Applying Tensor Decomposition Methods to Neural Population Recordings

Tadsawan Tansuwan

Master of Science
Artificial Intelligence
School of Informatics
University of Edinburgh
2021

Abstract

Advancement in the neuroscience field enables us to acquire large-scale neuronal activity data and comprise them into a tensor. A recent study has implemented tensor component analysis (TCA) to digest complex datasets in an unsupervised manner into a set of three low-dimensional descriptions with neurological aspects: neuron, temporal and trial factors. This dissertation analyzes the attention switching dataset by using TCA to extract the trial-to-trial variability and inspect the underlying neural dynamics. We aim to explore how the neuronal activity changes as the mice switch attention within the session and how improvements in sensory behavior are associated with the changes in neural circuits. Firstly, we implemented a pipeline for data preprocessing and evaluating the three factors from standard and non-negative TCA. Next, we demonstrate the selectivity metrics and refitting procedure we introduced in this dissertation while recovering a neural dynamic in the trial factor of the TCA model that associate with attention switching between stimuli in two mice. Finally, we discuss TCA's potential in extracting task-related variables and propose future directions on investigating attention switching data and TCA.

Acknowledgements

I would like to thank my advisor, Angus Chadwick, for his valuable guidance throughout this project. Lastly, I would like to thank my family and friends that support me throughout this rough year.

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.

(*Tadsawan Tansuwan*)

Table of Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Objective 2

1.3 Overview 2

2 Background **3**

2.1 Attention switching dataset 3

2.2 TCA 5

3 Methodology **8**

3.1 Tensor decomposition 8

3.2 Data preprocessing 10

3.2.1 Outlier detection 10

3.2.2 Normalization 11

3.3 Evaluation 11

3.3.1 Visual interpretability 12

3.3.2 Selectivity metrics 13

3.3.3 Performance metrics 14

3.4 Refitting 16

4 Results and Discussion **17**

4.1 Exploring TCA configuration 17

4.2 Underlying neural dynamic of attention switching 23

4.3 Robustness of TCA 27

4.4 Separation in task selectivity and additional component 29

5 Conclusion **31**

5.1 Future work 32

Bibliography **34**

A Additional Plots **37**

A.1 Additional evaluation metric 37

A.2 Additional components of TCA models 40

Chapter 1

Introduction

1.1 Motivation

We constantly switch our attention toward different tasks in everyday life without thinking about how we switches attention naturally. The neuronal changes that occurred within milliseconds as we switch our attention are never truly seen, as it is unknown how the neuronal activity changes occur when the attention switches. With the advancement in technology within the neuroscience field, it enables us to document different parts of the brain activity by recording a large number of neurons across multiple trials [4]. However, the complexity of the neural recording imposed a challenge in digesting such large-scale data into an interpretable result that can assist in identifying the task variable in the neural dynamics in our brain.

The dimensionality reduction method has been introduced to digest highly complex neural data in an unsupervised manner into low-dimensional descriptions while being interpretable in neurological aspects. Tensor component analysis (TCA) had previously shown to successfully extract low-dimension descriptions of mice's prefrontal cortex during navigation task and the motor cortex during the brain-machine interface learning task while retaining neurological variability that reflected rapid circuit, the time within a trial, and across trials changes [19]. The method allows the researcher to perform dimensionality reduction while maintaining the original dimension of the data, where the data is in a three-dimensional array, also known as a tensor, with a dimension of amount of neurons, neuronal recording at each time points within a trial, and the number of trials. Nevertheless, TCA has never been explored in the attention switching task dataset before. This dissertation primarily explores how the neuronal activity changes as the mice switch attention within the session and how improvements

in sensory behavior are linked to change in neural circuits by observing the TCA models.

1.2 Objective

This project aims to recover the neural dynamics of attention switching tasks that unfold across trials and time within trials using TCA. The hypothesis is based on [19] that suggested TCA can decompose acquired firing-rate neural recording data of neural activity recorded at each time point within multiple trials to three low-dimensional descriptions of each dimension of the data. While there are multiple interesting neurological aspects in the attention switching task dataset, for this dissertation, we focus on finding the neural dynamics of attention switching across the session of trials. Thus, the objective that we aim to accomplished are the following objectives:

- Find the underlying neural dynamic of attention switching in unsupervised manner by using TCA.
- Analyze TCA models interpretability both quantitatively and qualitatively.
- Identify a pattern recovered by TCA with possible neurological signature and attempt to associate with a task variable.

1.3 Overview

The dissertation is structured in three main chapters. Chapter 2 gives an overview of the attention switching data and describing the tensor component analysis procedure. Chapter 3 outlines the pipeline for processing the data and performing TCA. The pipeline includes normalization, outlier detection to retain only relevant cells, different tensor decomposition methods, and evaluation for TCA models. The TCA models from Chapter 3 are then analyzed based on the selectivity metrics for the trial factors, performance metrics, and mainly the visual interpretation. The results aiming for the objective are uncovered in Chapter 4, while discussing the potential of selectivity metrics and neurological signatures recover by TCA. Finally, Chapter 5 discusses the conclusion of the experiment with TCA on attention switching data and potential future work. Additional material can be found in Appendix, which includes the components from TCA models and performance metrics from experimenting with TCA.

Chapter 2

Background

2.1 Attention switching dataset

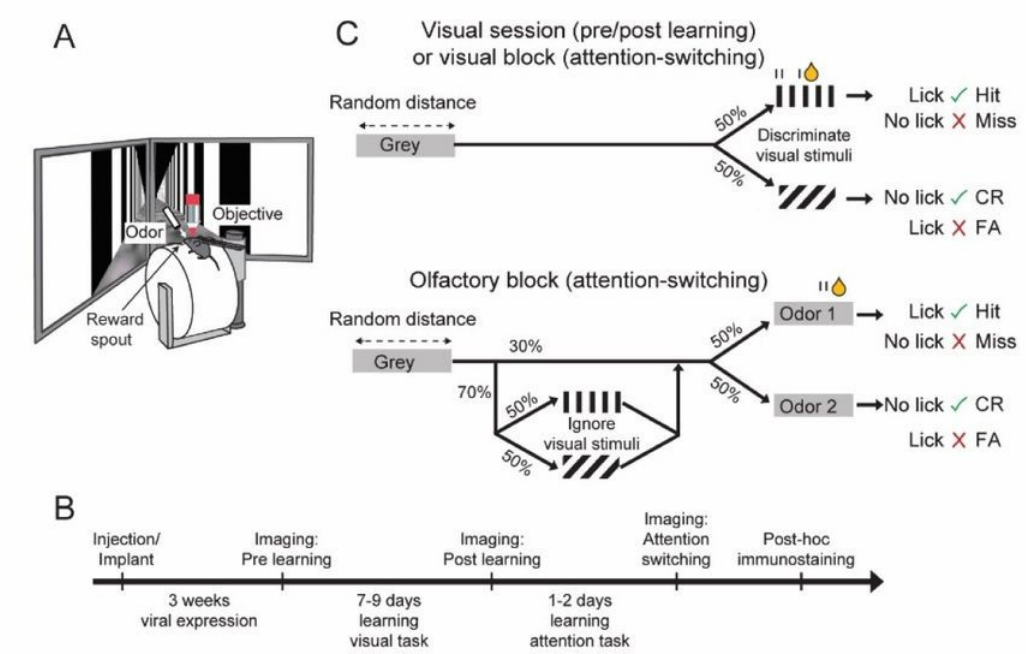

Figure 2.1: (A) Experimental setup for mice with virtual reality and odor. (B) Experimental timeline of the mice. (C) The behavioral task for the mice. Top, visual discrimination: Mice were rewarded for when vertical gratings were presented and not when angled gratings were presented. Olfactory discrimination: mice were rewarded for licking when odor 1 was presented and not when odor 2 while vertical or angled gratings were presented. [13]

Attention selectively enhances the processing of relevant stimuli, while suppress

the responses to non-preferred stimuli [12]. Early sensory cortical representation often changes when animals focus their attention on task-relevant features. It is currently unknown how the cortical changes actually occur when the attention switches, as mentioned in Section 1.1. In order to study the occurrence of attention switching, researchers recorded the calcium indicator of the population of the neurons in the primary visual cortex (V1) to monitor the cell activity in the subset of neurons as the mice learnt to ignore the two visual stimuli and performed an attention switching task to responds to the olfactory stimuli [14].

They had tracked excitatory cells: pyramidal excitatory pyramidal. The three inhibitory interneurons alongs the local excitatory neuron population are monitored as well, which included:(PYR) neurons, parvalbumin (PV), somatostatin (SOM), and vasoactive intestinal peptide (VIP) positive interneurons. To answer how the cells interact with each other during attention switching task, they had monitored the interaction between the inhibitory interneurons along with the excitatory cells, as the mice performed the experiment. The amount of recorded cells of each type is shown in Table 2.1.

Cell type	Attention
PYR	5813
PV	477
SOM	245
VIP	365

Table 2.1: Population of each type of cells capture in attention dataset. [14]

The mice underwent behavioral tasks, where they were first exposed to visual sessions to learn to discriminate the visual stimuli (Figure 2.1C top). All mice were motivated by thirstiness and hunger as they were on food restriction before the experiment. They were rewarded if they responded to the vertical stimulus and no rewarded if exposed to angled stimulus. Later, the mice were exposed to visual stimuli and followed by the olfactory stimuli (Figure 2.1C bottom). However, the mice were not rewarded for the visual stimuli but were rewarded based on the olfactory stimuli instead during the olfactory session. Therefore, the mice were expected to ignore the visual stimuli and switch their attention to the olfactory stimuli for their reward.

The task switches across the trials the mice participated in. Hence, the mice's attention was assumed to be switching throughout the session, and the selectivity of the cells would increase their activity as they encountered the rewarded stimuli and sup-

pressed as non-rewarded stimuli were encountered. The experiment was conducted on multiple mice. However the dataset we would be exploring contains six different mice and nine sessions in total. Each session was considered an individual set of data due to the inconsistency in the number of cells and trials recording. The dataset is structured in a three-dimensional array, a tensor, where the dimensions correspond to neuron N by time points within each trial T by experimental trials K. Tensor component analysis (TCA) would be suitable for extracting neurological descriptions of the dataset and analyzing the possible underlying to attention switching through the across-trial variability that TCA would capture.

2.2 TCA

The advancement in neuronal recording allows us to document large number of neurons at higher precision over a long period of time [11, 10], which results neural recording to contain multiple dynamics of neurons over various of behavioral trials. This imposed a challenge of analyzing and interpret the dynamic within the neural recording. Dimensionality reduction method had been used commonly in neuroscience to extract low-dimensional descriptions of neural recording. The standard of dimensionality reduction method would be principle component analysis (PCA). While, PCA can decompose highly complex data to its low-dimensionality representation in an unsupervised manner. It does have disadvantages regarding decomposing the common structure of the neuronal recording and recovering the across-trial variable. Commonly, a recording of N neurons over K experimental trials with T time points within each trials is represented in $N \times T \times K$ third-order tensor. It requires preprocessing to be done to the dimensionality of the tensor to be reduced to a second-order matrix in order to fit into PCA. One approach can be to average multiple trials and fit into PCA to decompose to R dimensions [5], thereby reducing the data tensor into $R(N+T)$ summary of trial-averaged data. Averaging the trials would dismantle the trial-to-trial dynamic and treat trial variability as irrelevant noise. However, variability can reveal important neural dynamics, such as switching in attention through the trials, which we want to explore using the dataset described in Section 2.1.

Furthermore, PCA do not guarantee uniqueness [3] as it restrains on an assumption that the factors are orthogonal. There is infinity amount of low-dimension factors with a transformation matrix to create the same reconstruction of the data. Hence, it increases the vulnerability the result from the PCA to be incorrectly describing the

data. Tensor decomposition analysis (TCA) [19] is introduced as the alternative dimensionality reduction method as it can decompose the neuronal data while including the trial-to-trial dynamic within the data in the low-dimensionality descriptions, and avoid the need for a matrix conversion, and being constrain in orthogonality to yield unique solution.

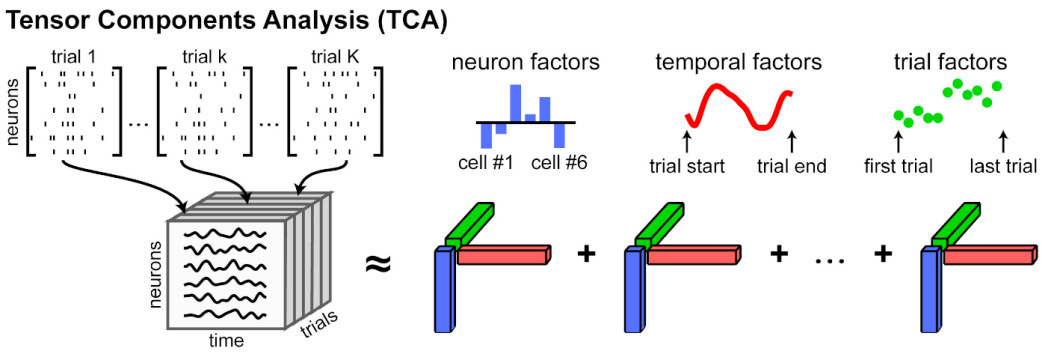

Figure 2.2: Schematic of TCA. The data is a third-order tensor with dimensions $N \times T \times K$. TCA approximates the data as a sum of outer products of three vectors which produces the corresponding factors (neuron, temporal and trial factors) describing the changes across the data. [19]

For this version of TCA, the process involves the CANDECOMP/PARAFAC (CP) decomposition [1] to identify a set of low-dimensional components which describes variability along the three axes. The CP decomposition factorizes a tensor X into a sum of outer products of vectors. Hence, a tensor decomposes to a sum of R rank one tensors as:

$$X \approx \sum_{r=1}^{R} w^r \circ b^r \circ a^r \tag{2.1}$$

where ($\circ$) represents the outer product and w^r, b^r, a^r are vectors of each axes' dimension. Applying TCA on the attention switching dataset from Section 2.1 results biologically representation of each vectors, as follows:

- Neuron factor w^r represents an assembly of cells with rapid.
- Temporal factor b^r denotes on how the recorded cell ensembles are working before and after onset stimuli in every single trial.
- Trial factor a^r represents slower across-trial dynamic as the mice switched between task conditions.

Figure 2.2 illustrates the decomposition, where the neuron factors w^r are represented in blue, temporal factors b^r are in red, and trial factors a^r are in green.

Unlike PCA, TCA is able to embed the trial-to-trial variability into R components' trial factors, while PCA cannot capture due to the data restriction. Additionally, TCA is unique up to rescalings and permutations when factors are all linearly independent (W, B, A have full column rank) [9]. Thus, the condition for uniqueness is much weaker than PCA, which increases the credibility of the decomposed factors. With this, applying TCA to the attention switching dataset can achieve a decomposition in which captures the across-trials attention switches and within-trials temporal changes while yielding a unique solution without restriction and can be applied directly to the data in an unsupervised manner.

