-1} \tag{B.9}$$

$$\mu'_{\mathbf{v}_t}(h_t) := \Sigma'_{\mathbf{v}_t} \Big(\big(\sigma_o^2\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top\big)^{-1}\mathbf{A}^{-1}\mathbf{o}_t + \big(\sigma_v^2\mathbb{I}\big)^{-1} f(h_t) \Big) \tag{B.10}$$

1. $\forall$ PROBABILITY DISTRIBUTIONS $p(x) : \int_{\mathcal{R}} p(x)dx = 1$

2. Recursive definition:

$$\omega(h_t) = a \sum_{h_{t-1}} p(h_t|h_{t-1})\, p(h_{t-1}, \mathbf{o}_{1:t-2})\, \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_{t-1}; f(h_{t-1}), \Sigma'_{\mathbf{o}_t}\big)$$

3. GAUSSIANS ARE CLOSED UNDER MULTIPLICATION

$$\mathcal{N}\big(x; a, A\big)\, \mathcal{N}\big(x; b, B\big) = \mathcal{N}\big(x; c, C\big)\, Z$$

WITH $C = \big(A^{-1} + B^{-1}\big)^{-1}$, $c = C(A^{-1}a + B^{-1}b)$ AND $Z = \mathcal{N}\big(a; b, A + B\big)$

B.5 Derivation of normalisation constant $\mathcal{Z}$

$$p(\mathbf{o}_{1:t}) = \sum_{h_t} \int_{-\infty}^{\infty} p(h_t, \mathbf{v}_t, \mathbf{o}_{1:t}) d\mathbf{v}_t \tag{B.11a}$$

$$\overset{B.4}{=} \sum_{h_t} \int_{-\infty}^{\infty} a\, p(h_t, \mathbf{o}_{1:t-1})\, \mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t, \mathbf{o}_t), \Sigma'_{\mathbf{v}_t}\big)\, \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big)\, d\mathbf{v}_t \tag{B.11b}$$

$$= \sum_{h_t} a\, p(h_t, \mathbf{o}_{1:t-1})\, \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big) \int_{-\infty}^{\infty} \mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t, \mathbf{o}_t), \Sigma'_{\mathbf{v}_t}\big)\, d\mathbf{v}_t \tag{B.11c}$$

$$= a \sum_{h_t} p(h_t, \mathbf{o}_{1:t-1})\, \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big) \tag{B.11d}$$

$$=: a\, \mathcal{Z} \tag{B.11e}$$

B.6 Derivation of posterior distributions from filtering results

Using

$$p(\mathbf{o}_{1:t}) = \sum_{h_t} \int_{-\infty}^{\infty} p(h_t, \mathbf{v}_t, \mathbf{o}_{1:t}) d\mathbf{v}_t = a\, \mathcal{Z} \tag{B.12}$$

the following posterior distributions are derived:

1. Joint posterior over latents:

$$p(h_t, \mathbf{v}_t | \mathbf{o}_{1:t}) = \frac{p(h_t, \mathbf{v}_t, \mathbf{o}_{1:t})}{p(\mathbf{o}_{1:t})} = \frac{p(h_t, \mathbf{v}_t, \mathbf{o}_{1:t})}{\sum_{h_t} \int_{-\infty}^{\infty} p(h_t, \mathbf{v}_t, \mathbf{o}_{1:t}) d\mathbf{v}_t} \tag{B.13a}$$

$$= \frac{1}{\mathcal{Z}}\, \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big)\, \mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t, \mathbf{o}_t), \Sigma'_{\mathbf{v}_t}\big)\, \omega(h_t) \tag{B.13b}$$

2. Spatial posterior:

$$p(h_t | \mathbf{o}_{1:t}) = \int_{-\infty}^{\infty} p(h_t, \mathbf{v}_t | \mathbf{o}_{1:t}) d\mathbf{v}_t = \frac{\int_{-\infty}^{\infty} p(h_t, \mathbf{v}_t, \mathbf{o}_{1:t}) d\mathbf{v}_t}{\sum_{h_t} \int_{-\infty}^{\infty} p(h_t, \mathbf{v}_t, \mathbf{o}_{1:t}) d\mathbf{v}_t} \tag{B.14a}$$

$$= \frac{1}{\mathcal{Z}} \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big)\, \omega(h_t) \tag{B.14b}$$

