Master ATIAM, IRCAM

Master's Thesis

Laboratoire Des Systèmes Perceptifs

# Task-relevant adaptation of phase-locked neuronal responses

*Author:*
Martin Averseng

*Supervisors:*
Pr. Shihab Shamma
Ph.D. Yves Boubenec

*Stimulus reconstruction in ferrets primary auditory cortex*

28/03/2015 - 23/08/2015

# Contents

## Acknowledgments

I would like to thank Pr. Shihab SHAMMA for supervising me during this internship. I also would like to thank Pr. Daniel PRESSNITZER for his help in finding an internship for me at the LSP. Thanks also to the team who recorded the data used in this work. Finally, I would like to express my great gratitude to Ph.D. Yves BOUBENEC for his very precious and patient help all along the internship, for spike-sorting all the recordings, for the valuable discussions we had on the data, for reviewing some 100 versions of this report, and for welcoming me so warmly at ENS !

## Abstract

Task engagement can rapidly shape neuronal responses to match behavioural requirements. Here we studied the neuronal encoding of simple sounds (periodical click trains) in the primary auditory cortex (A1) of ferrets trained on a simple task (low vs high rate). We recorded the neuronal responses in two conditions, when animals were passively listening to the stimuli or actively engaged in the task. We asked how the phase-locked component of the neuronal encoding was modified between these two conditions and whether changes observed were relevant to the task. We show that click trains were robustly encoded in neuronal responses through a phase-locked representation even at rates up to 30Hz. Somewhat surprisingly, when engaging in the task, this phase-locked component was substantially degraded. However, neuronal encoding exhibited task-relevant and behaviourally-driven stimulus adaptation in the active state: adaptation to click trains was decreased for target stimuli, indicative of an enhancement of the relevant stimulus in A1.

## Introduction

Studying changes induced by task-engagement in neuronal responses is crucial to understand how the brain processes its inputs to match environmental requirements. In simple go/no-go tasks, information transiting into the brain is converted from a faithful and unbiased encoding of the stimulus delivered by sensory receptors, to an abstract and task-relevant representation under the form of decisions. Different parts of the brain are involved in this processing. Primary sensory cortices mainly encode stimulus features, for example through tonotopic organization [**?**] and spectro-temporal selectivity [**?**] in the primary auditory cortex, or orientation selectivity in the primary visual cortex [**?**]. Upper areas progressively extract task-relevant information allowing to make decisions (higher auditory cortex [**?**], LIP [**?**], pre-frontal cortex (involved in short-term memory) [**?**, **?**], secondary somatosensory cortex [**?**]). However, plasticity occurs early in this treatment: task-relevant changes in neuronal responses have often been reported already in primary cortices. Early plasticity could be a useful mechanism to enhance the contrast between stimuli to be discriminated [**?**, **?**]. Here, we studied data from ferrets auditory cortex, a model that has been increasingly used in auditory studies, particularly in A1 [**?**, **?**, **?**] but also in higher areas [**?**]. Rapid plasticity has been shown to occur in ferrets primary auditory cortex in the spectral domain [**?**], but similar mechanisms have not been reported in the temporal domain. Here we used the same task as in [**?**], but with click trains instead of pure tones. We thus expected to observe analogue plasticity in the neuronal tunings during behaviour.

Stimulus adaptation, defined as the decay in the neuronal response to a periodical stimulus, is a salient property of primary cortices [**?**, **?**] and is thought to be linked to task-engagement: active states are characterised by lower adaptation than passive states [**?**]. We show that although overall phase-locked encoding of the neuronal responses was surprisingly decreased during task performance, adaptation was specifically reduced for target click trains, thus showing plasticity of the temporal representations in A1. In a go/no-go task, this decrease in adaptation results in an enhancement of the contrast between the neuronal representations of target and reference stimuli [**?**].

Neuronal data analysis was performed using the stimulus reconstruction (SR) method. SR has been successfully used in recent auditory studies [**?**, **?**, **?**] to analyse recorded neuronal responses. It performs a linear filtering on the neuronal data designed to project them back into the stimulus domain. That is, spikes are considered in terms of their correlations with a chosen stimulus representation, and not in terms of neuronal dynamics, or average spike rate in a window. SR is also an alternative to usual reverse correlation methods where a filtering is performed on the stimulus representation to predict the observed neuronal data. In the latter method, neuronal properties are deduced and compared using those forward filters [**?**]. With SR, one can compare reconstructed stimuli, which provide a simple visualization of the stimulus as it is encoded by the neuronal population. SR has been used to address the question of what features of a stimulus are faithfully encoded in a population and which stimulus representation best reflects the neuronal encoding [**?**]. Recently, SR also allowed to predict the audio stream attended by humans subjects in cocktail party environments, using electroencephalogram recordings [**?**]. Here, we propose a simple non-linear version of the model. It is equivalent to a one-

layer feed-forward artificial neural network and is designed to take into account the possible saturation of information in the neuronal population. Possible advantages over linear models are discussed.