Chapter 3

Methodology

This chapter describes the pipeline developed for preprocessing, extracting low-dimensional description and evaluating the attention switching data, introduced in Section 2.1. The pipeline consists of following steps: outlier detection, normalization, applying TCA, evaluation, and refitting. These processes were applied onto the attention switching dataset, introduced in Section 2.1. The structure of the attention switching dataset was in a three-way array, known as a tensor, where the dimensions are neurons N by time points within-trials of the neuron T by trials K. The data contained two stimuli conditions, which were as follows: vertical and angled visual grating. The task that the mice were assigned were visual and olfactory tasks. For simplicity, we had concatenated the data into a sequence, as follows: vertical grating in visual block, vertical grating in olfactory block, angled grating in visual block, and angled grating in olfactory block.

3.1 Tensor decomposition

In [19], they had introduced a tensor decomposition method, CANDECOMP/PARAFAC (CP) decomposition [1], to apply on the third-order tensor data of neural recording $N \times T \times K$, where the data was a recording of N neurons over K experimental trials and the neural activity was recorded at T time points within each trial. The decomposition is the process in TCA, as introduced in Section 2.2. The process was carried out in Tensortools [19] and code was writting in Python 3 [18]. We explored the two different CP decomposition method: standard CP decomposition and non-negative CP decomposition. The goal for both method were to compute a CP decomposition with

R components that best approximates X [8], which is defined as:

$$\underset{\hat{X}}{\text{minimize}} \quad \|X - \hat{X}\|_F^2, \quad \hat{X} = \sum_{r=1}^{R} w^r \circ b^r \circ a^r \tag{3.1}$$

where X denotes the data and $\widehat{X}$ denotes the tensor approximated by the TCA, and the error is computed by the squared Euclidean norm ($\|\cdot\|_F^2$) of between the two tensors. Fitting TCA to data is a nonconvex problem, which means that each run does not guarantee to reach global minima. Hence, TCA can be only be fitted by an iteration optimization algorithm. Although it may be stuck in the suboptimal local minima, most optimization fits converged to similar reconstruction errors [19]. Alternative least-square (ALS) [8] is used as the iteration optimization algorithm for both of the CP decomposition methods. The goal is to compute a CP decomposition with lowest error in Equation 3.1.

$$W \longleftarrow \underset{\tilde{w}}{\text{argmin}} \sum_{ntk} (X_{ntk} - \sum_{r} \tilde{w}_n^r b_t^r a_k^r)^2 \tag{3.2}$$

The ALS approach reduces the problem to a linear least-square problem as seen in Equation 3.2. The update rule is optimizing the neuron factors W, while temporarily fixing the within-trial factors B and trial factors A. The procedure continues by repetitively optimizing one factor and fixed the rest of the factors. Since each iteration of ALS can only fitted with R component at a time, the experiment tested component number ranging from 1 to 20. This is done until the objective function (Equation 3.1) converges. Furthermore, the two CP decomposition methods are fitted into different version of the data, introduced in Section 3.2.2, as shown in Section 3.2.

Standard CP decomposition Standard CP decomposition includes the negativity in the fitting. The original data with no normalization, z-score and rescaled by minimum and maxiumum normalized version of the data are fitted into standard CP decomposition.

Non-negative CP decomposition This version of standard CP decomposition is only constrained to positive orthant during the optimization. Hence, the objective function in Equation 3.1 adapts the non-negative constraint on the factors (W, B, A). It has been proven that non-negative CP decomposition (NCP) has stronger uniqueness conditions when compared to the standard CP decomposition [15], which reduces the chance of CP degeneracies occurring when fitting the data [2]. The data rescaled by minimum and maximum values is fitted into the non-negative CP decomposition.

3.2 Data preprocessing

Outlier detection and normalization were carried out in the data preprocessing section of the pipeline. The process was carried out in NumPy[6], and code was written in Python 3 [18]. The aim of applying data preprocessing is to eliminate any irrelevant noise within the data and accommodate the data to be applied in different versions of tensor decomposition methods.

3.2.1 Outlier detection

Algorithm 1 Signal-to-noise ratio outlier process

Require: neuronal recording tensor in $N \times T \times K$ dimension

$x \leftarrow$ neuronal recording tensor

for each cells i in x's N dimension **do**

  $mean_{condition} \leftarrow [\,]$

  for each condition j in stimuli and task conditions set **do**

    $value \leftarrow$ extracts i activity at all time points at condition j

    for m each time points after and before stimuli **do**

      $time_m \leftarrow$ computes the mean of $value$ at each time points m

    end for

    $mean_condition$ records $time_m$

  end for

  $signal_{var} \leftarrow$ compute the variance of the means of each conditions (recorded in $mean_{condition}$

  $noise_{var} \leftarrow$ compute the variance of overall cell i activity in all conditions

  $snr \leftarrow \frac{signal_{var}}{noise_{var}}$

end for

Outlier were detected and excluded from the data using signal-to-noise ratio (SNR), as seen in Equation 3.3. SNR is often used in neuroscience community for discriminating relevant signals from noise [16]. For this project, SNR is used to discriminate which cells are relevant or not. The process involves calculating SNR on each cell and at certain threshold, lower SNR cells are excluded. We assumed that the data might contain irrelevant signals which we presumed as an error from the recording or signals not associated with the attention switching task. Hence, the lower SNR described cells with low fraction of their total variance correlated with task variables, which should be

excluded. The SNR is defined as:

$$snr = \frac{\text{signal}}{\text{noise}} \tag{3.3}$$

where the signal is the variance of the means of a cell activity at each time points in each stimuli condition, and the noise is the total variance of a cell activity in all stimuli conditions. After computing the SNR of each cells, we sorted the cells by its SNR and removed at certain threshold from lowest to highest. The threshold ranged from 5% to 30%. The selected cells' recordings were excluded from the dataset. Algorithm 1 shows how the outlier detection method operates and use SNR to remove cells with low SNR.

3.2.2 Normalization

Normalization is used to reduce variability within the data and allows non-negative CP decomposition to be explored. The data normalization methods are min-max and z-score normalization. Both approaches are suitable for this data since the minimum and maximum values were specific to each dataset, and z-score only consists of scaling the data to the mean and standard deviation of 1.

For each set of data, the minimum and maximum values of the whole recording session were used to rescale each individual cell activity in the data to be between zero and one. The rescaling method is defined as:

$$rescale = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{3.4}$$

where x denotes an individual value in the data and $x_{min/max}$ indicates the minimum and maximum value of the whole data. Similarly, z-score normalization used the mean and standard deviation of the whole data to compute the z-score of each individual value. The z-score method is defined as:

$$zscore = \frac{x - \bar{x}}{\sigma_x} \tag{3.5}$$

where x denotes an individual value in the data and $\bar{x}, \sigma_x$ denote the mean and standard deviation of the whole data. Both approaches were used to explore the different normalization methods on the data.

3.3 Evaluation

This section describes the evaluation metrics used to assess the results of the tensor decomposition in Chapter 4. Each evaluation process we chose focuses on a different

aspect of the results and to answer the objective of this project - finding the underlying structure of attention switching. The corresponding aspect of each evaluation are focusing on are, as follow:

- Visual interpretability will focus on the neurological interpretability of each factor in the components.
- Selectivity metrics will focus on stimulus and task discrimination in the trial factor.
- Performance metrics will be overviewing the fitting of tensor decomposition method.

Other metrics were utilized to provide support on where we should focused on when visually inspecting the components, therefore the most important evaluation for this project is the visual interpretability of the components in neuroscience.

3.3.1 Visual interpretability

TCA produced three factors (neuron factor, temporal factor, and trial factor) with an arbitrary component number. We visualized the three factors of the TCA model and visually examined each individual component's factors, mainly the trial factor, and attempted to associate the pattern with any possible task variable that might influence it. For the neuron factor and temporal factor, we assessed for relevant component(s) to interpret how neural activity is modulated. This process required expert knowledge on analysing neuron activity to assist on how to interpret the modulation.

For interpretability in neuroscience aspect, the weight of the neurons in the neuron factors are order from negative to positive to illustrate the cells in a continuous spectrum. In addition, the temporal factors were marked before and after stimuli onset to show the modulation of the neural activity between the two periods. Finally, we concatenated the trials in the switching sequence between visual and olfactory blocks, while the first two blocks are the vertical stimulus and the last two blocks are the angled stimulus. The concatenation sequence was created to assist in discriminating the pattern between the two stimuli and the task blocks. We displayed the trial factor in two versions: ordered from the concatenation sequence and ordered by the trial sequence from the recording.

3.3.2 Selectivity metrics

The mice were switching their task across multiple trials in a session. Hence, the attention switches would occur throughout the trials. We created the stimulus selectivity and task selectivity metrics to see the trial variability between stimuli conditions that the mice were in. The goal of the stimulus selectivity is to measure how the trial factor of the component can discriminate between the vertical and the angled stimuli in the visual block s_v and olfactory block s_o. If the value of the stimulus selectivity in the visual or olfactory block is high, there is a difference in neural activity between the vertical or angled grating during the visual or olfactory block. Thus, stimulus selectivity is associated with the mice's task to switch between vertical and angled stimuli. The vertical stimulus is when the mice received their reward, more in Section 2.1. The stimulus selectivity are defined as:

$$s_v = \frac{\bar{ve} - \bar{an}}{\sqrt{0.5 \cdot (\sigma_{ve}^2 + \sigma_{an}^2)}}, \text{ where } ve \text{ and } an \text{ denote the vertical and angled stimulus in visual block} \tag{3.6}$$

$$s_o = \frac{\bar{ve} - \bar{an}}{\sqrt{0.5 \cdot (\sigma_{ve}^2 + \sigma_{an}^2)}}, \text{ where } ve \text{ and } an \text{ denote the vertical and stimulus in olfactory block} \tag{3.7}$$

Both Equation 3.6 and 3.7 are computing the difference in mean relative to the standard deviation of the stimuli activity. The standard deviation σ is considered as a measurement of the amount of variability in the stimuli, and the difference in the mean is considered as the signals. Stimulus selectivity in visual block is more focused, as the data was recorded in visual cortex, as introduced in Section 2.1, we want to see the attention switching between the stimuli.

The task selectivity (s_t) measures the separation between the visual and olfactory block over both vertical and angled stimuli in the trial factor of the component. We created the task selectivity similar to the stimulus selectivity to observe the difference in neural activity during the visual and olfactory blocks. If the value is high in task selectivity, it denotes a difference in activity between the visual and olfactory block in the trial factor of that component. The task selectivity is associated with mice switching between visual and olfactory blocks during their experiment. If the mice successfully learned how to switch their attention, they would switch toward the olfactory and ig-

nore the visual task. The metric is defined as:

$$s_t = \frac{\bar{vi} - \bar{ol}}{\sqrt{0.5 \cdot (\sigma_{vi}^2 + \sigma_{ol}^2)}}, \text{ where } vi \text{ denotes the visual block and } ol \text{ denotes the olfactory block} \tag{3.8}$$

Similar to the stimulus selectivity in Equation 3.6 and 3.7, the task selectivity is computing the signals in the visual and olfactory blocks against the standard deviation that is considered as the noise that occurs during those task blocks.

3.3.3 Performance metrics

Due to the nature of TCA, fitting TCA to data does not guarantee global minima convergence [8]. Thus, we ran multiple instance of TCA with multiple number of component to validate the performance of each component number. Following [19] on evaluating multiple instances from fitting TCA, the reconstruction error score, the similarity score between the models of the same component number, and cross-validation are used to assist in choosing the number of components and tensor decomposition method.

Reconstruction error score Both method of TCA (mentioned in Section 3.1) were fit to minimize the squared reconstruction error in Equation 3.1. However for the reconstruction error score to be interpretable, it is normalize to be on scale of zero to one, which is defined as:

$$\frac{\|X - \widehat{X}\|_F^2}{\|X\|_F^2} \tag{3.9}$$

where X and $\widehat{X}$ denotes data and a tensor approximated by TCA and $\|\cdot\|_F^2$ denotes the squared Frobenius norm [19]. It takes values between zero and one, with one representing poor fit and zero representing perfect reconstruction. We visualized the reconstruction error of multi-instance of each component number in a single plot called 'error plot.' If the error starts to decrease at a certain component number, we could use that as an indicator of which component number to use [19].

Similarity score Similarity score computes by comparing multiple of same component number TCA models. We wanted to compared to ensure whether each replication of each component number results similar model. If the similarity starts to decrease, we could use that as an indicator of which component number to use [19].

Before the comparison, the columns of factors were normalized to unit Euclidean length and the scalings were absorb into λ_r for each component. The rescaled estimate

of the data is defined as:

$$\widehat{X}_{ntk} = \sum_{r=1}^{R} \lambda_r w_n^r b_t^r a_k^r \tag{3.10}$$

With the vectors w^r, b^r, a^r constrained to be unit length, the similarity score between the two fitted TCA models was the comparison of the angles between latent factors [17]. The comparison between two TCA models, W, B, A and W', B', A', the score is defined as:

$$\max_{w \in \omega} \frac{1}{R} \sum_{r=1}^{R} [(1 - \frac{|\lambda_r - \lambda_{w(r)}|}{\max(\lambda_r, \lambda_{w(r)}}) (w_r^T w'_{w(r)} \cdot b_r^T b'_{w(r)} \cdot a_r^T a'_{w(r)})] \tag{3.11}$$

where ω denotes the set of all permutations of factors and w is one of the permutation. To ensure the comparison is fair, we computed the scores of all possible permutations of the model and select the highest score. For example, a model with component number of three ($R = 3$), the set of computations are $\omega = (1,2,3), (1,3,2), (2,3,1), (3,2,1)$. The score scales between zero and one, with one representing the two models are identical and zero representing the two models are different. We followed a guideline given in [19] that similarity score within the range of 0.6-0.8 are considered to have very similar result.

Cross validation To evaluate the risk of overfitting within the model, we performed cross validation on the model. We randomly masked 20% of the data for training set by applying a masking tensor M. The process of masking the tensor is defined as:

$$\min_{W,B,A} \|M * (X - \widehat{X})\|_F^2 \tag{3.12}$$

where it performs entrywise multiplication (Hadamard product) with masking tensor M. Fitting into TCA, this was solved by using the optimization update rules with missing data in dependent variable. The normalized reconstruction error on training set was computed as:

$$\frac{\|M * (X - \widehat{X})\|_F^2}{(M * X)_F^2} \tag{3.13}$$

For testing, the normalized reconstruction error was computed as:

$$\frac{\|(1 - M) * (X - \widehat{X})\|_F^2}{\|(1 - M) * X_F^2\|} \tag{3.14}$$

where only the unmasked part of the tensor was considered in the computation of the error.

3.4 Refitting

During our observation on the TCA models, the neuron factor in the TCA models, demonstrates in Chapter 4, were noticeably skewed. The skewness in multiple neuron factors were concerning since it might signify that only a small amount of neurons were responsible for creating the unique pattern(s). Therefore to confirm the robustness of the data, we created a refitting method to estimate the percentage of cells that were responsible for creating a pattern with high stimulus selectivity in both visual and olfactory block, introduced in Section 3.3.2.