3. Visual posterior:

$$p(\mathbf{v}_t | \mathbf{o}_{1:t}) = \sum_{h_t} p(h_t, \mathbf{v}_t | \mathbf{o}_{1:t}) = \frac{\sum_{h_t} p(h_t, \mathbf{v}_t, \mathbf{o}_{1:t})}{\sum_{h_t} \int_{-\infty}^{\infty} p(h_t, \mathbf{v}_t, \mathbf{o}_{1:t}) d\mathbf{v}_t} \tag{B.15a}$$

$$= \frac{1}{\mathcal{Z}} \sum_{h_t} \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big)\, \mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t), \Sigma'_{\mathbf{v}_t}\big)\, \omega(h_t) \tag{B.15b}$$

4. Spatial predictive posterior:

$$p(h_{t+1}|\mathbf{o}_{1:t}) = \sum_{h_t} p(h_{t+1}|h_t)\, p(h_t|\mathbf{o}_{1:t}) \tag{B.16a}$$

$$= \frac{1}{\mathcal{Z}} \sum_{h_t} \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big)\, p(h_{t+1}|h_t)\, \omega(h_t) \tag{B.16b}$$

5. Visual predictive posterior:

$$p(\mathbf{v}_{t+1}|\mathbf{o}_{1:t}) = \sum_{h_t, h_{t+1}} p(\mathbf{v}_{t+1}|h_{t+1}) p(h_{t+1}|h_t) p(h_t|\mathbf{o}_{1:t}) \tag{B.17a}$$

$$= \frac{1}{\mathcal{Z}} \sum_{h_t, h_{t+1}} \mathcal{N}\big(\mathbf{v}_{t+1}; f(h_{t+1}), \sigma_v^2 \mathbb{I}\big)\, p(h_{t+1}|h_t)\, \omega(h_t) \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \mathbb{I}\big) \tag{B.17b}$$

B.7 Derivation of DDC encoding for $p(\mathbf{v}_t|\mathbf{o}_{1:t})$

The following derivation follows the argumentation used in Section A.1. For detailed explanations of the steps, please refer to that section.

$$r_i(\mathbf{v}_t) = \int_{-\infty}^{\infty} \phi_i(\mathbf{v}_t) p(\mathbf{v}_t|\mathbf{o}_{1:t})\, d\mathbf{v}_t \tag{B.18a}$$

$$= \int_{-\infty}^{\infty} \phi_i(\mathbf{v}_t) \left[\frac{1}{\mathcal{Z}} \sum_{h_t} \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \sigma_o^2 \mathbf{A}^{-1}(\mathbf{A}^{-1})^\top + \sigma_v^2 \mathbb{I}\big) \mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t, \mathbf{o}_t), \Sigma'_{\mathbf{v}_t}\big)\, p(h_t, \mathbf{o}_{1:t-1}) \right] d\mathbf{v}_t \tag{B.18b}$$

$$= \frac{1}{\mathcal{Z}} \int_{-\infty}^{\infty} \sum_{h_t} \left[\phi_i(\mathbf{v}_t)\, \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \sigma_o^2 \mathbf{A}^{-1}(\mathbf{A}^{-1})^\top + \sigma_v^2 \mathbb{I}\big) \mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t, \mathbf{o}_t), \Sigma'_{\mathbf{v}_t}\big)\, p(h_t, \mathbf{o}_{1:t-1}) \right] d\mathbf{v}_t \tag{B.18c}$$