# 1 Material and Methods

## 1.1 Experimental design

**Recording set-up:** We recorded neuronal responses in the primary auditory cortex of two awake adult female ferrets trained on a simple click train discrimination task. During recordings, animals were head-restrained to ensure stability. Neuronal activity was recorded using 24-site Plexon linear electrodes. Neuronal responses were spike-sorted using a High-dimensional cluster analysis with a Masked EM Algorithm by Shabnam N. Kadir, Dan F.M. Goodman and Kenneth D. Harris (KlustaKwik). This algorithm dissociates spikes from neurons located near the same recording site, according to spike dynamics. Neurons with average firing rates that did not exceed 2 spikes/s during at least one of the sounds were not included in the analysis. Neuronal data were processed as PeriStimulus Time Histograms (PSTH). Spikes timings in each trial were binned and aligned to the click train onset. To increase the precision in the SR method, we used PSTH with overlapping bins. Bin sizes and steps were chosen after several iterations of the analysis and set to the values that yielded the best convergence in the reconstruction filters. We used bins of 10 ms and steps of 1 ms.

**Structure of the experiment:** The task was composed of trials in which a reference click train was played a random number of repetitions (1-7) followed by a target click train with a different rate. Target click trains were used as a warning signal in the task. Reference click trains did not provide information about the position of target click trains and were thus irrelevant in the task. Each click train was preceded by a neutral sound and followed by a silence period of 800 ms. The first 400 ms were neutral and called "post-click silence". The last 400 ms were called go (resp. no-go) windows when they followed a reference (resp. target) click train. In the text, reference or target click train presentations are indifferently called blocks, since the term "trial" refers to a succession of 1 to 8 blocks. To keep classical notations however, we used the term single-trial instead of single-block. Each session was composed of a succession of 30 to 40 trials, with fixed reference and target rates. In every session, reference click trains were presented 150 times, and target click trains between 30 and 40 times. The structure of a trial is represented in Figure 1A.

Click train rates were separated in two categories, "low rates" (6Hz-16Hz) and "high rates"(20Hz-36Hz). The reference and target rates were always randomly chosen in distinct categories. For one ferret (Avocado), reference click trains were always low rates and target click trains, high rates. The other ferret (Lemon) learnt an opposite decision rule (see Figures 1B-C). We thus could assume that differences observed in the responses to click trains within the same category came from their identity in the task (reference or target) and not from differences between the two animals. With this assumption, we could efficiently decide whether task-engagement and stimulus identity had or not an influence on a given feature of the neuronal response. Data of other animals will be included in later work to ensure the validity of this assumption.

**Passive and engaged states:** Ferrets were water restrained prior to every experiment, and recordings were made successively in passive and engaged conditions. In the engaged condition, ferrets could lick a spout in front of their head. Licks were rewarded with water when they occurred during sound, post-click silence and go windows but punished with mild tail electric shocks during no-go windows. In the passive condition, the electrical apparatus was switched off and the water spout removed. No rewards nor punishment were provided to the animal: ferrets were left free to passively listen to the sounds, rest, and possibly sleep. We recorded 9 passive and 9 engaged sessions for both ferret.

| Reference/target | Neutral sound | Click train | Post-stim silence | Go/No-Go |
|---|---|---|---|---|
| Reference 1 |  |  |  | Go  |
| Reference 2 |  |  |  | Go  |
| Reference 3 |  |  |  | Go  |
| ... |  |  |  | Go  |
| Reference n ($1 \leq n \leq 7$) |  |  |  | Go  |
| Target |  |  |  | No-Go  |

**(A)** Structure of a trial

| Click train rates for Avocado (Hz) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Reference | 15 | 7 | 6 | 7 | 6 | 6 | 6 | 8 | 6 |
| Target | 30 | 36 | 32 | 26 | 24 | 36 | 36 | 33 | 28 |

**(B)** Rates of reference and target for first animal (Avocado)

| Click train rates for Lemon (Hz) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Reference | 32 | 36 | 30 | 20 | 26 | 20 | 26 | 28 | 28 |
| Target | 16 | 6 | 6 | 6 | 6 | 8 | 8 | 8 | 9 |

**(C)** Rates of reference and target for second animal (Lemon)

**Fig. 1** – (A) Structure of a trial. Each trial is a succession of several reference sounds followed by the target sound. Each sound is composed of an initial silence (400 ms, not represented), followed by a neutral sound (1.25 s), a click train (750 ms), a neutral post-click silence(400 ms), and a go/no-go window (400 ms of silence). A session is made of 30 to 40 such trials. (B) and (C) Table of rates (in Hz) for the first animal (Avocado, B), and the second animal (Lemon, C).

## 1.2 Stimulus Reconstruction

### 1.2.1 Presentation of the method

We used the stimulus reconstruction (SR) method to analyse experimental neuronal data. SR allows to quantify to what extent the timing information in a time-varying signal can faithfully be extracted from a multi-channel code. As in classical machine learning algorithms, a subset of the neuronal response (training set) is used to find the best possible decoder, from the vector of spikes timings to the signal. Then, reconstruction quality is assessed by comparing the proximity of the original stimulus to its reconstruction from a separate subset (test set). The usual approach for SR is the linear model in which a direct linear mapping [**?**, **?**] is assumed between the signal and the neuronal response. Let

$$S = \begin{pmatrix} s(0) \\ s(1) \\ \vdots \\ s(T) \end{pmatrix}$$

the input signal and $\forall i \in [|1, N|]$

$$\begin{pmatrix} r_i(0) \\ r_i(1) \\ \vdots \\ r_i(T) \\ \vdots \\ r_i(T+\tau) \end{pmatrix}$$

the i-th neuronal response, where $\tau$ is a positive integer called maximal lag in the SR method. The linear mapping assumption takes the form of the following model for the encoding of the stimulus:

$$s(t) = \sum_{i=1}^{N} \sum_{\delta=0}^{\tau} g_i(\delta) r_i(t+\delta) \tag{1}$$