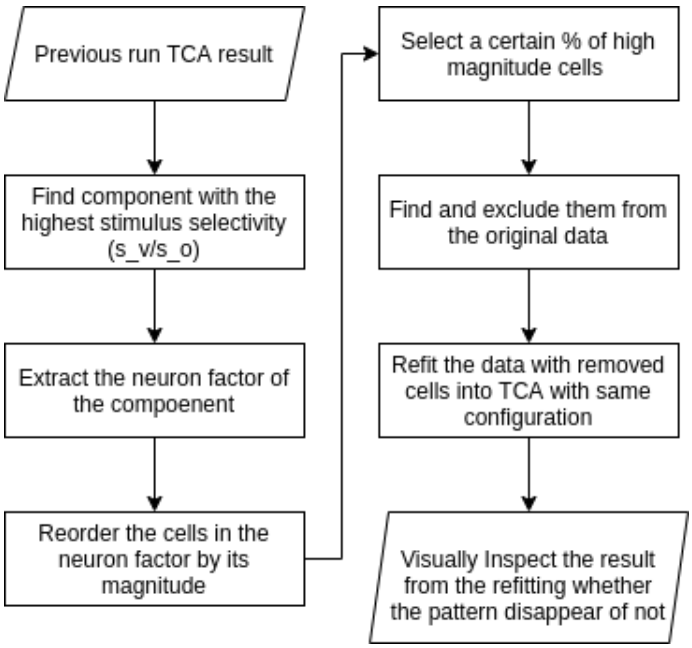

Figure 3.1: Flow diagram of the refitting procedure

The flow diagram for the refitting procedure can be seen in Figure 3.1. The previously fitted TCA result with a component with high stimulus selectivity was chosen to perform the refitting. First, the neuron factor of the component was extracted from the component. Then, the cells in the neuron factor were reordered by their magnitude. At each threshold, the high magnitude cells were selected. The threshold range was ranging from 10% to 50% of cells selected. Next, the selected cells were located in the original data and removed completely. Finally, the processed data were fitted into the same decomposition, normalization, and outlier detection method. The results were visually inspected to see whether the pattern disappears at which threshold.

Chapter 4

Results and Discussion

The previous chapter outlined the pipeline that were applied to the attention switching dataset. The pipeline included normalization and outlier detection applied to the dataset, different tensor decomposition methods, and evaluation methods.

In Section 4.1, we report on TCA results with the variety of component numbers and configuration of the data to find which configuration is the most suitable for this experiment. Section 4.2 aims to analyze and identify attention switching dynamics within the trial factors of the TCA models from each session of each mice. Next, Section 3.4 investigates the robustness of the TCA models by performing the refitting process, introduced in Section 3.4. Finally, Section 4.4 reports on a high task selectivity and additional components with interesting neural dynamic in the trial factors.

4.1 Exploring TCA configuration

In Section 3.1, the experiment explores the standard CP and non-negative CP (NCP) decomposition. We examined each result's factors and evaluation metrics to see which decomposition method with each outlier detection and normalization method provided the most biological meaningful insights of the neural activity pattern in the attention switching task. First, we explored which decomposition methods would be more suitable for the attention switching task. Our dataset includes negative neuronal signals, and we assume that the negative signals might contain necessary information regarding attention switching, such as the mice might react negatively when the reward was not presented. Hence, CP decomposition is assumed to be a suitable method for the interpretability of negative signals. However, non-negative CP decomposition with our normalized version of the data, introduced in Section 3.2.2, this method might

improve the fitting of the data and result in better evaluation metrics. We found that non-negative CP and standard decomposition do not differ much according to the similarity score plot, reconstruction error plot, and the visual results in TCA models, more in evaluation in Section 3.3.

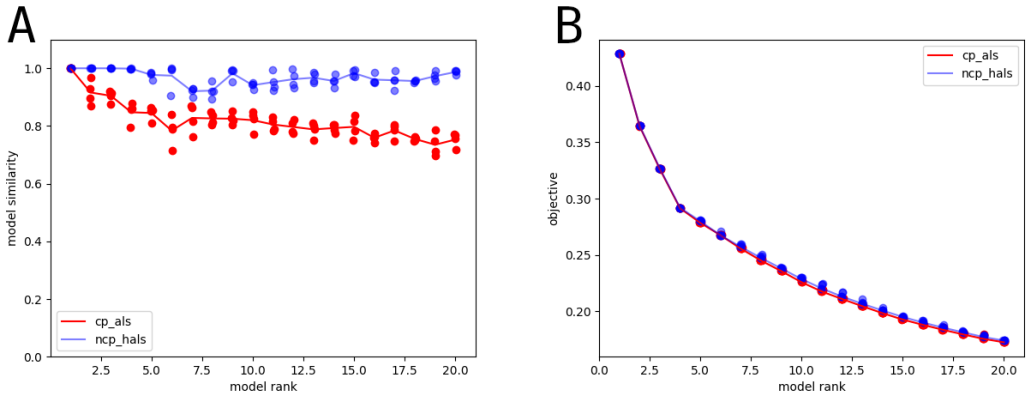

Figure 4.1: (A) Reconstruction error (objective) on Y-axis of each component number on X-axis(model rank) between non-negative CP (blue) and standard CP decomposition (red). (B) Similarity score plot of each component number (model rank) between non-negative CP (red) and standard CP decomposition (blue). The fitted with rescaled version of the data with exclude 5% of low signal-to-noise (SNR) cells was fitted into both CP decomposition method. Each dots corresponds to different optimization run.

For comparison between the decomposition method, the experiment was conducted by fitting rescaled version of the data into standard and non-negative CP decomposition methods. Figure 4.1A illustrates reconstruction error (objective) of each component number (model rank) between the decomposition methods. There is no difference in the reconstruction error between the two methods. However, for the similarity score between the same component number runs, Figure 4.1B illustrates that non-negative CP decomposition results in a higher overall similarity score. The similarity score remained constant throughout the number of components. This demonstrates that non-negative CP decomposition converged more consistently due to the stronger uniqueness conditions of the method when compared to the standard CP decomposition method [15]. Similar founding had been seen in [19], where their experiment results similar reconstruction error and non-negative CP gained a higher overall similarity score.

Furthermore, we found that the non-negative CP decomposition does not differ much from the standard CP decomposition in the visual illustration of the TCA mod-

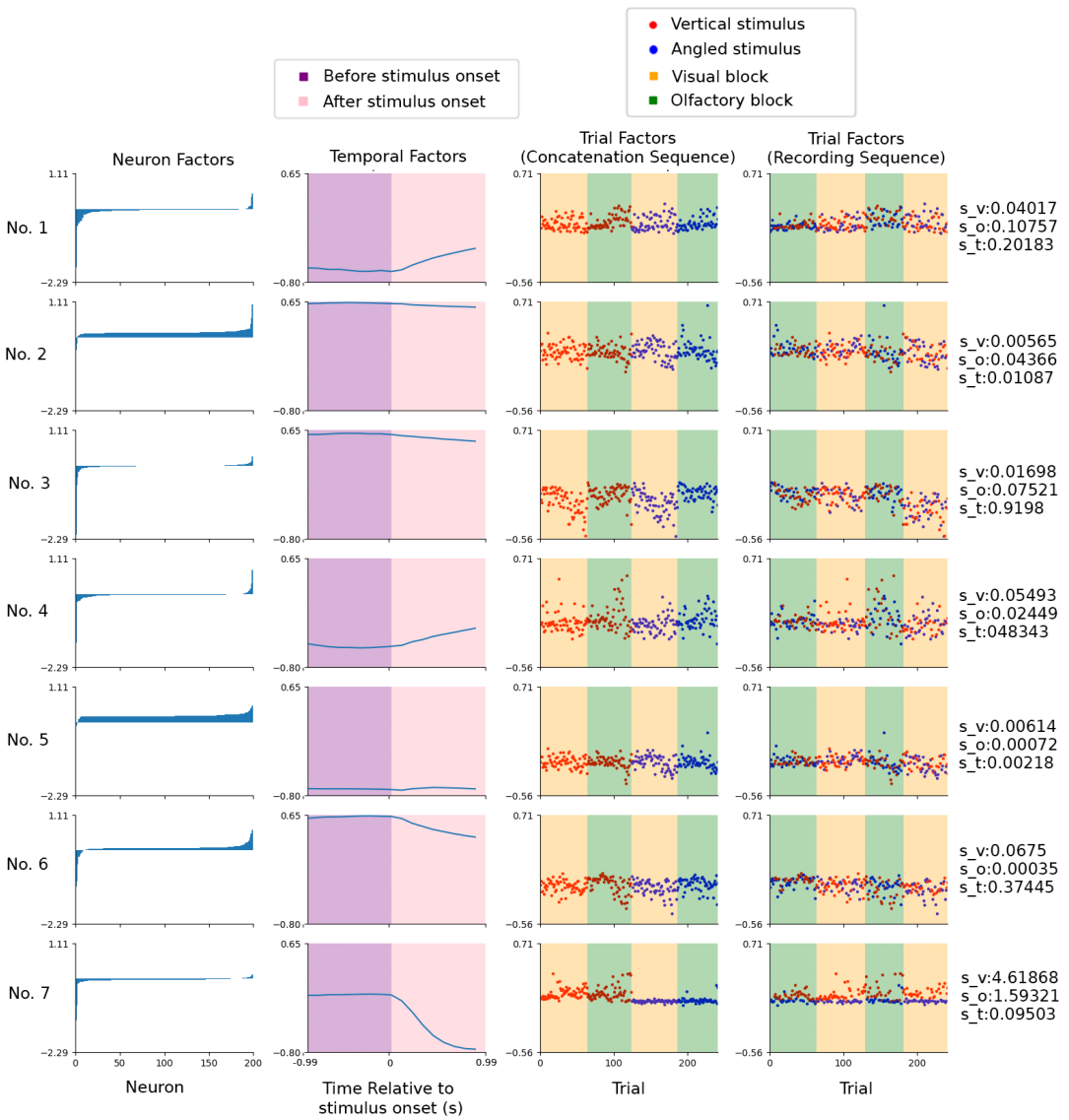

Figure 4.2: Example of 7-component standard TCA model fitted with rescaled data. The columns are presenting: neuron factor (first; left column), temporal factor (second column), and trial factor ordered by concatenation sequence (third column) and recording sequence (fourth column).

els. Non-negative TCA model's trial and temporal factor in component 7 in Figure 4.3 is identical to standard TCA model's trial and temporal factor in component 7 in Figure 4.2. The only difference is that the non-negative TCA forced the constraint on the factors, which removed the negative space, and the pattern is identical despite the neuron and temporal factors are inverted. Factors TCA is invertible. Hence, standard CP decomposition and non-negative CP decomposition do not differ drastically in similarity scores, reconstruction scores, and visual low-dimensionality descriptions. For

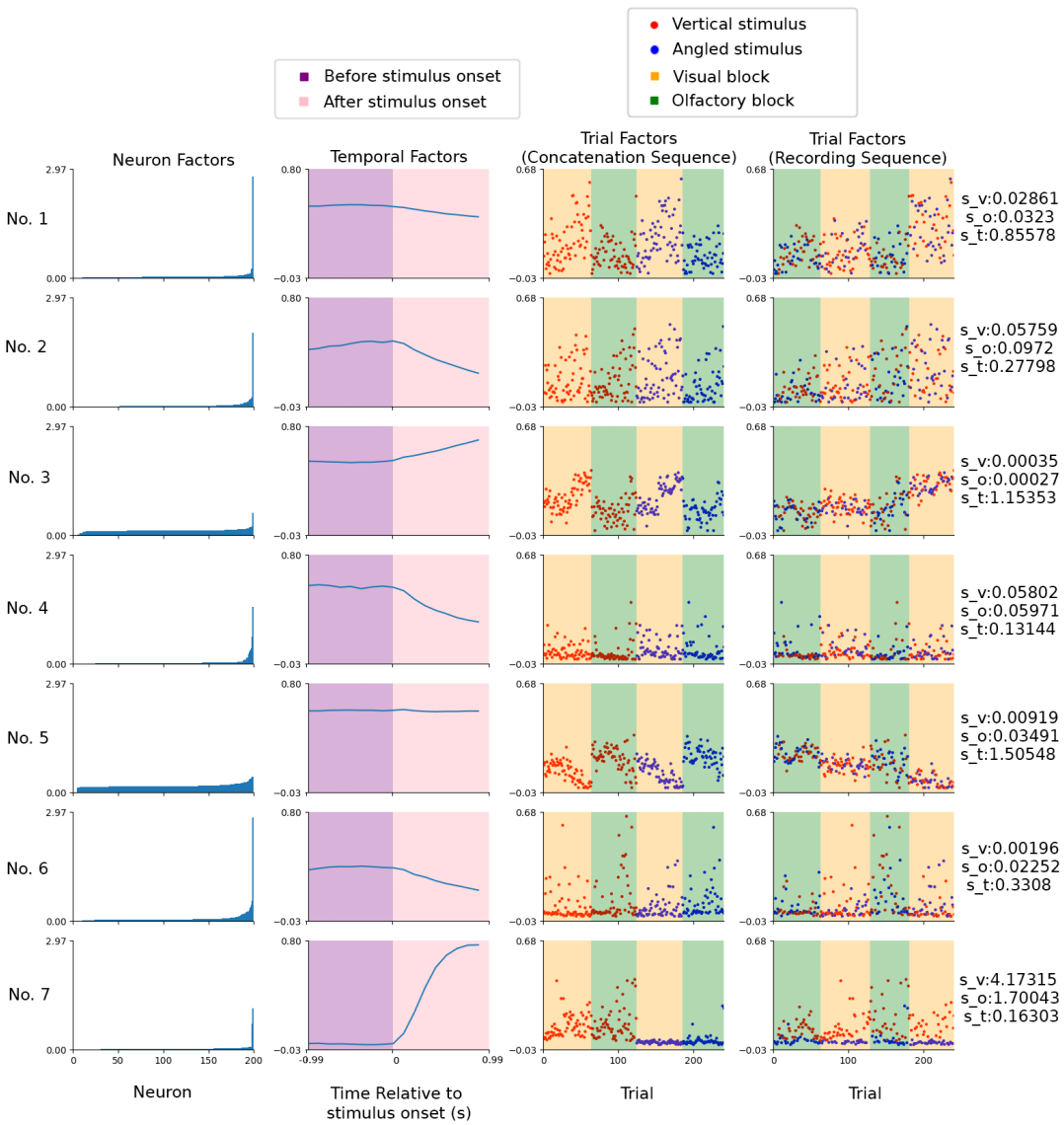

Figure 4.3: Example of 7-component non-negative TCA model fitted with rescaled data. The columns are presenting: neuron factor (first; left column), temporal factor (second column), and trial factor ordered by concatenation sequence (third column) and recording sequence (fourth column).

simplicity and biological interpretability, we used the standard CP decomposition as our default method to keep the negative signals for the biological interpretability of the results.