$$= \frac{1}{\mathcal{Z}} \sum_{h_t} \left[\int_{-\infty}^{\infty} \phi_i(\mathbf{v}_t)\, \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \sigma_o^2 \mathbf{A}^{-1}(\mathbf{A}^{-1})^\top + \sigma_v^2 \mathbb{I}\big) \mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t, \mathbf{o}_t), \Sigma'_{\mathbf{v}_t}\big)\, p(h_t, \mathbf{o}_{1:t-1})\, d\mathbf{v}_t \right] \tag{B.18d}$$

$$= \frac{1}{\mathcal{Z}} \sum_{h_t} \left[\mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \sigma_o^2 \mathbf{A}^{-1}(\mathbf{A}^{-1})^\top + \sigma_v^2 \mathbb{I}\big)\, p(h_t, \mathbf{o}_{1:t-1}) \int_{-\infty}^{\infty} \phi_i(\mathbf{v}_t)\, \mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t, \mathbf{o}_t), \Sigma'_{\mathbf{v}_t}\big)\, d\mathbf{v}_t \right] \tag{B.18e}$$

$$\overset{(4.4)}{=} \frac{1}{Z}\sum_{h_t}\Bigg[\mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \sigma_o^2\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top + \sigma_v^2\mathbb{I}\big)\, p(h_t, \mathbf{o}_{1:t-1}) \int_{-\infty}^{\infty}\left(\rho_{i0} + \sum_{k=1}^{K_i}\rho_{ik}\,\psi_{ik}(\mathbf{v}_t)\right)\mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t,\mathbf{o}_t), \Sigma'_{\mathbf{v}_t}\big)\, d\mathbf{v}_t\Bigg] \tag{B.18f}$$

$$= \frac{1}{Z}\sum_{h_t}\Bigg[\mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \sigma_o^2\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top + \sigma_v^2\mathbb{I}\big)\, p(h_t, \mathbf{o}_{1:t-1}) \left(\int_{-\infty}^{\infty}\rho_{i0}\,\mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t,\mathbf{o}_t), \Sigma'_{\mathbf{v}_t}\big)\, d\mathbf{v}_t + \int_{-\infty}^{\infty}\sum_{k=1}^{K_i}\rho_{ik}\,\psi_{ik}(\mathbf{v}_t)\,\mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t,\mathbf{o}_t), \Sigma'_{\mathbf{v}_t}\big)\, d\mathbf{v}_t\right)\Bigg] \tag{B.18g}$$

$$= \frac{1}{Z}\sum_{h_t}\Bigg[\mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \sigma_o^2\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top + \sigma_v^2\mathbb{I}\big)\, p(h_t, \mathbf{o}_{1:t-1}) \left(\rho_{i0}\int_{-\infty}^{\infty}\mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t,\mathbf{o}_t), \Sigma'_{\mathbf{v}_t}\big)\, d\mathbf{v}_t + \sum_{k=1}^{K_i}\int_{-\infty}^{\infty}\rho_{ik}\,\psi_{ik}(\mathbf{v}_t)\,\mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t,\mathbf{o}_t), \Sigma'_{\mathbf{v}_t}\big)\, d\mathbf{v}_t\right)\Bigg] \tag{B.18h}$$

$$\overset{(B.19)}{=} \frac{1}{Z}\sum_{h_t}\Bigg[\mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \sigma_o^2\mathbf{A}^{-1}(\mathbf{A}^{-1})^\top + \sigma_v^2\mathbb{I}\big)\, p(h_t, \mathbf{o}_{1:t-1}) \left(\rho_{i0} + \sum_{k=1}^{K_i}\frac{\rho_{ik}\sqrt{|\Gamma_{ik}|}}{\sqrt{|\Gamma_{ik}+\Sigma'_{\mathbf{v}_t}|}} \exp\Big(-\frac{1}{2}\big(\mu_{ik}-\mu'_{\mathbf{v}_t}(h_t,\mathbf{o}_t)\big)^\top\big(\Gamma_{ik}+\Sigma'_{\mathbf{v}_t}\big)^{-1}\big(\mu_{ik}-\mu'_{\mathbf{v}_t}(h_t,\mathbf{o}_t)\big)\Big)\right)\Bigg] \tag{B.18i}$$