Let

$$R_i = \begin{pmatrix} r_i(0) & r_i(1) & \cdots & r_i(\tau) \\ r_i(1) & r_i(2) & \cdots & r_i(1+\tau) \\ \vdots & \vdots & \vdots & \vdots \\ r_i(t) & r_i(t+1) & \cdots & r_i(t+\tau) \\ \vdots & \vdots & \vdots & \vdots \\ r_i(T) & r_i(T+1) & \cdots & r_i(T+\tau) \end{pmatrix}$$

the lagged neuronal response of neuron i, and

$$G_i = \begin{pmatrix} g_i(0) \\ g_i(1) \\ \vdots \\ g_i(\tau) \end{pmatrix}$$

the i-th filter matrix. Let

$$R = (R_1 \quad R_2 \quad ... \quad R_N)$$

and

$$G = \begin{pmatrix} G_1 \\ G_2 \\ \vdots \\ G_N \end{pmatrix}$$

Then equation (1) becomes

$$S = RG \tag{2}$$

The SR method finds the best fit $\hat{G}$ from the neuronal response $R$ to the original stimulus $S$ in a least square sense. $\hat{G}$ is called the reconstruction filter. Applying the reconstruction filter on new neuronal response $R'$ allows to map them to a list of reconstructed stimuli. When training filters, we separated the set of stimuli in two halves. A filter was trained on each half and tested on the other half. Thus, reconstruction filters for reference click trains (150 in each session) were trained on 75 stimuli while target reconstruction filters (30 to 40 in each session) only on 15 to 20 stimuli. Using same sizes for training set will be done in later work to remove bias due to better training on reference. However, please note that since the same procedure was applied in passive and engaged conditions, modulations from passive to active cannot be influenced by this difference, and consequently, our results are still valid.

The previous model allows to compute reconstructed stimuli based on the aggregation of responses from several neurons. It can also be observed that the reconstructed stimuli are the sum of the neuronal contributions as

$$S = \sum_{i=1}^{N} S_i$$

where

$$S_i = R_i G_i$$

This stated, it is therefore possible to isolate a reconstructed stimulus for each neuron, which is useful to quantify single-neurons properties based on those contributions. We defined a ranking of neurons among each session based on the relative importance of their contributions $S_i$ to the summed reconstructed stimulus $S$. In each session, the first neurons of the ranking were thus the most phase-locked neurons in the session. By restricting our analysis to those neurons, we could isolate the phase-locked component of the neuronal population response. In each session, 15 single-units were typically found among which 3 displayed robust phase-locking in the range of rates of the session. Unless stated otherwise, analysis using SR were restricted to the three most phase-locked neurons in each session (n=54 neurons).

### 1.2.2 Improving reconstruction by using single-trial information

It is common in neuroscience studies to average neuronal data over trials. This is done in order to uncover the underlying distribution of a pseudo-random spiking process, in a very variable inter-trial context. However, we will show here that taking separately the responses in each trial greatly improves the quality of the reconstruction filter.

It can be shown that given a set of responses $R^1, R^2, ..., R^K$, the matrix $\hat{G}$ that minimizes the error function

$$e_{mean}(\hat{G}) = ||S - \left(\frac{1}{K}\sum_{k=1}^{K} R^k\right)\hat{G}||^2 \tag{3}$$

is given by

$$\hat{G} = K\left((\sum_{k=1}^{K} R^k)^t \sum_{k=1}^{K} R^k\right)^{-1}\left(\sum_{k=1}^{K}(R^k)^t\right) S \tag{4}$$