We explored the normalization and outlier detection method options. The results were examined by using our evaluation methods introduced in Section 3.3. We compared the outlier excluded by using the signal-to-noise ratio (SNR) and the original

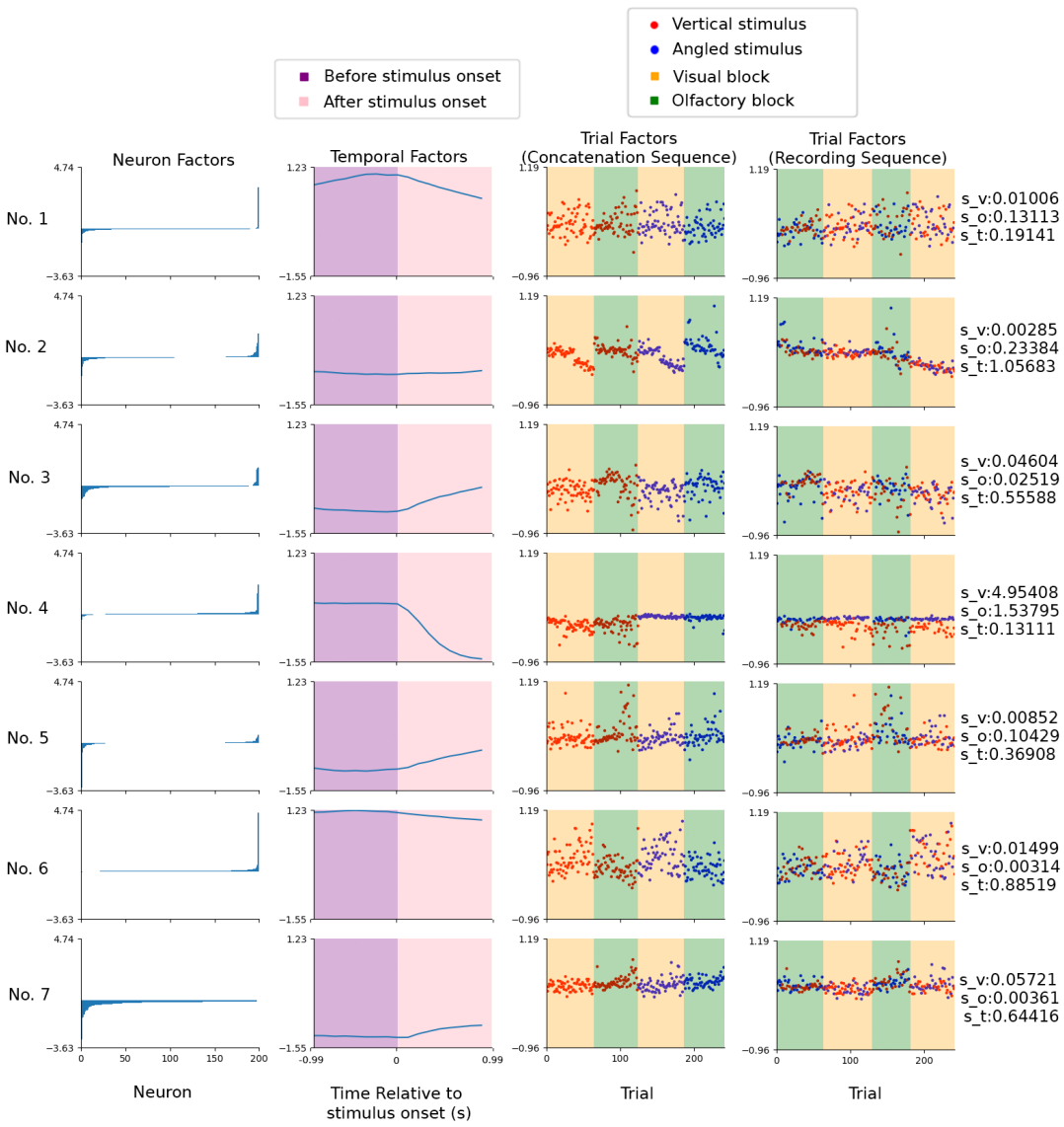

Figure 4.4: The unnormalize data with 5% of low signal-to-noise (SNR) cells excluded was fitted into standard TCA. The columns are presenting: neuron factor (first; left column), temporal factor (second column), and trial factor ordered by concatenation sequence (third column) and recording sequence (fourth column).

version of the data. We mainly focus on the visible results from the fittings since there is no difference in reconstruction error and similarity score plots. This is shown in the experiment with normalized data for a fair comparison. The reconstruction and similarity plots for data without outlier exclusion are displayed in Figure A.1, A.2 and with excluding 5% of low SNR cells are displayed in Figure 4.1.

Using the standard CP decomposition to decompose the data, we found that using

standard CP decomposition with excluding 5% of low SNR cells reveals the dynamic neural patterns in lower component numbers than with no outlier exclusion. Figure 4.4 illustrates the TCA models with unnormalized and 5% of low SNR cells excluded data. Trial factors in component 4 already recovered an interesting neural dynamic of attention switching, more in Section 4.2. However, in Figure A.5 had not recovered any biological significant patterns in trial factors. This suggests that excluding outliers is beneficial for reducing computational time and possibly removed irrelevant cells for attention switching tasks, as it requires a lower component number to reveal the unique patterns. We had tested excluding 10% of low SNR cells, however we did not find any significant difference between 5% and 10% exclusion, the trial factor in Figure 4.4 component 4 is also recovered in 10% exclusion, as shown in Figure A.7 component 1. Since we did not find any difference, we used 5% removal of low SNR cells to retain as much information in the data while optimizing the TCA runs. Similarly, z-score normalization data did not result in any significant visual difference when compared to rescaled and unnormalized data, as illustrated in Figure A.8. Hence, we explored the data with standard TCA and unnormalized data with 5% of low SNR cells excluded since normalization did not improve the visualization of TCA models.

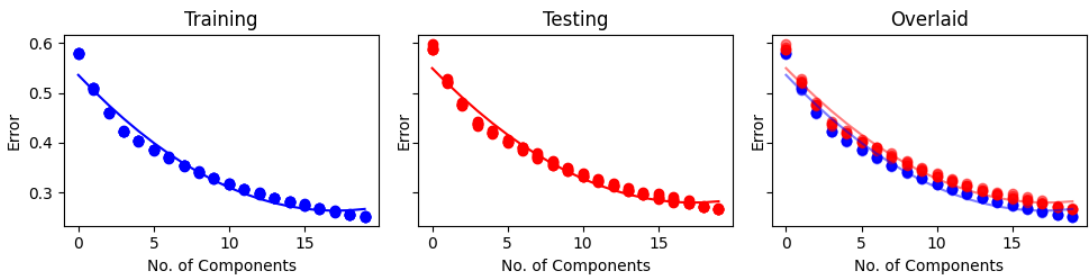

Figure 4.5: Normalized reconstruction error in training set (blue) and test set (red) for cross validation performed on TCA with standard CP decomposition, and remove 5% low SNR cells, while holding 20% of data in training set.

For selecting the component number, we evaluated based on the visual results, similarity score, reconstruction error score, and cross-validation. However, we did not find any component number that drastically decreased in the similarity score and reconstruction error, as illustrated in Figure A.3 and A.4. Following [19]'s evaluation, we tested cross-validation to see whether TCA models would show drastic error after performing cross-validation throughout the component numbers. The result of cross-validation did not signify any drastic changes in the error of each component number after masking 20% of the data and test the model, as shown in Figure 4.5. It was

expected that cross-validation since TCA does not easily overfit as it uses very few parameters to reduce a large dataset into a compact description [19]. Thus, we rely on the visual biological interpretability to determine the component number to use. However, with a visual inspection, we found the number of components above 7 is when dynamic neural patterns start to reveal in most of the session data we have, and above 13 is when patterns in the three factors start to repeat. Figure 4.4 illustrates an interesting trial component at component 7, which is further discussed in Section 4.2. The same pattern was recovered in the 13-components TCA model at component 4, as shown in Figure 4.6, while the same pattern was not yet recovered in the 5-components TCA model, as shown in Figure A.6. Thus, for the results, we use results ranging from 7 to 13 component numbers from the standard CP decomposition method with unnormalized data and 5% low SNR cells exclusion as our default setting. This setting was used to all mice when evaluating.

4.2 Underlying neural dynamic of attention switching

We next examined the neural, temporal, and trial factors visually to see whether the TCA with the configuration as in Section 4.1 can provide an insightful summary of the attention switching activity in different mice. Each mouse had multiple sessions and we treated each session as an individual set of data due to the inconsistent in neuron and trial recording number. Each TCA model result components that identified an assemble of cells (neuron factor; left column), within-trial modulation (temporal factor; second column), and trial-to-trial variability (trial factor; third column and fourth column). The trial factors were arranged in concatenation sequence (third column) and in recording sequence to visualize the mice's stimulation (fourth column). Additionally, all the conditions of the trial are color-coded, and the period before and after stimulus onset are marked in temporal factor. This assist in the visual interpretability of the TCA models according to the attention switching task. To investigate on attention switching, the temporal and trial factors were the main focus since the mice switched their tasks across multiple trials in a session. The evaluations we focused on were the stimuli selectivity and visual evaluation, introduced in Section 3.3. The stimuli selectivity helped to provide an estimate which trial factor of a component may contain a neural dynamic that describes the attention switching between stimuli, while the visual evaluation was used to verify the neural dynamic presented in the pattern. We found that TCA recovered a pattern in the trial factor, which recovered the neural dynamics

of attention switching between the vertical and angled stimuli, and this pattern was found in two out of six mice in the dataset.

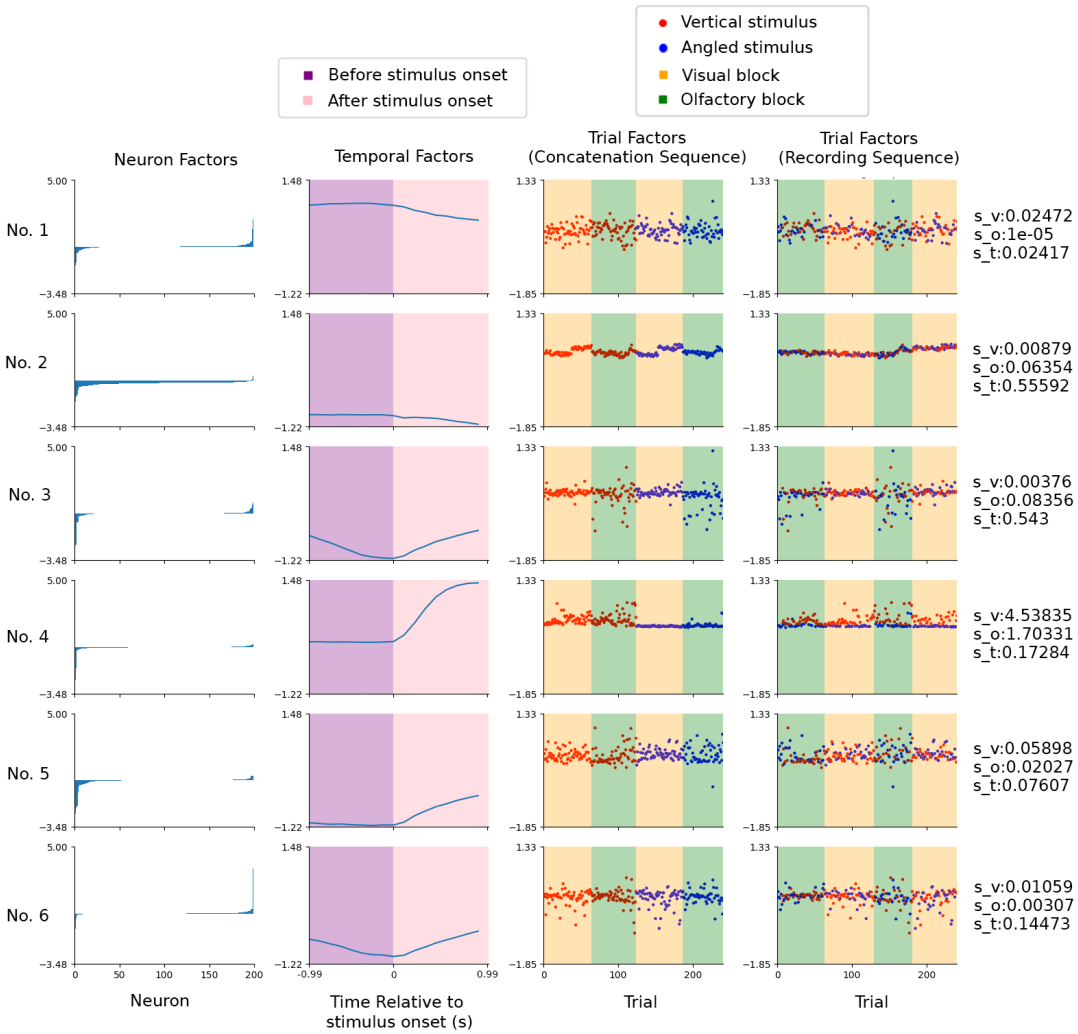

Figure 4.6: Six low-dimensional components out of thirteen components of Mouse *A*, each containing a neuron factor (first; left column), temporal factor (second column), and trial factor ordered by concatenation sequence (third column) and recording sequence (fourth column). For each trial factor, the behavioural conditions are color-coded based on the meta-data labels (see legend above). For each temporal factor, the before and after onset stimulus are color-coded based on the meta-data labels (see legend above). The configuration for this model mentioned in Section 4.1. Additional components are shown in: Figure A.9

The pattern was first found in one of the mice (for simplicity, we called it Mouse *A*). The trial factor that is associated with the neural dynamics of attention switching

between stimuli in Mouse *A* is presented in component 4 of Figure 4.6 (the remaining components contains irrelevant patterns for this section and are shown in Figure A.9). The trial factor has the highest stimulus selectivity in both visual and olfactory blocks, introduced in Section 3.3.2, with the visual higher than the olfactory's stimulus selectivity. High stimulus selectivity in the visual block indicates that the trial component revealed a pattern with the difference in neural activity between the vertical and angled stimulus during the visual block. This indicates the mouse being attentive toward one of the stimuli. We assumed that it was expected for the stimulus selectivity in visual to be higher since the data was recorded in the visual cortex, more in Section 2.1. Thus, the stimulus selectivity metric successfully assisted in finding the component with high neural selectivity toward the vertical stimulus while the neural activity remains suppressed in the angled stimulus. This can be seen as the angled stimulus (blue dots) appears to remain constant, while the vertical stimulus (red dots) appears to be highly and positively active in both the concatenation sequence and recording sequence (third and fourth column) in Figure 4.6.

The selectivity toward vertical stimulus and suppression in neural activity toward angled stimulus suggest that the pattern is describing the mouse was attentive toward the vertical stimulus and ignored the angled stimulus. This led us to speculate that the reward influenced this pattern. Vertical stimulus is when mice were rewarded during the visual session. However, we did not analyze the reward conditions in the session. Additionally, the temporal factor (second column) shows heighten of neural activity within trial after the stimulus onset. This suggests the heighten of neural activity after the stimulus is shown within the trials. With switching between the stimulus in the trial factor and activity heighten after the stimulus onset in the temporal factor, we concluded that TCA successfully recovered the neural dynamics of attention switching between stimuli.