1. Using the results derived in Section A.1, the following holds:

$$\int_{-\infty}^{\infty}\rho_{ik}\,\psi_{ik}(\mathbf{v}_t)\,\mathcal{N}\big(\mathbf{v}_t; \mu'_{\mathbf{v}_t}(h_t,\mathbf{o}_t), \Sigma'_{\mathbf{v}_t}\big)\, d\mathbf{v}_t \overset{(A.2h)}{=} \frac{\rho_{ik}\sqrt{|\Gamma_{ik}|}}{\sqrt{|\Gamma_{ik}+\Sigma'_{\mathbf{v}_t}|}} \exp\Big(-\frac{1}{2}\big(\mu_{ik}-\mu'_{\mathbf{v}_t}(h_t,\mathbf{o}_t)\big)^\top\big(\Gamma_{ik}+\Sigma'_{\mathbf{v}_t}\big)^{-1}\big(\mu_{ik}-\mu'_{\mathbf{v}_t}(h_t,\mathbf{o}_t)\big)\Big) \tag{B.19}$$

B.8 DDC equations for HHMM

The following equations give the DDC encoding for the HHMM:

1. Spatial posterior:

$$r_i(p(h_t|\mathbf{o}_{1:t})) = \sum_{h_t} \phi_i(h_t)\, p(h_t|\mathbf{o}_{1:t}) = \frac{1}{\mathcal{Z}} \sum_{h_t} \phi_i(h_t) \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big)\, \omega(h_t) \tag{B.20}$$

2. Spatial predictive posterior:

$$\begin{aligned} r_i(p(h_{t+1}|\mathbf{o}_{1:t})) &= \sum_{h_{t+1}} \phi_i(h_{t+1})\, p(h_{t+1}|\mathbf{o}_{1:t}) \\ &= \frac{1}{\mathcal{Z}} \sum_{h_t, h_{t+1}} \phi_i(h_{t+1}) \mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big)\, p(h_{t+1}|h_t)\, \omega(h_t) \end{aligned} \tag{B.21}$$

3. Visual posterior:

$$\begin{aligned} r_i(p(\mathbf{v}_t|\mathbf{o}_{1:t})) &= \int_{-\infty}^{\infty} \phi_i(\mathbf{v}_t) p(\mathbf{v}_t|\mathbf{o}_{1:t})\, d\mathbf{v}_t \\ &\overset{B.7}{=} \frac{1}{\mathcal{Z}} \sum_{h_t} \Bigg[\mathcal{N}\big(\mathbf{A}^{-1}\mathbf{o}_t; f(h_t), \Sigma'_{\mathbf{o}_t}\big)\, \omega(h_t) \\ &\qquad \left(\rho_{i0} + \sum_{k=1}^{K_i} \rho_{ik} \sqrt{(2\pi)^D |\Gamma_{ik}|}\, \mathcal{N}\Big(\mu_{ik}; \mu'_{\mathbf{v}_t}(h_t), \Gamma_{ik} + \Sigma'_{\mathbf{v}}\Big) \right) \Bigg] \end{aligned} \tag{B.22}$$

4. Visual predictive posterior:

$$\begin{aligned} r_i(p(\mathbf{v}_{t+1}|\mathbf{o}_{1:t})) &= \int_{-\infty}^{\infty} \phi_i(\mathbf{v}_{t+1}) p(\mathbf{v}_{t+1}|\mathbf{o}_{1:t})\, d\mathbf{v}_{t+1} \\ &= \frac{1}{\mathcal{Z}} \sum_{h_{t+1}} \Bigg[p(h_{t+1}|\mathbf{o}_{1:t}) \Bigg(\rho_{i0} + \\ &\qquad \sum_{k=1}^{K_i} \rho_{ik} \sqrt{(2\pi)^D |\Gamma_{ik}|}\, \mathcal{N}\Big(\mu_{ik}; f(h_{t+1}), \Gamma_{ik} + \sigma^2_{\mathbf{v}} \mathbb{I}\Big) \Bigg) \Bigg] \end{aligned} \tag{B.23}$$

Appendix C

Supplementary figures

This appendix presents figures that are not directly related to the findings of the thesis but important for the presented arguments.