Similarly, the minimum of the error function

$$e_{single}(\hat{G}) = \sum_{k=1}^{K} ||S - R^k\hat{G}||^2 \tag{5}$$

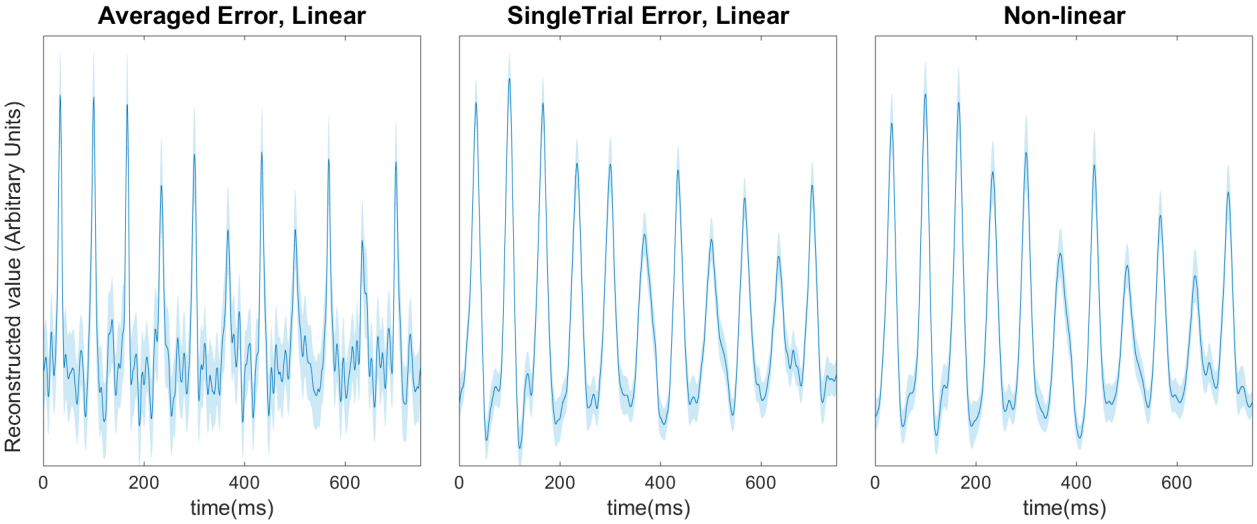
is met by

$$\hat{G} = \left(\sum_{k=1}^{K}(R^k)^t R^k\right)^{-1}\left(\sum_{k=1}^{K}(R^k)^t\right) S \tag{6}$$

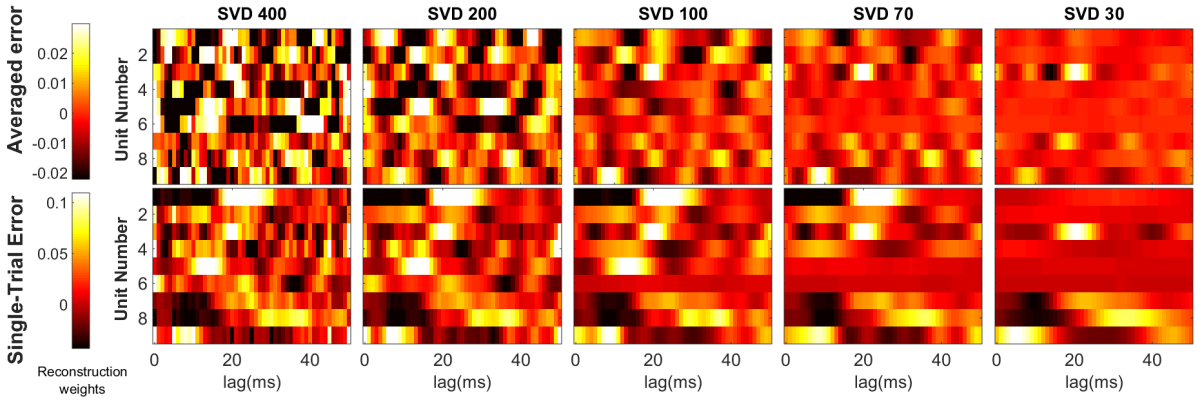
Due to the stochastic nature of the responses, inversions in the previous formula can always be performed. In the first case, errors in the single-trial reconstructions $R^k\hat{G}$ can cancel each other. This generally leads to poor convergence in the filter weights (seen as big jumps in the weights of the filter). Oppositely, the second error function prevents errors to cancel, yielding sparser and smoother filters, and less noisy reconstructed stimuli as can be observed in Figure 2A (left two panels). We call averaged error model the model using error function $e_{mean}$ in equation (3) and single-trial error model the model using error $e_{single}$ function in equation (5). All analysis in the present work was performed using the single-trial error model.

### 1.2.3 Optimal prior and flat prior reconstructions

The method we are using is referred to in [**?**] as the optimal prior SR, as opposed to the flat prior method. Shortly, optimal prior SR makes optimal use of the inner stimulus correlations to enhance fit quality. A stimulus only partly encoded in neuronal responses could be fully decoded in this framework due to information redundancy in the stimulus. This prior knowledge of the stimulus correlations could be encoded somewhere else in the brain (perhaps in superior areas) allowing the neuronal encoding to be more parsimonious (this is actually thought to be an important role of adaptation, [**?**]). However, it could also well be that prior actually implemented in the brain be not optimal, in which case




**(A)** Average reconstructed click trains for the three methods (linear averaged error and single trial error, non-linear).




**(B)** Effect of de-noising on reconstruction filters.

**Fig. 2 – Comparison of reconstruction models** (A) Average reconstructed low-rate (15 Hz) click train of an example session using from left to right: linear averaged error model, linear single-trial error model, non-linear model. Non-linear and single-trial linear models both achieve better generalisation than linear averaged error model. (B) Reconstruction filters obtained in the averaged error model (top row) and in the single-trial error model (bottom row). Each column corresponds to a given level of de-noising in the singular-value decomposition (SVD), from left to right with increasing level of de-noising (decreasing rank in the SVD). In each box, a row corresponds to one recorded neuron and colors represent the weights of the reconstruction filter at the corresponding lag. Single-trial error model yields smoother and sparser filters than averaged error model. It can also be seen that too much de-noising removes the contribution of the top neuron in the single-trial model, despite the fact that this neuron seems to be reliably encoding the stimulus (based on the weights obtained for higher ranks).

reconstructed stimuli could over-estimate the quality of the stimulus encoding. On the other hand, flat prior SR, not described here, does not assume inner correlations in the stimuli. In our analysis, we used optimal prior SR because of the quantity of correlations present in click trains. This represented an appreciable increase of reconstructed stimuli usability. However, optimally reconstructed stimuli cannot be compared when the original stimulus contains different inner correlations as is the case for our low rate and high rate click trains. This is why direct comparison between reconstructed click trains of different rates is not done in our study.