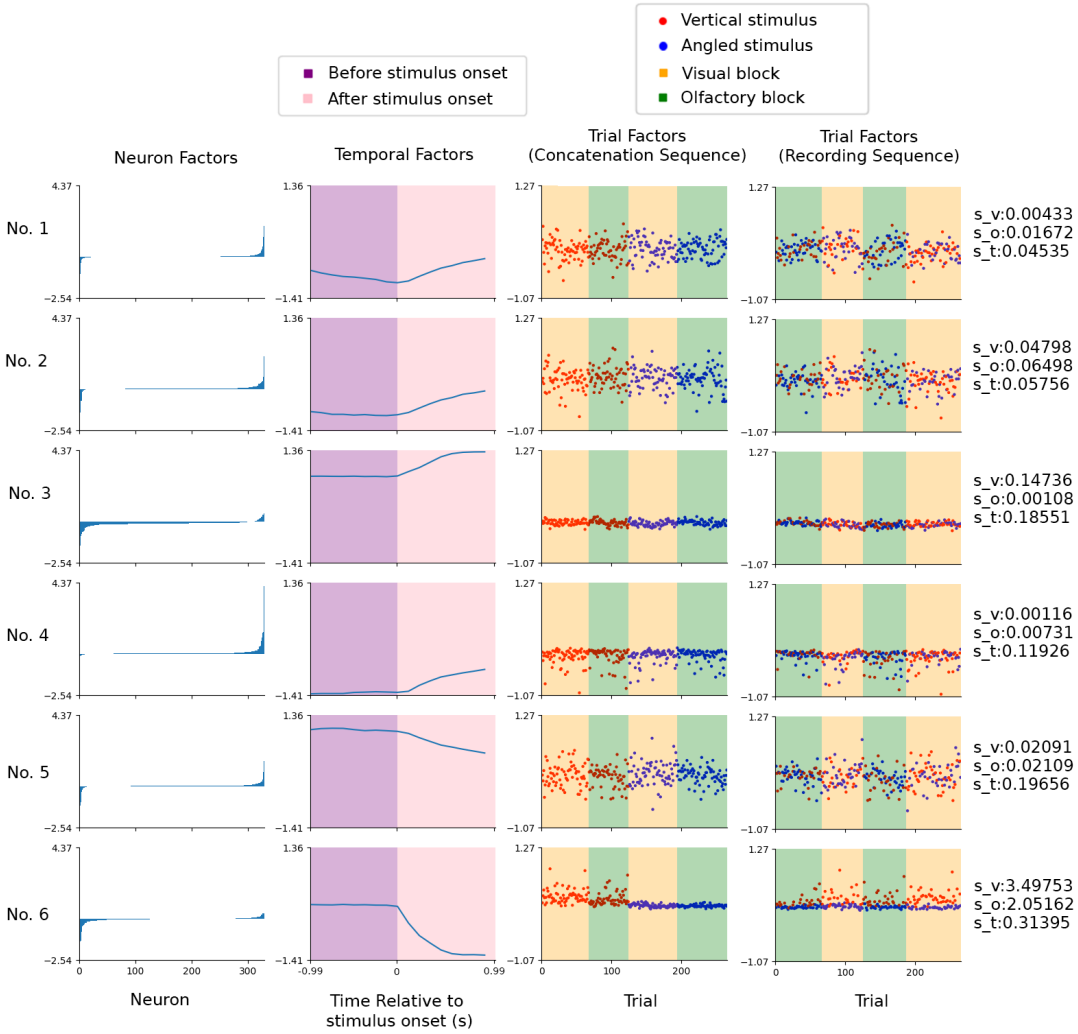

Figure 4.7: Six low-dimensional components out of ten components of Mouse *B*. Same figure setting and TCA configuration as Figure 4.6. Additional components are shown in: Figure A.10

Furthermore, the pattern in trial factor associated with attention switching between stimuli appeared in another mouse (for simplicity, we called the second mouse, Mouse *B*). Figure 4.7 shows six components from 10-component TCA model (the remaining components are shown in Figure A.10. The identical pattern in trial factor we found are shown in Figure 4.6 component 4 for Mouse *A* and in Figure 4.7 component 6 for Mouse *B*. Despite that, the temporal factor Mouse *A* increases after the stimulus onset (pink background) and decreases in Mouse *B*. However it is not concerning since the factors are invertible. We interpreted the temporal factor to be identical. It was interesting how there is a common neural dynamic in two different mice which were

extracted using TCA.

4.3 Robustness of TCA

After visualizing different mice TCA models, we found the neuron factors showed the cells to be skewed to either very positively (Figure 4.7, component 4) or very negatively in value (Figure 4.6, component 3). We assumed that TCA revealed only subset of cells that influence each components' trial and temporal factors. We are particularly interested in the pattern in Figure 4.6 component 4, and Figure 4.7 component 6, as the neuron factors often appeared to be negatively influence by a subset of cells. To verify the robustness of TCA models, refitting method was used, introduced in Section 3.4, to investigate the robustness of components identified by TCA.

Based on the previous TCA model, the neuron factor of a component with highest stimulus selectivity in the trial factor was chosen. The magnitude of the cells in that neuron factor would be removed from 10% to 50%, increase incrementally by 10%. The goal was to remove a percentage of cells according to their magnitude and observe when the pattern disappeared. The experiment was tested on Mouse A as the pattern with high stimulus selectivity appeared to be very strongly dominated throughout TCA models with R components ranging from eight to twenty. We found that $R = 19$ components allows the pattern to remain, even after 20% of the cells were removed as shown in Figure 4.8, component 19, until 40% of the cells were removed, illustrate in Figure 4.9. Other refitting result of 10% and 30% are shown in Figure A.12 and A.13. Furthermore, we applied refitting to previously fitted TCA models R components (ranging from eight to twenty) of Mouse A that contained the pattern with high stimulus selectivity in the trial factor. The pattern in trial factor with high stimulus selectivity remained after removing 50% of the cells in TCA models with a component number below 17, as shown in Figure A.14. This suggested that multi-cells were responsible for influencing the underlying dynamic pattern. Hence, TCA recovered a robust component.

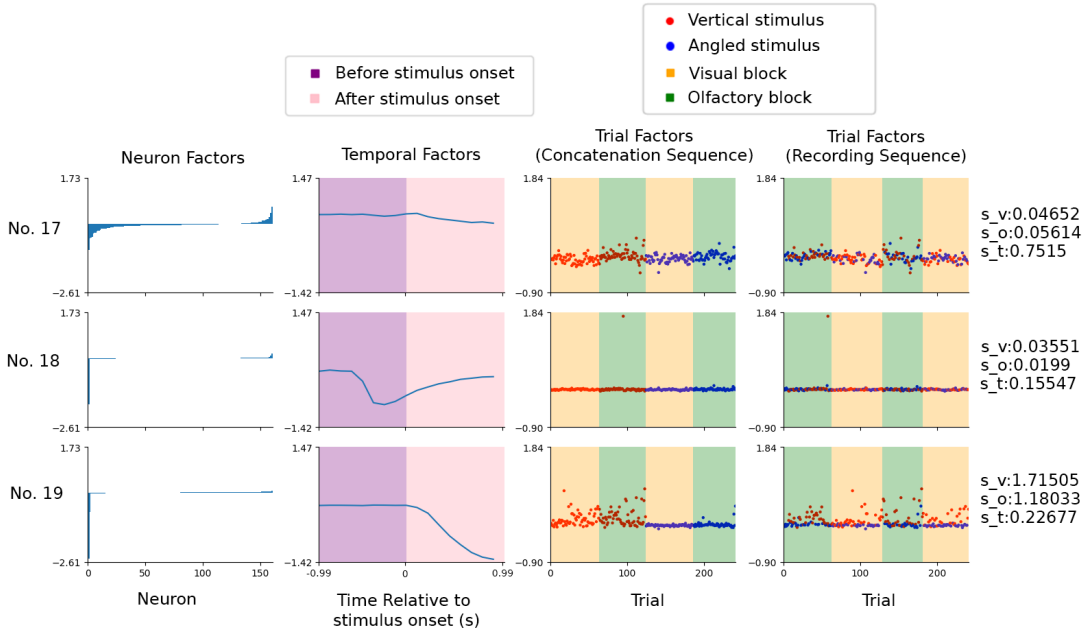

Figure 4.8: Three low-dimensional components out of nineteen components of Mouse *A* refitting result after removing 20%. Additional components are in Figure A.11A.

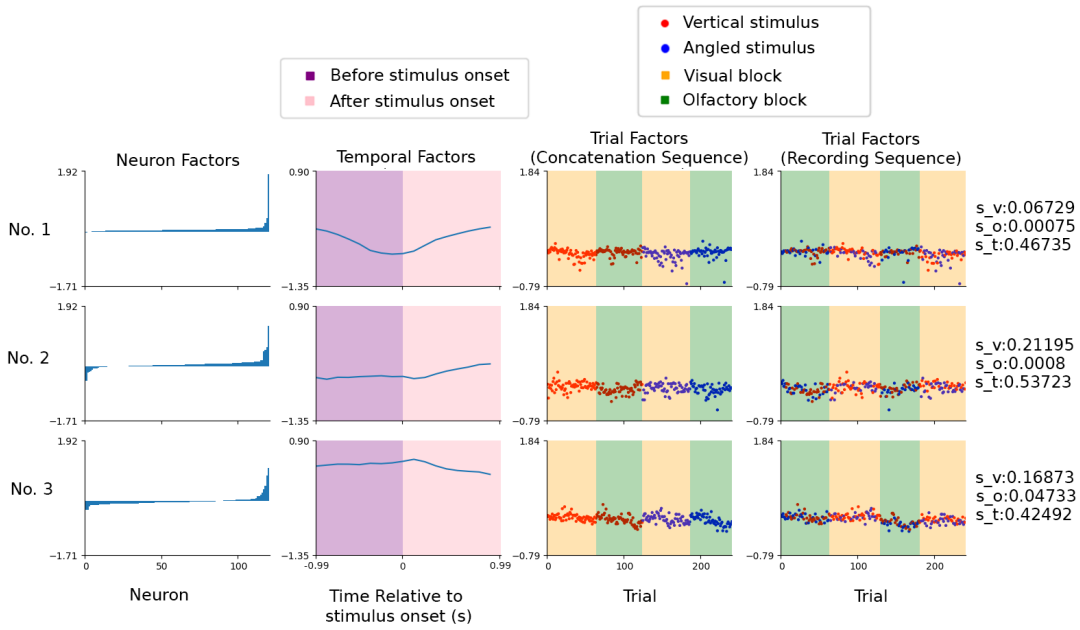

Figure 4.9: Three low-dimensional components of Mouse *A* refitting result after removing 40%. Additional components are in Figure A.11B.

4.4 Separation in task selectivity and additional component

Besides the high stimulus selectivity being revealed by TCA, we found an additional component with high task selectivity, introduced in Section 3.3.2. Similar to the stimulus selectivity, which measures the separation between stimuli in the visual or olfactory block, the task selectivity measures the separation between the olfactory and visual block in the trial factor. Figure 4.10 shows three components from a 13-component TCA model in Mouse *A* (the remaining components are shown in Figure 4.6 and Figure A.9). The component with high task selectivity is component 2. There is an increase of activity when the trials are in the olfactory block (green) and steady during the visual block in the ordering sequence of the trial factor. In addition to TCA recovering a component associating with switching between stimuli, TCA is proven to recover a component that switches between visual and olfactory tasks. The task selectivity metric assists in discovering which component that TCA recovers can be associated with task discrimination.

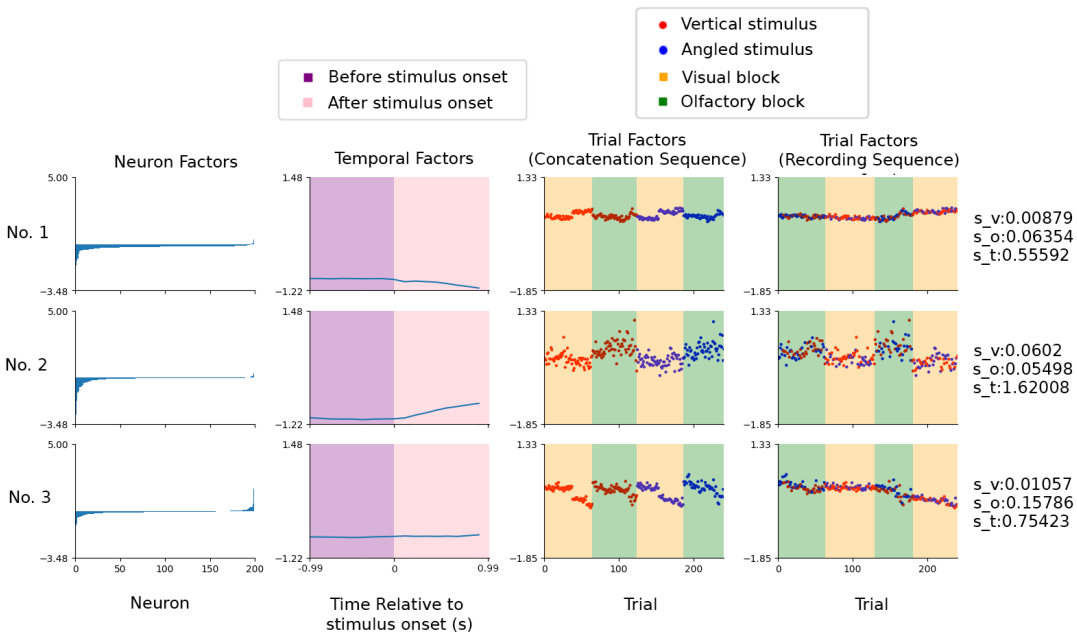

Figure 4.10: Three components of Mouse *A*, where component 2 shows a high task selectivity in trial factor, and component 1 and 3 are the unknown component with possible association with thirsty of the mouse.

From exploring other components, we further found interesting components that

switched abruptly during the session. We speculated that the switched was motivation of the mouse reached satiety. The mouse was motivated by hunger, as stated in Section 2.1. Therefore, the mouse might become satisfy with its hunger or thirsty toward the end of the session. Figure 4.10 shows additional two components (component 1 and 3) that are not correlated to any selectivity metrics or show any task variable in trial factor ordered by the concatenation sequence (third column). However, in the trial factor ordered by the recording sequence (fourth column), component 1 peaks toward the end of the trial, and component 3 decreases toward the end of the session. We speculated that component 3 is associated with the thirstiness of the mice. Therefore, the drops toward the end of the trial might signify that the mouse was satisfied with its thirstiness. For component 1, the component may be describing the arousal of the mouse, which is currently unknown which task variable to be associated with. TCA reveals multiple task variables, including switching between stimuli, olfactory and visual block, and another unknown variable that needs further discovery.

Chapter 5

Conclusion

The main goal of this dissertation was to recover the neuronal activity changes as mice switch attention between tasks inside the visual cortex by extracting a low-dimensional description of the neural recording using TCA. The complexity of the neural recording of the attention switching dataset had imposed a challenge on how to explore the neuronal activity changes within the mice while looking at how improvements in sensory behavior are linked to change in neural circuits. The dimensionality reduction technique had proven to be effective at digesting neuronal data in an unsupervised manner into a set of low-dimensional descriptions while retaining the trial-to-trial modulation. TCA [19] was introduced as the technique to reduce neuronal data into three sets of interpretable biological factors: the neuron, temporal and trial factors.

To explore the TCA with the attention switching data, we implemented a pipeline in Python [18] to prepare the data for the TCA. The pipeline included: non-negative and standard CP decomposition methods, data preprocessing, evaluations and refitting. The signal-to-noise ratio (SNR) was used for the outlier detection method to remove low SNR cells, as they might be irrelevant cells in the attention switching task. In evaluation, we created selectivity metrics to quantitatively estimate which trial factors has a pattern associated to attention switching between stimuli or the task condition. Refitting was introduced by us to measure the robustness of the component from the TCA models by removing percentage of cells, according to its magnitude, and observe the visual results whether the pattern disappear or not.