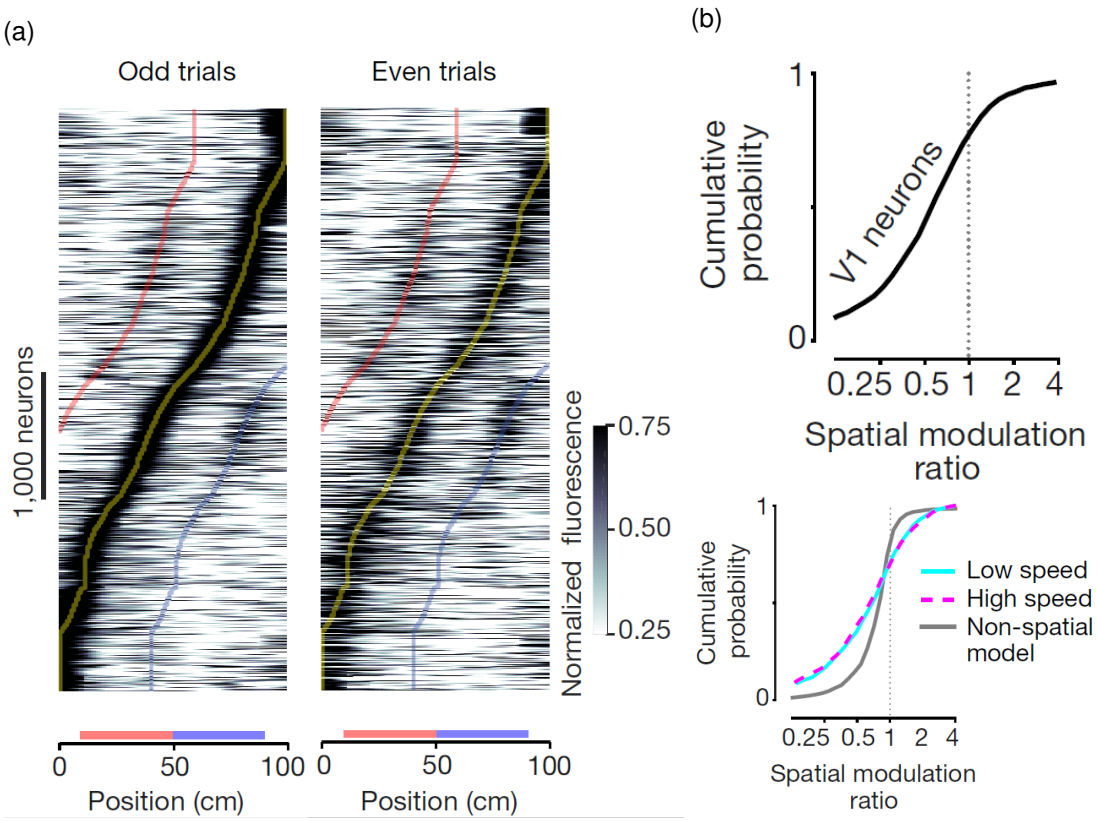

Figure C.1: Figures 1e, 1f, 1g, and 1h taken from Saleem et al. (2018). (a) V1-neurons sorted by peak position. (b) Spatial modulation ratio averaged over all trials and neurons (top) and only low and high speed trials compared to a purely visual receptive-field-based model (bottom).