### 1.2.4 Taking baseline activity into account

For the majority of the neurons recorded, spikes do not occur only during stimulus, but a baseline firing rate, constant over the experiment, even during silence periods, is observed. This means that some spikes might have to be ignored when trying to decode the stimulus since they seem not to be informative of stimulus presence. To take this into account, a constant term can be added to the model in equation (1) as

$$s(t) = \sum_{i=1}^{N} \sum_{\delta=0}^{\tau} g_i(\delta)[r_i(t+\delta) - b_i]$$

The vector of parameters $(\hat{G}, b)$ where $b = (b_1, b_2, ...b_N)$ that minimizes the error functions shown previously is never unique. In both the averaged error and single-trial error models, the optimal vector $(b_1, b_2, ..., b_N)$ can be chosen in an affine hyperplane orthogonal to the vector

$$(\bar{G}_1, \bar{G}_2, ..., \bar{G}_N)$$

where

$$\bar{G}_i = \sum_{\delta=1}^{\tau} g_i(\delta)$$

Meaning that the set of baselines $b_i$ can be replaced by only one parameter

$$\hat{B} = \sum_{i=1}^{N} b_i \bar{G}_i$$

and the error models can then be replaced by:

$$e(\hat{G}, \hat{B}) = ||S - \left( \frac{1}{K} \sum_{k=1}^{K} R^k \right) \hat{G} + \hat{B}||^2$$

for the averaged error model and

$$e(\hat{G}, \hat{B}) = \sum_{k=1}^{K} ||S - R^k \hat{G} + \hat{B}||^2$$

for the single-trial error model, where the parameters $(\hat{G}, \hat{B})$ can be found analytically. Once the parameter $\hat{B}$ that minimizes the error is known, it is not possible to retrieve uniquely the original model parameters $(b_1, b_2, ..., b_N)$, because $\hat{B}$ only reflects a summed baseline activity. When studying single-neuron reconstructions, this can be problematic since now separating the contributions of each neuron in the constant $\hat{B}$ is not possible. This is of special importance because one of the most visible effects of task-engagement in the neuronal response is a robust increase of baseline activity. We thus wanted to compare the passive and engaged responses once this baseline had been removed, if possible. In the present analysis, measured baselines were not included in the model by lack of time. To overcome the flaws of having only a global baseline, we estimated a baseline *a posteriori* for each neuron, in order to remove bias in the analysis when comparing neurons. The estimated baselines were chosen so that the reconstructed click trains had zero mean over time.

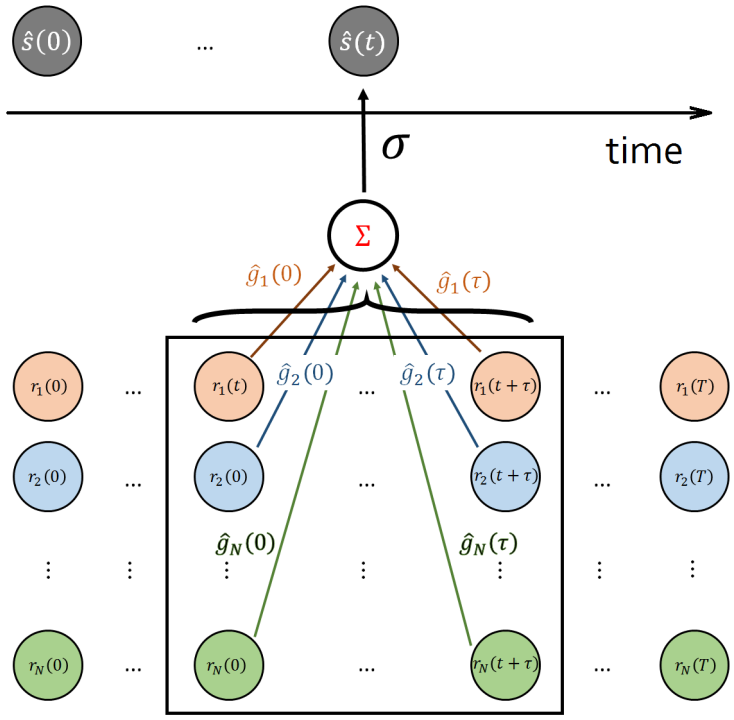



**Fig. 3** – Non linear reconstruction is equivalent to a sliding single-layer feed-forward neural network using the neuronal response as an input. The reconstructed stimulus at time $t$ is the output of the network using responses spanning from $t$ to $t+\tau$.