With the objective in Section 1.2, we explored the TCA models to investigate their capability to digest attention switching tasks while retaining the trial-to-trial variability. First, we explored the configuration for TCA. In Section 4.1, we demonstrated that standard CP decomposition along with SNR outlier methods did optimize the TCA

fitting as it requires less component number to be fitted for TCA to recover interpretable neural dynamics. The difference between standard CP decomposition and non-negative CP decomposition was mainly the performance metric. However, based on the visual evaluation, there was no difference. Then, in Section 4.2, we observed the usage of stimulus selectivity and the attractive biological component that had high stimulus selectivity. The trial factor was mainly focused on investigating the modulation across the session as the mouse switched through the task. We found a trial factor of a component with high selectivity toward the vertical stimuli while suppressed toward the angled stimuli. The pattern was associated with mice switching attention between the vertical and angled stimuli. It was concluded that stimulus selectivity could estimate which component had the potential neural dynamic of attention switching between stimuli, and TCA could recover the neural dynamic through the trial factors. Next, Section 3.4 investigated the refitting method, introduced in Section 3.1. We found that identified components by TCA were very robust since the pattern with high stimulus selectivity in trial factor could withstand after 50% of the cells were removed when the component number was lower than 17.

Finally, we identified additional components with our task selectivity metrics and other interesting trial factors that we speculate their association with satiety in Section 4.4. The contribution of this project consist of the selectivity metrics, refitting process and proven capability of TCA in extracting neurological meaning of the data in multiple low-dimensional descriptions.

5.1 Future work

In Section 2.1, the dataset included the pre-post learning session of the mice. The extension of additional observation between learning and switching using TCA could result in interesting underlying neural dynamics to fully visualize the neuronal changes between learning and switching [7, 14]. Furthermore, mentioned in Section 4.2, we did not explore the reward condition of the mice during the session. This might help to understand the influence of the patterns in the trial factor more in-depth and associate more strongly with the neural dynamics of the attention switching we found in this project. Besides the additional condition and data, we only explored popular CP decomposition methods. However, there are alternative optimizers that could improve the result of TCA. For example, non-negative CP decomposition using the Block Coordinate Descent Method [20] could improve the quality of the decomposition and go

beyond the limitation of the ALS optimizer. We conclude that TCA can extract task-related structures and reduced them into rank-1 modulation of neural activity tensor.

Bibliography

[1] J Douglas Carroll and Jih-Jie Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, 35(3):283–319, 1970.

[2] Pierre Comon, Xavier Luciani, and André LF De Almeida. Tensor decompositions, alternating least squares and other tales. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 23(7-8):393–405, 2009.

[3] John P Cunningham and M Yu Byron. Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, 17(11):1500–1509, 2014.

[4] Ashesh K Dhawale, Rajesh Poddar, Steffen BE Wolff, Valentin A Normand, Evi Kopelowitz, and Bence P Ölveczky. Automated long-term recording and analysis of neural activity in behaving animals. *Elife*, 6:e27702, 2017.

[5] Peiran Gao and Surya Ganguli. On simplicity and complexity in the brave new world of large-scale neuroscience. *Current opinion in neurobiology*, 32:148–155, 2015.

[6] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[7] Adil G Khan, Jasper Poort, Angus Chadwick, Antonin Blot, Maneesh Sahani, Thomas D Mrsic-Flogel, and Sonja B Hofer. Distinct learning-induced changes in stimulus selectivity and interactions of gabaergic interneuron classes in visual cortex. *Nature neuroscience*, 21(6):851–859, 2018.

[8] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[9] Joseph B Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear algebra and its applications*, 18(2):95–138, 1977.

[10] Henry Lütcke, David J Margolis, and Fritjof Helmchen. Steady or changing? long-term monitoring of neuronal population activity. *Trends in neurosciences*, 36(7):375–384, 2013.

[11] Yael G Maguire, Mikhail G Shapiro, Thaddeus R Cybulski, Joshua I Glaser, Dario Amodei, P Benjamin Stranges, Reza Kalhor, David A Dalrymple, Dongjin Seo, Elad Alon, et al. Physical principles for scalable neural recording. *Frontiers in computational neuroscience*, 7:137, 2013.

[12] Carrie J McAdams and John HR Maunsell. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. *Journal of Neuroscience*, 19(1):431–441, 1999.

[13] Jasper Poort, Katharina A Wilmes, Antonin Blot, Angus Chadwick, Maneesh Sahani, Claudia Clopath, Thomas D Mrsic-Flogel, Sonja B Hofer, and Adil G Khan. Learning and attention increase visual response selectivity through distinct mechanisms. *bioRxiv*, 2021.

[14] Jasper Poort, Katharina A Wilmes, Antonin Blot, Angus Chadwick, Maneesh Sahani, Claudia Clopath, Thomas D Mrsic-Flogel, Sonja B Hofer, and Adil G Khan. Learning and attention increase visual response selectivity through distinct mechanisms. *bioRxiv*, 2021.

[15] Yang Qi, Pierre Comon, and Lek-Heng Lim. Uniqueness of nonnegative tensor approximations. *IEEE Transactions on Information Theory*, 62(4):2170–2183, 2016.

[16] S. R Schultz. Signal-to-noise ratio in neuroscience. *Scholarpedia*, 2(6):2046, 2007. revision #137197.

[17] Giorgio Tomasi and Rasmus Bro. A comparison of algorithms for fitting the parafac model. *Computational Statistics & Data Analysis*, 50(7):1700–1734, 2006.

[18] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009.

[19] Alex H Williams, Tony Hyun Kim, Forea Wang, Saurabh Vyas, Stephen I Ryu, Krishna V Shenoy, Mark Schnitzer, Tamara G Kolda, and Surya Ganguli. Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis. *Neuron*, 98(6):1099–1115, 2018.

[20] Yangyang Xu and Wotao Yin. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on imaging sciences*, 6(3):1758–1789, 2013.

Appendix A

Additional Plots

A.1 Additional evaluation metric

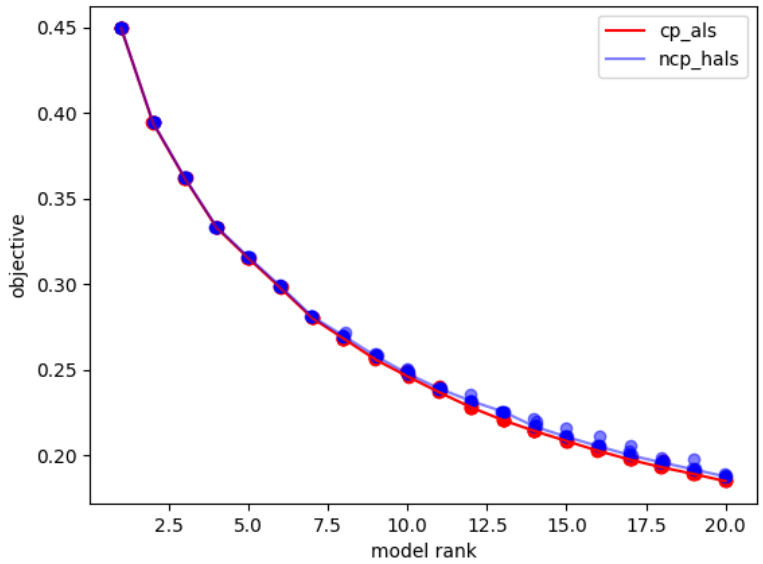

Figure A.1: Reconstruction error (objective) plot of each component number (model rank) between non-negative CP (blue) and standard CP decomposition (red), where both were fitted with rescaled version of the data without outlier exclusion. Each dots corresponds to different optimization run

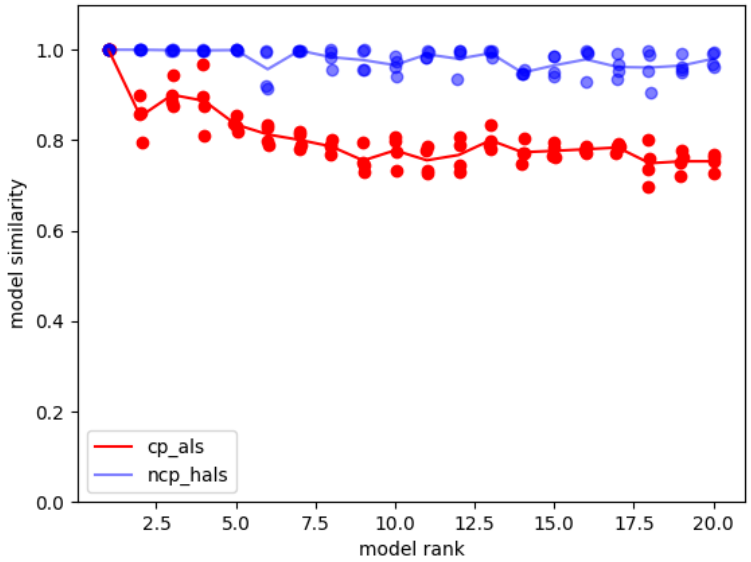

Figure A.2: Similarity score plot of each component number (model rank) between non-negative CP (blue) and standard CP decomposition (red), where both were fitted with rescaled version of the data without outlier exclusion. Each dots corresponds to different optimization run

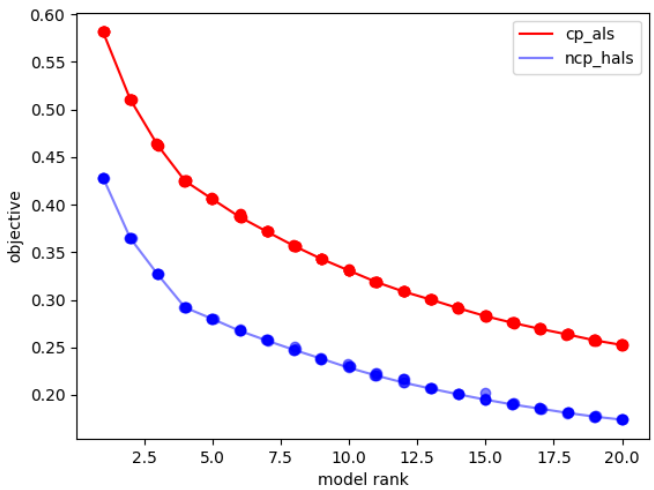

Figure A.3: Reconstruction error (objective) plot of each component number (model rank) between non-negative CP (blue) and standard CP decomposition (red), where rescaled data is fitted into non-negative CP and unnormalized data is fitted in standard CP. Both data version excluded 5% of low SNR cells. Each dots corresponds to different optimization run

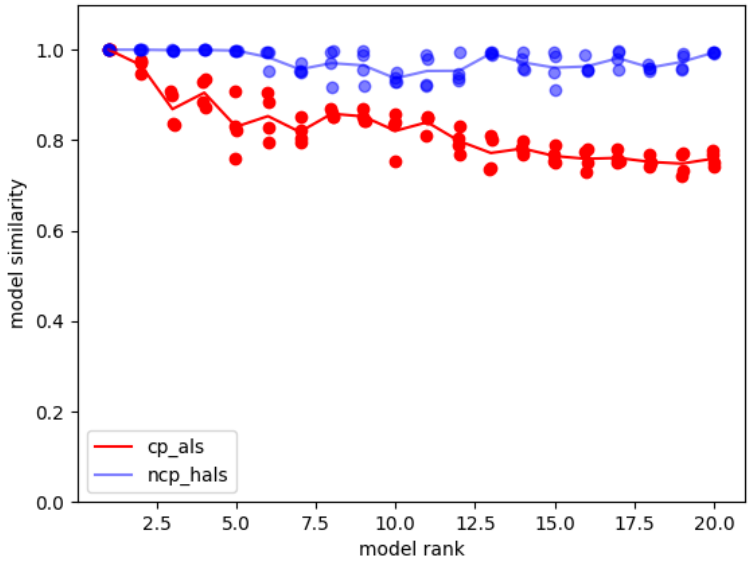

Figure A.4: Similarity score plot of each component number (model rank) between non-negative CP (blue) and standard CP decomposition (red), where rescaled data is fitted into non-negative CP and unnormalized data is fitted in standard CP. Both data version excluded 5% of low SNR cells. Each dots corresponds to different optimization run

A.2 Additional components of TCA models

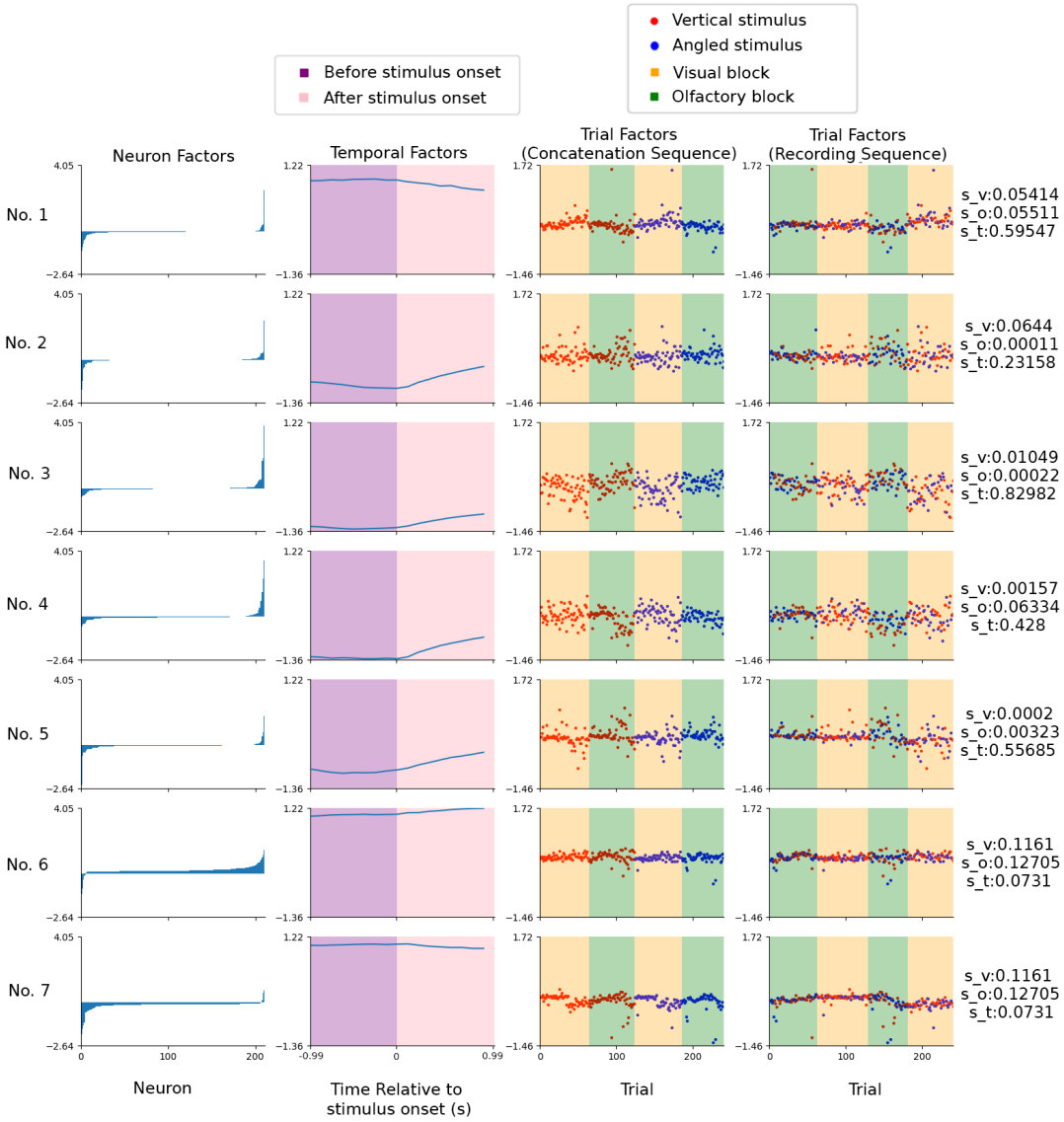

Figure A.5: The unnormalize with no outlier exclusion was fitted into standard TCA. The columns are presenting: neuron factor (first; left column), temporal factor (second column), and trial factor ordered by concatenation sequence (third column) and recording sequence (fourth column).