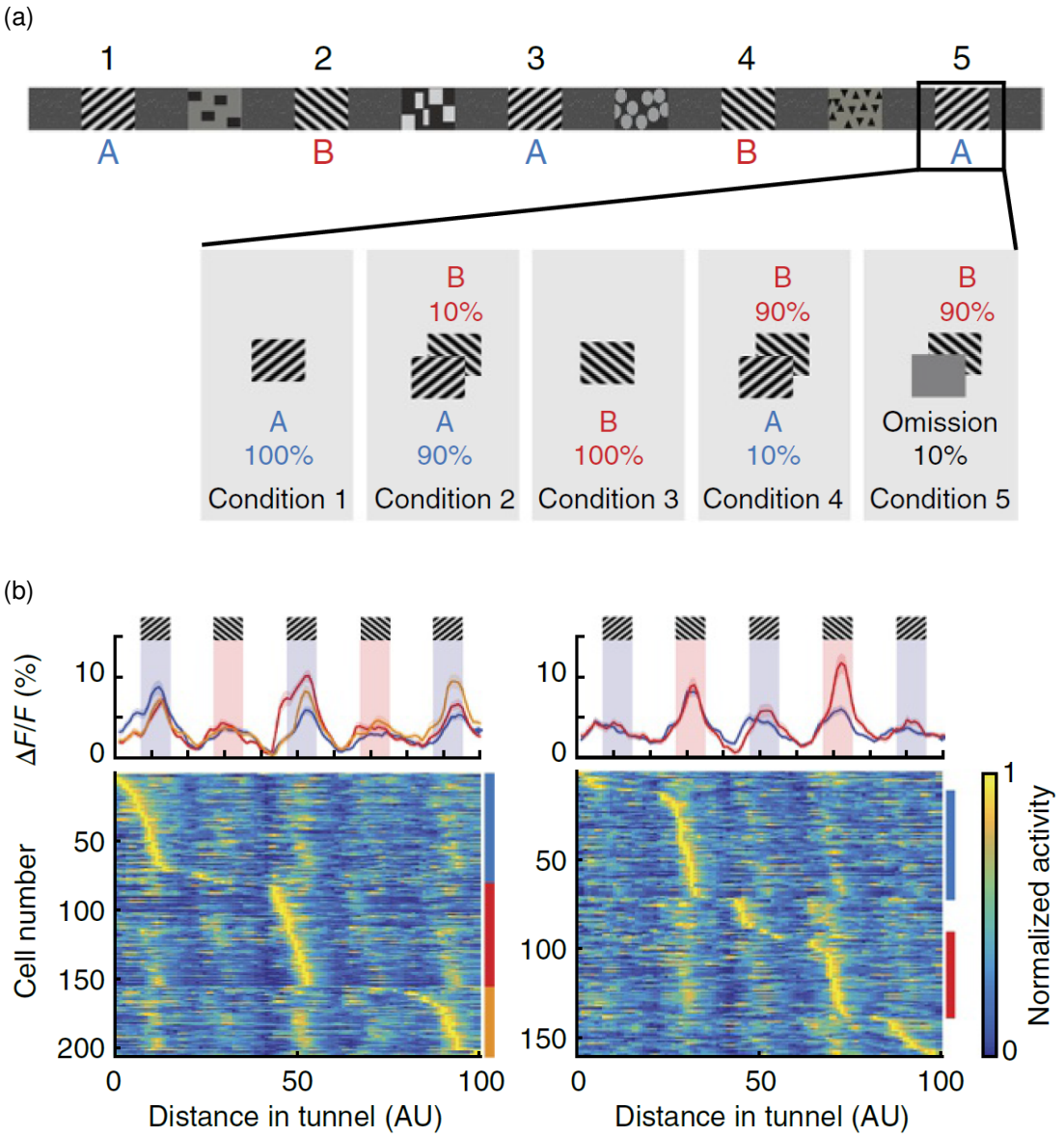

Figure C.2: Figures 1b and 1g taken from Fiser et al. (2016) (a) Corridor setup. (b) A-selective neurons (left) and B-selective neurons (right) in the hippocampus (CA1) sorted by peak position.