### 1.2.5 Non-linear reconstruction

Linear SR models assume a direct linear mapping between the stimulus and the response. Here we investigated the possibility of modifying the error function to include a non-linearity. We wanted to model the fact that neuronal population can saturate in information (see *Discussion*). Our non-linear model is given by

$$s(t) = \sigma\left(\sum_{i=1}^{N}\sum_{\delta=0}^{\tau} g_i(\delta) r_i(t+\delta)\right) \tag{7}$$

Where $\sigma$ is the usual sigmoid function, defined by

$$\sigma(x) = \frac{1}{1+\exp(-x)}$$

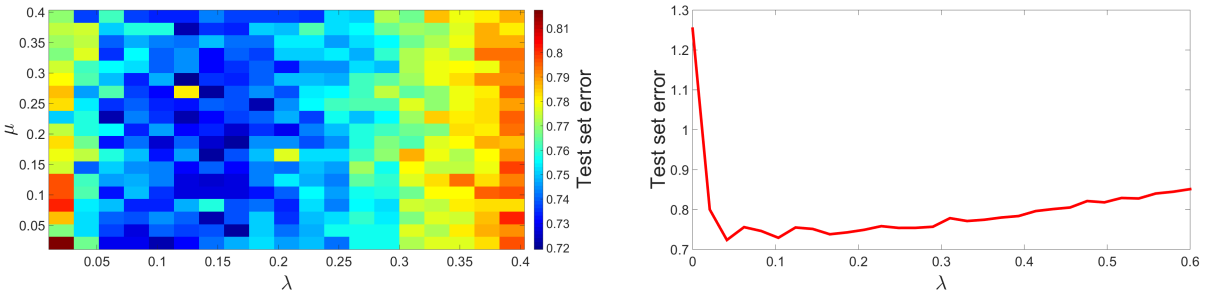
Weights can be fitted by gradient descent on a training set and be used to compute the reconstructed stimuli on a test set.

This model is equivalent to a simple single-layer feed-forward neural network (Figure 3). Taking the baseline into account as discussed before would be equivalent to adding a bias on our network.

Comparing linear with non-linear SR has not yet been done completely. Some elements are given in the *Results* section. All results presented in this work that make use of the SR method have been obtained with the linear model, but the qualitative results also hold when using non-linear SR.

### 1.2.6 Over-fitting

Both linear and non-linear models are subject to over-fitting, meaning that the trained filter yields accurate reconstructions on the training set but generalizes poorly on the test set. To avoid this behaviour, different strategies can be adopted. In the linear model, a singular-value decomposition (SVD) can be performed in the inversion in equations (4, 6), see Figure 2. This method requires the user to estimate the adequate noise level. However, a similar processing is not possible for non-linear




**(A)** Test set error as a function of $\lambda$ and $\mu$ **(B)** Test set error for $\mu = 0$

**Fig. 4** – Over-fitting in the non-linear model. A filter was trained on an example session for different parameters of $\lambda$ and $\mu$ in equation (8), and over-fitting was estimated as the mean square error between reconstructed and original stimuli in the test set. (A) Evolution of the test set error as a function of $\lambda$ and $\mu$. $\lambda$ is the most influential parameter. (B) Test set error for $\mu = 0$ plotted as a function of $\lambda$. The high value for $\lambda = \mu = 0$ shows the importance of penalization.

models. An alternative possibility is to add terms in the error function to penalize filters that are not sparse or smooth. The error function becomes:

$$e_1(\hat{G}) = e(\hat{G}) + \lambda \sum_{i=1}^{N} \sum_{\delta=0}^{\tau} |g_i(\delta)|^2 + \mu \sum_{i=1}^{N} \sum_{\delta=0}^{\tau-1} |g_i(\delta+1) - g_i(\delta)|^2 \tag{8}$$

that can still be solved analytically for linear models and by gradient descent in the non-linear models. We used SVD for linear SR due to conditioning problems when trying to solve analytically the previous equation, whereas penalization was used for non-linear SR. It is noteworthy that penalization neatly improved the non-linear method even for small values of $\lambda$ and $\mu$, as shown on Figure 4. In future work, we hope to overcome the troubles we had applying penalization on the linear model in order to provide an objective way to compare the two models (since they will use the same error function).

### 1.2.7 Passive and optimal encoding

We used the SR method in order to compare features of the neuronal response in the passive and engaged state. There are two ways of comparing the responses.

First, one can use the passive reconstruction filter to decode both the response of the passive and the engaged session. Suppose the neuronal response $R$ is linked to a stimulus $S$ by $R = HS$ where $H$ is a transfer function. Changes in the transfer function due to task engagement will cause changes in the response according to $\Delta R = \Delta HS$. This change is projected back to stimulus domain [**?**]: we assume that the transfer function $H$ does not change and instead compute the change in the stimulus $\Delta S$ that would cause the same change in the response. This can be written as $\Delta R = H\Delta S$. Since H is the forward mapping from the stimulus to the neuronal response, one has simply $G = H^{-1}$. Using the reconstruction filter $\hat{G}$ measured in the passive condition, one has approximately $\Delta S = \hat{G}\Delta R$. Therefore, comparing the reconstructed stimuli using the passive filter on both passive and active data allows to see the changes in the neuronal tunings in the stimulus domain.

Within this approach, it is not possible to quantify the quality of encoding across the two conditions. Indeed since the encoding $H$ is likely to change, reconstructing response in the engaged condition with passive filter will result in lower reconstruction accuracy than using the engaged filter. To compare the quality of encoding across the two conditions, it is necessary to use matching filters for each session.

The first approach will be denoted by "passive SR" since it decodes the response using only the passive reconstruction filter. The second will be denoted by "optimal SR" (here optimal does not refer to prior knowledge on the stimulus). Passive SR highlights differences in the neuronal response: differences observed in the reconstructed stimuli directly reflect differences in the neuronal tuning. Optimal SR allows to compare reconstruction accuracy in the two conditions, and thus provide a reliable measure of phase-locking.