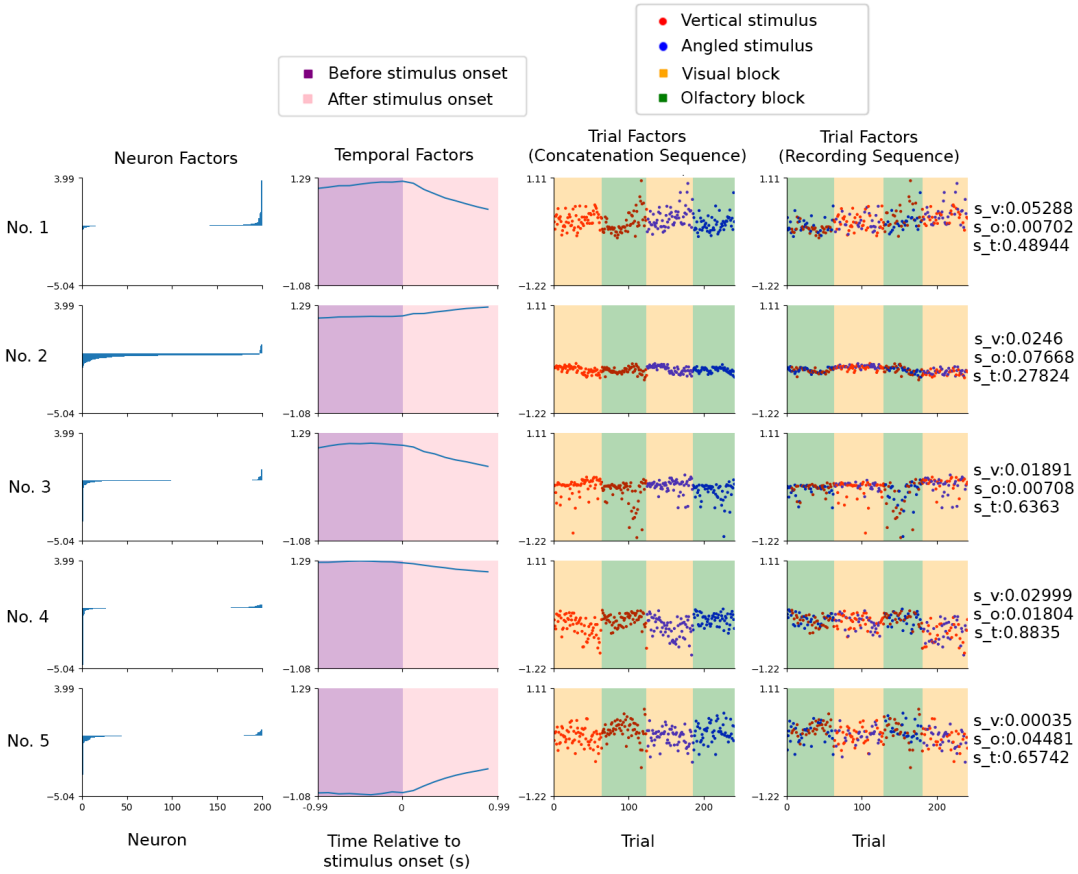

Figure A.6: The unnormalize data with 5% of low signal-to-noise (SNR) cells excluded was fitted into 5-components standard TCA.

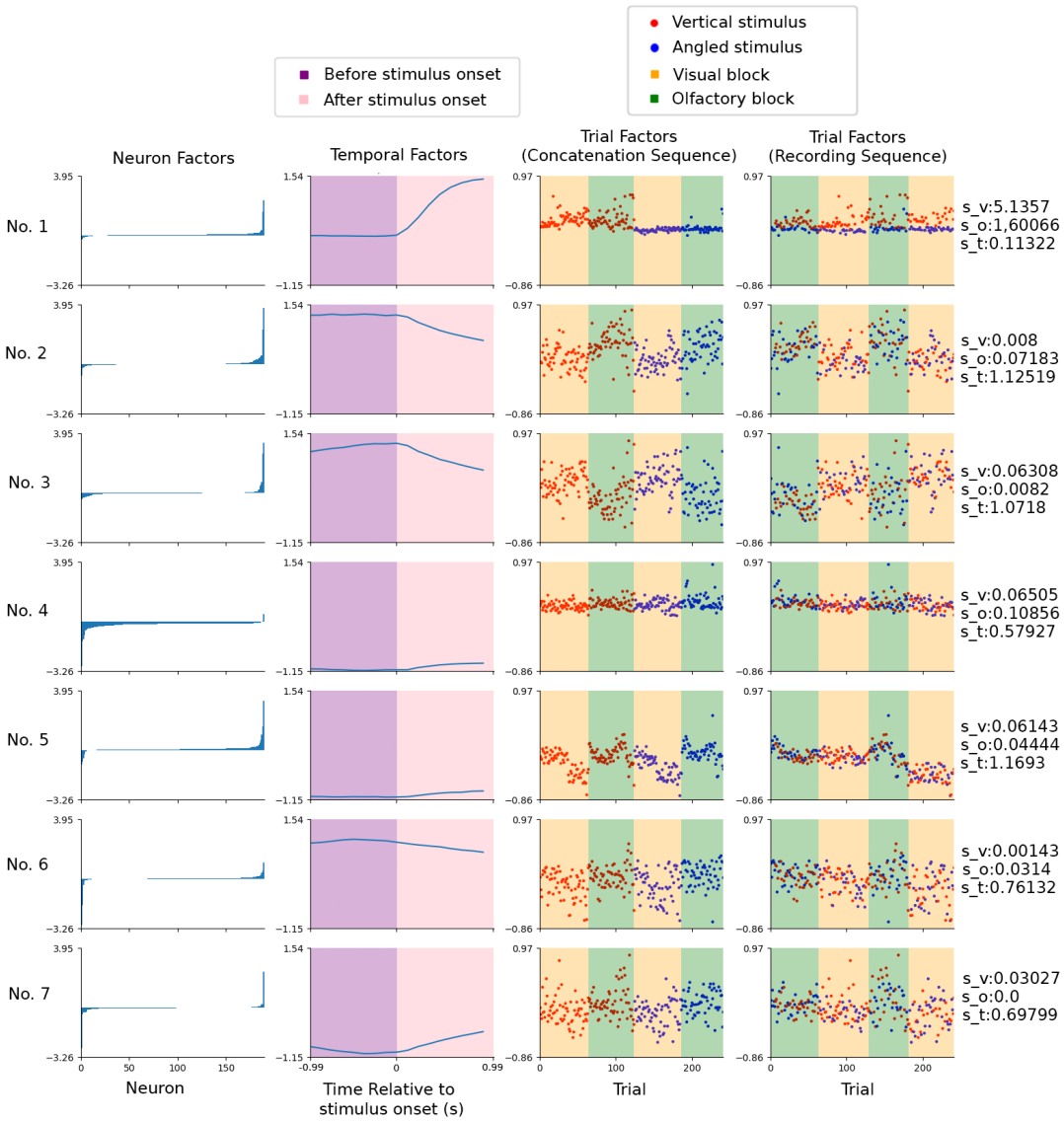

Figure A.7: The unnormalize data with 10% of low signal-to-noise (SNR) cells excluded was fitted into 7-components standard TCA.

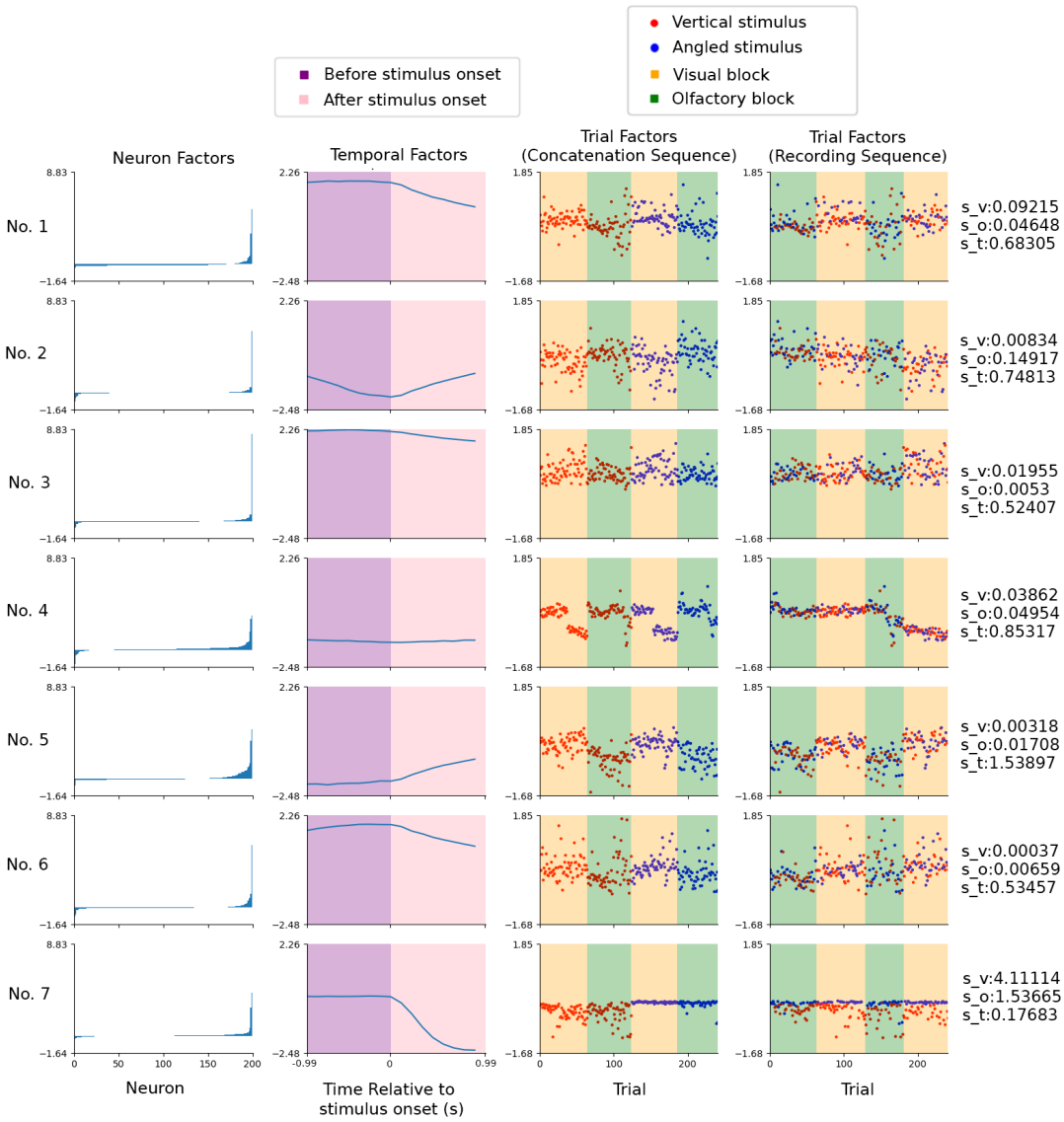

Figure A.8: The data normalized by z-score with 5% of low signal-to-noise (SNR) cells excluded was fitted into 7-components standard TCA.

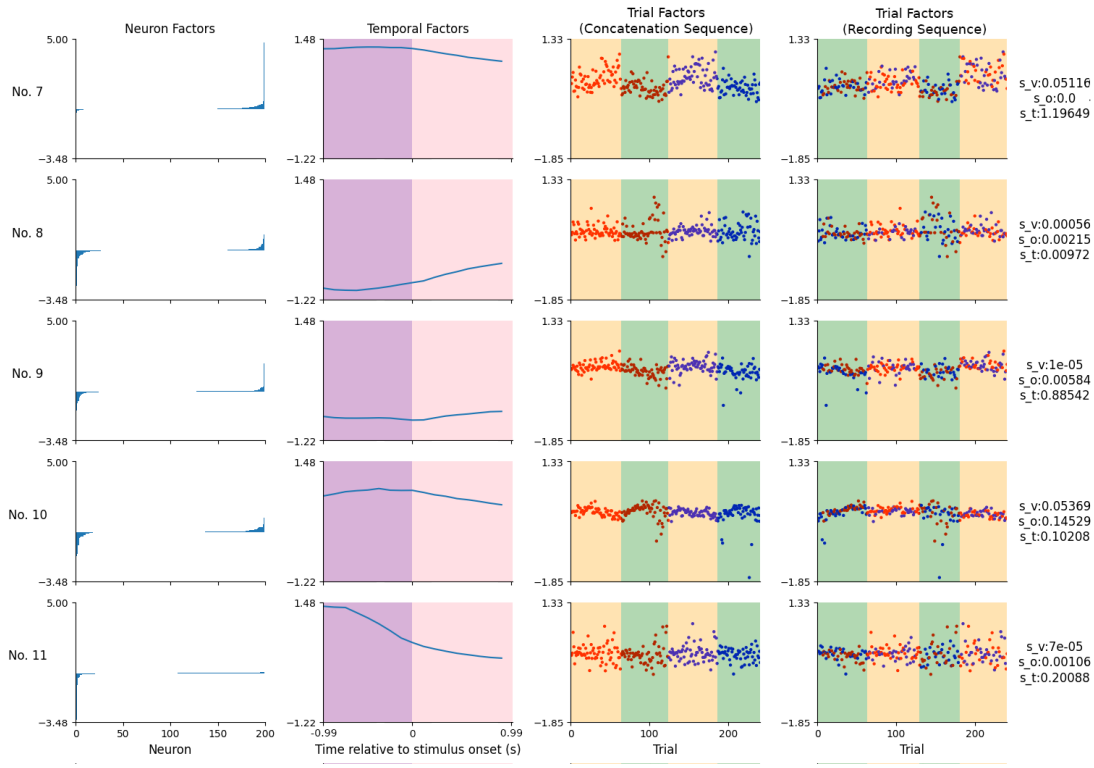

Figure A.9: Figure 4.6 remaining components from the decomposition of the Mouse *A* from the 13-component TCA model.