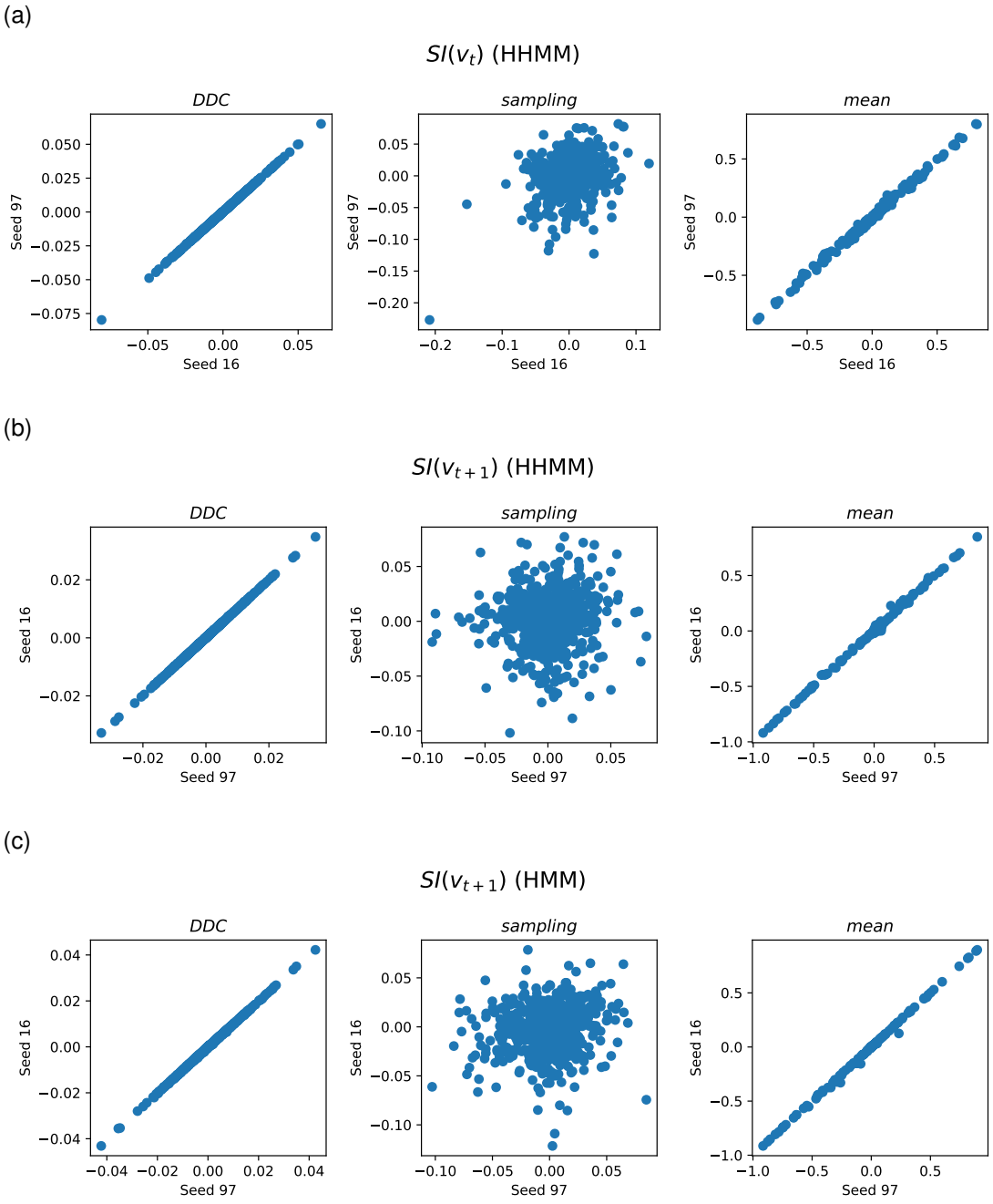

Figure C.3: Testing the number of trials for estimating the SI. 50 trials are sufficient for the DDC encoding and close to sufficient for the mean encoding as the SI values are aligned well to the diagonal. However, the SI values for the sampling conditions are not reliable for 50 trials.