### 1.2.8 Model and parameters used in the study

Since running SR over all sessions is computationally costly, we were not able to exhaustively explore the space of parameters. However, on every set of parameters, the general results exposed globally held, including those assessed with non-linear models. The parameters for which results are displayed here were chosen based on the global quality of reconstructed stimuli. We used the linear model to quantify all results, since comparison between non-linear and linear models is still going, and because it is the most classical approach. A comparison of linear and non-linear models is also discussed in more details later. The parameters explored for the two models were

- The maximal lag $\tau$ admitted in the models (1) and (7). This value reflects the time after an instant $t$ over which neuronal response is integrated to choose the reconstructed value at $t$. We used values of $\tau$ from 50 to 90 ms, according to observations on neuron dynamics. We retained 70ms.

- The SVD rank for linear SR. We compared values ranging from 50 to 90 and retained 70.

- The bin sizes and gaps in the PSTH when conditioning neuronal responses. We chose 10ms bins among 5, 10 and 15ms. The bin gaps retained were 1 ms, rather because lowest gaps would cause memory crashes on the machine.

- The smoothing and sparseness parameters for non-linear SR. We did not explore those parameters on all sessions (Figure 4 only shows over-fitting for one session due to computational cost). Values retained were $\lambda = 1$ and $\mu = 1$, which might be quite far from optimum (Figure 4 suggests using lower values).

Prior to this exploration, some tests had already been performed in order to reduce the window of exploration for each parameters.

## 1.3 Measures on reconstructed data

### 1.3.1 Measuring reconstruction accuracy

We defined several measures to quantify features of the neuronal response as observed in the stimulus domain. For each click in a click train, measures werre defined as follows:

Peak value: given a reconstructed click train, we defined the peak value as the highest value in the reconstruction of a click.

Reactivity: We wanted to quantify the reaction time of neurons between the two conditions. A direct comparison between the timing of the reconstructed and original click was not suited for the use of modulation index since it is not positive. We thus defined reactivity as follows

$$r(\hat{t}) = \exp\left(\frac{t - \hat{t}}{\tau_0}\right)$$

where $\hat{t}$ and $t$ are the timing of the peak value of the reconstructed click and of the original click respectively, and $\tau_0$ is a scaling parameter that was set to 3 ms. In the optimal encoding, reactivity is expected to be close to 1 since the delay of the neuronal responses is accounted for in the reconstruction filters, (thus $\hat{t}$ should be close to $t$ in average). However, when using the passive encoding to reconstruct the response in the engaged condition, values of the reactivity could potentially differ from 1, which would indicate a change in the neuron tunings during behaviour.

Width: We defined a width measure as a custom full width at half maximum of a reconstructed click. Shortly, width was estimated as the half of the time during which a click can be separated from silence with 70% confidence, based on the maxima and minima of the reconstructed click train.

Reconstruction error: We defined reconstruction error as the error $e(\hat{G})$ in equation (8) on the test set. This amounts to the mean square difference between the original and reconstructed click trains.

Values obtained were systematically ignored for all 4 measures when at least one of them was unrealistic (more than 25 ms between reconstruction peak and decoded click, more than 60 ms width), as sometimes occurred for poorly phase-locked neurons. All measures were computed for the block-averaged reconstruction of each click in a click train.

#### 1.3.2 Measuring adaptation

Finally, we studied adaptation, defined as the decay of the evoked neuronal response to a periodical stimulus. For a decay of the form $\exp(\lambda t)$, adaptation (resp. facilitation) is occurring if $\lambda < 0$ (resp. $\lambda > 0$). We defined the adaptation index as $a(\lambda) = \exp(-\lambda)$. Values above 1 of the adaptation index indicate that adaptation is occurring whereas values below 1 indicate facilitation. We estimated the decay rate $\lambda$ by fitting an exponential decay to the shape of the neuronal response to click trains.

#### 1.3.3 Modulation index

Results obtained for passive and engaged sessions were compared using modulation index. For a given positive measure $m$ and the values $m_p$ and $m_e$ taken by this measure in passive and engaged conditions respectively, the modulation index of this measure is defined as

$$\mathrm{MI}(m) = \frac{m_e - m_p}{m_e + m_p}$$

One has $-1 \leq \mathrm{MI}(m) \leq 1$. Modulation index were useful because the mean value of a measure could vary greatly among sessions, and the normalization by the sum was preferable.