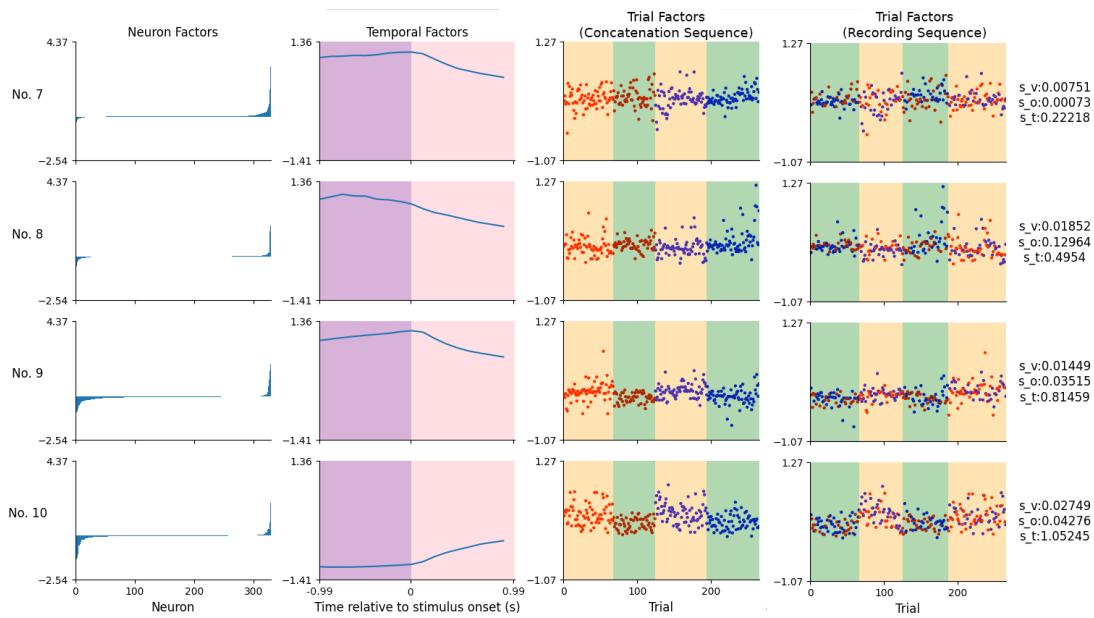

Figure A.10: Figure 4.7 remaining components from the decomposition of the Mouse *B* from the 10-component TCA model.

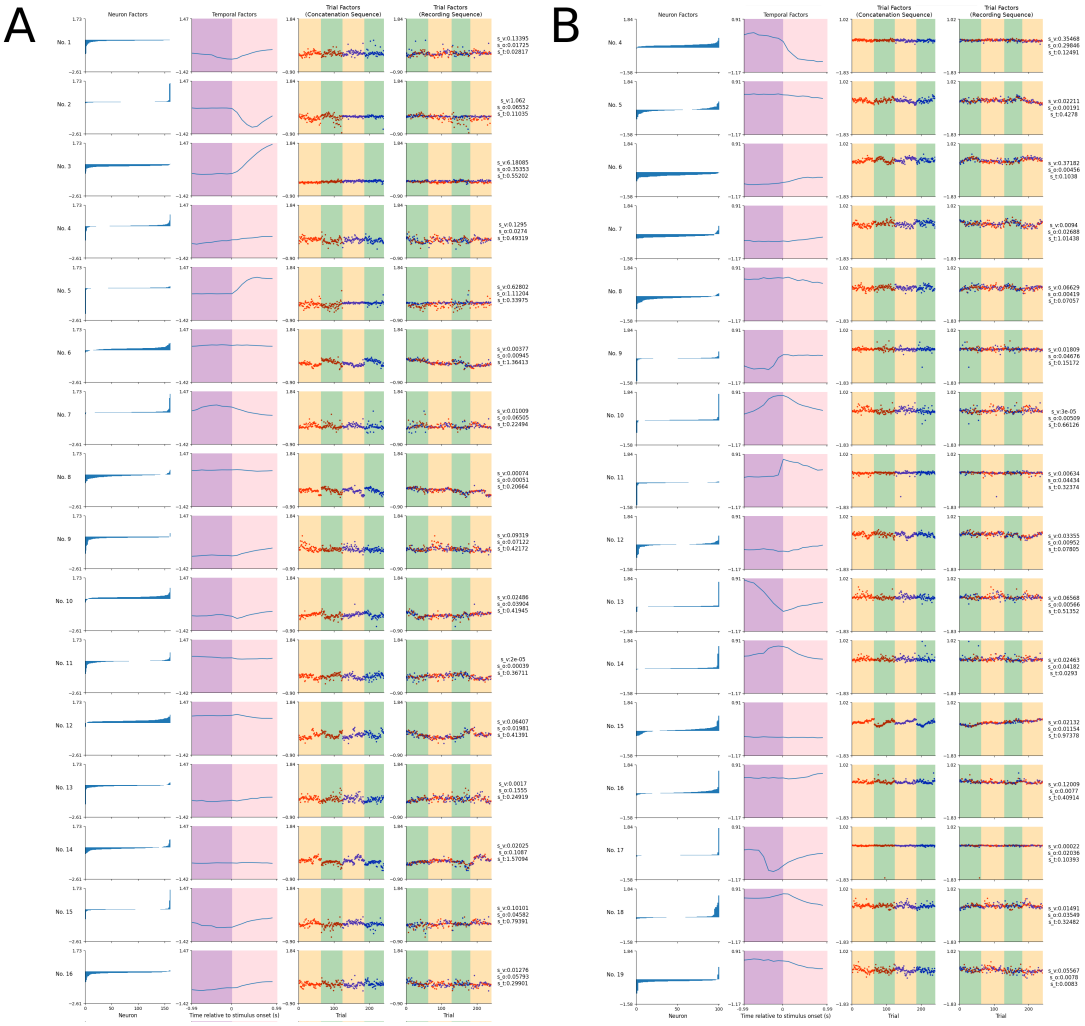

Figure A.11: Remaining components of decomposition after refitting of Mouse *A* from 19-component TCA model. (A) Figure 4.8 remaining refitting result after removing 20%. (B) Figure 4.9 remaining refitting result after removing 40%.

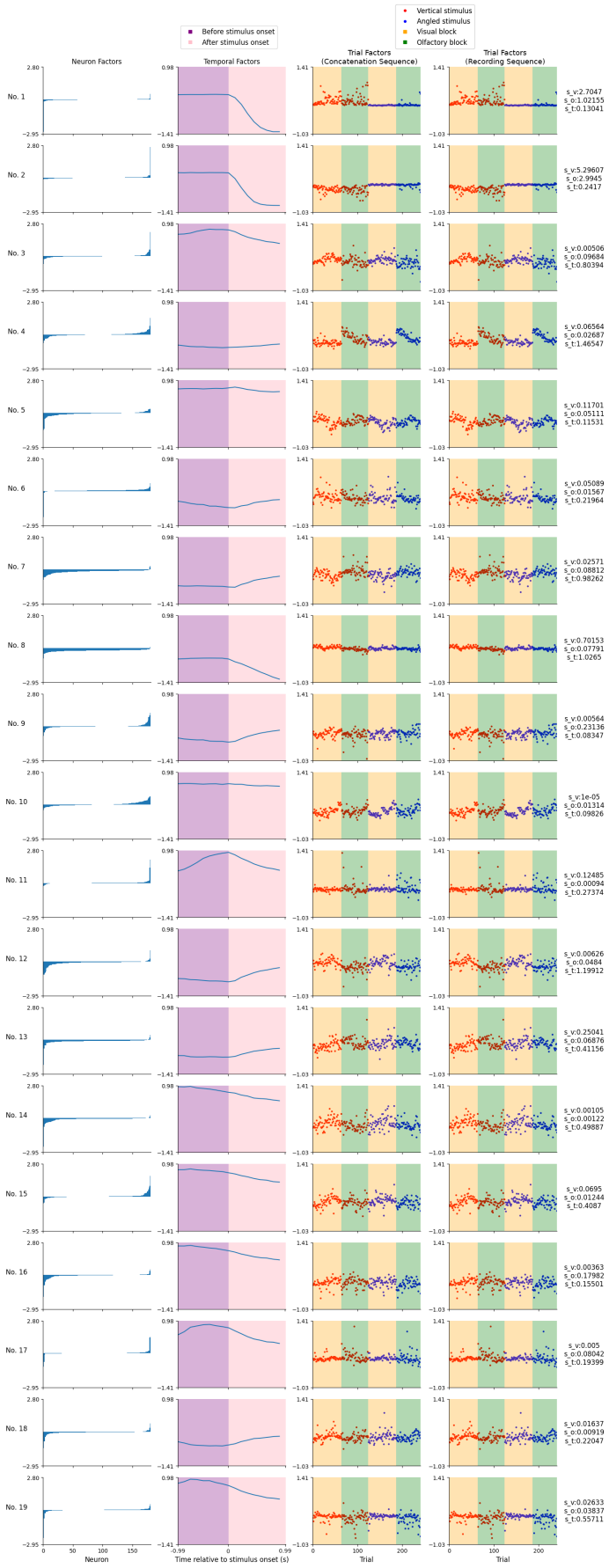

Figure A.12: Mouse *A* refitting 19-components results after removing 10% of cells.

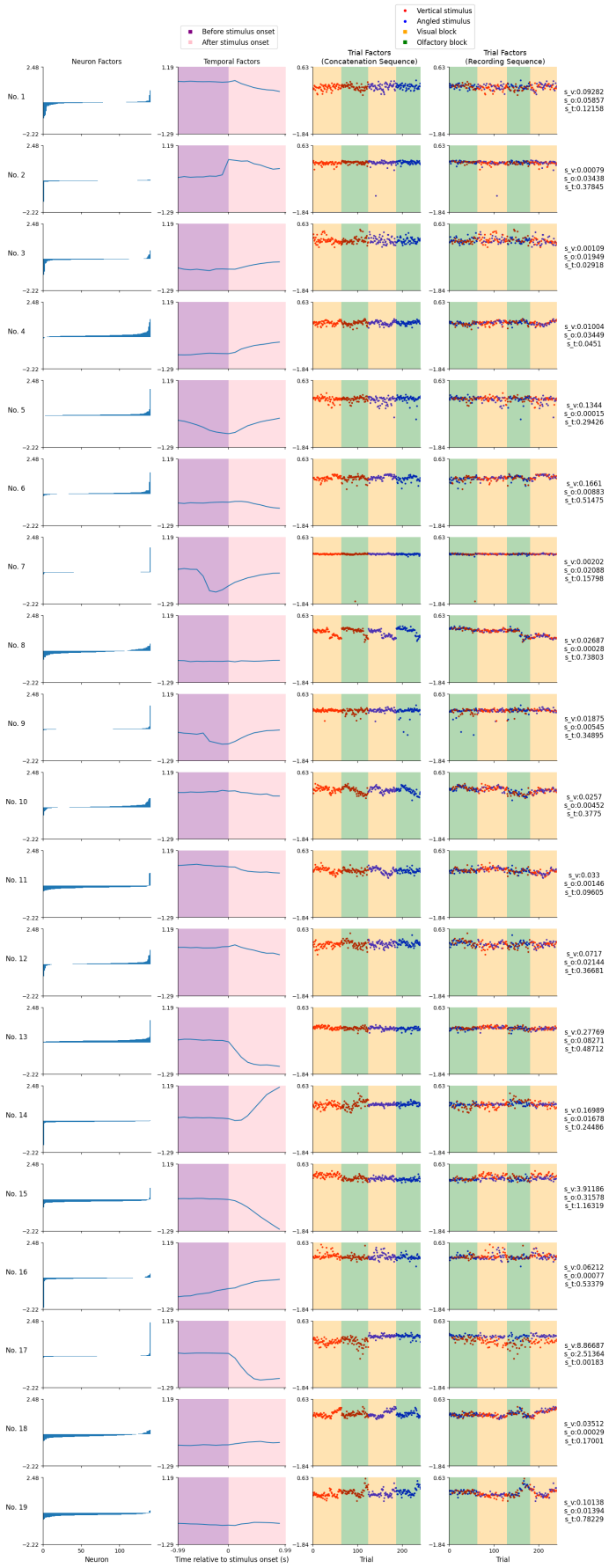

Figure A.13: Mouse A refitting 19-components results after removing 30% of cells.

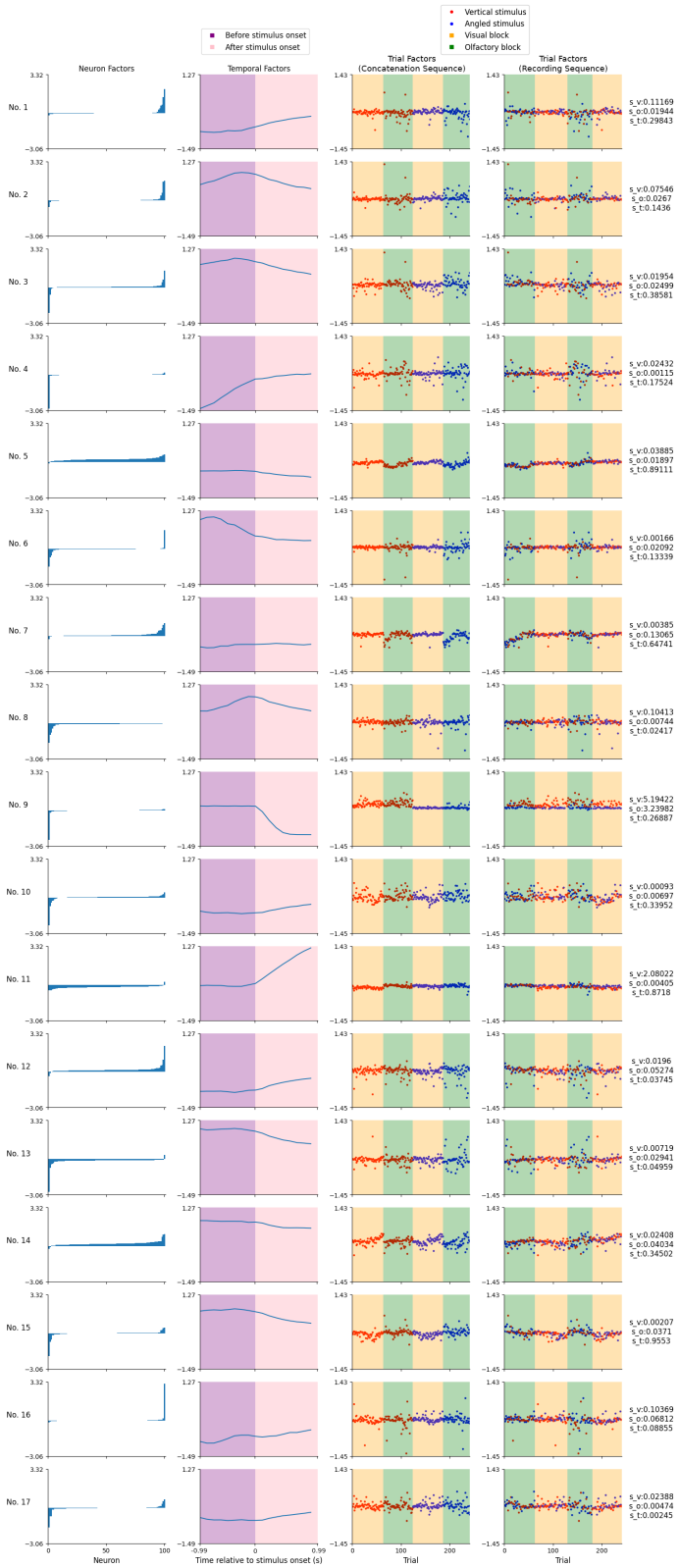

Figure A.14: Mouse *A* refitting 17-components results after removing 50% of cells.