#### 1.3.4 Statistical analysis

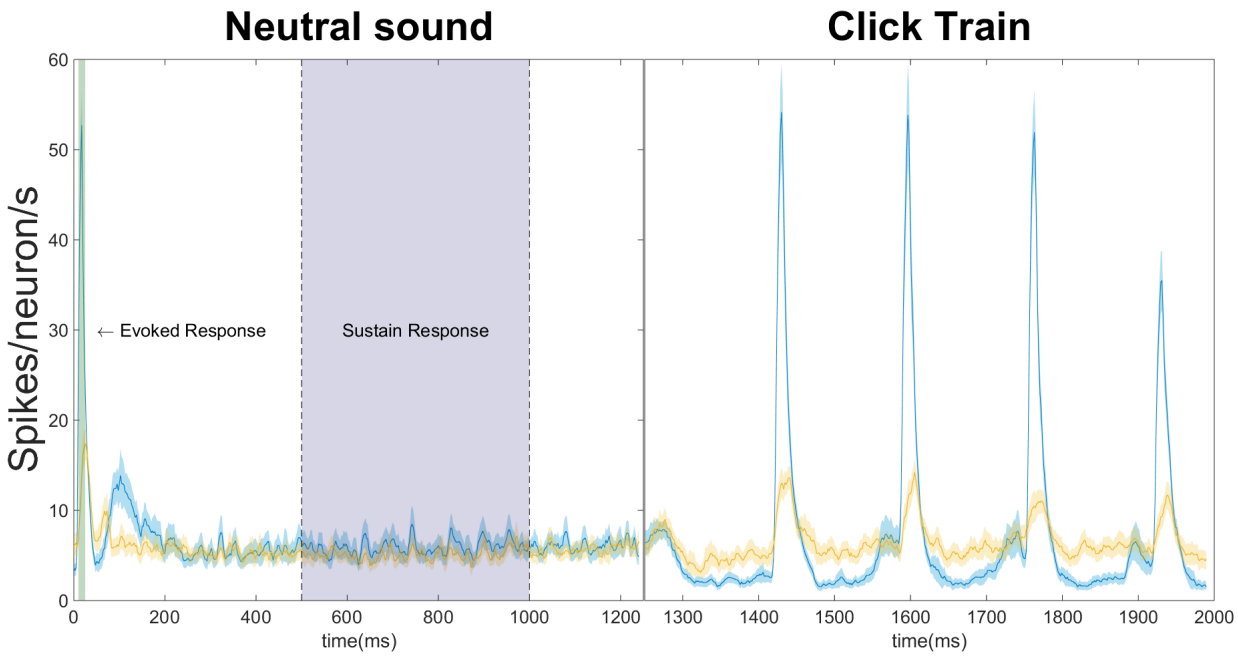
Significance of modulations from passive to engaged state were assessed using paired t-tests. The significance of the difference of amplitude of the modulations between several conditions (for example reference vs. target) was assessed using N-way analysis of variance. Statistical tests were performed with MATLAB and Statistics Toolbox Release 2014b. In the text, confidence intervals are given at the 95% level of significance, and are represented by black error bars or shaded areas in all Figures.

## 2 Results

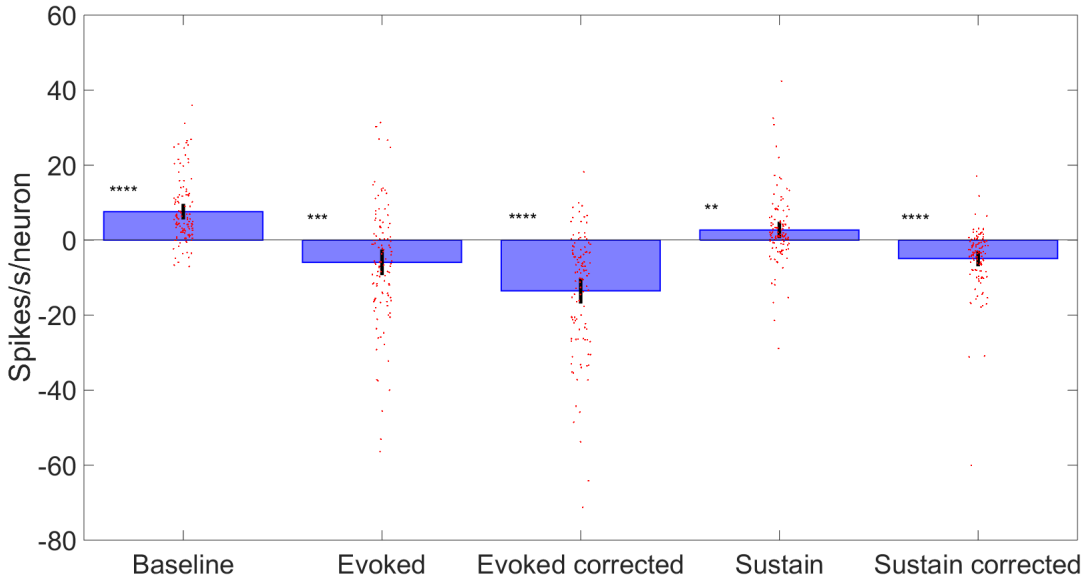
### 2.1 Phase-locked encoding is degraded during behaviour

Engaging in a task has a great impact on neuronal responses. Here we describe the changes that occur during behaviour. We show that a global degradation of the phase-locked component of the neuronal signal is observed.

**Spontaneous, evoked and sustained firing rate:** The most striking effect in the neuronal response when engaging in the task is a robust increase of spontaneous firing rates by an average 7.6 ± 1.6Hz, an increase by 56% ± 12% of the mean passive value ($p < 5e\text{-}15$, paired t-test), throughout all the session, as visible in Figures 5A,5B for an example session. Moreover, responses evoked by stimulus onsets were suppressed during behaviour (− 5.9Hz ± 3.0Hz, -11% ± 5.4%, $p < 5e\text{-}4$, paired t-test), therefore leading to a smaller difference between evoked and baseline firing rates (- 13.5Hz ± 2.9Hz, −32% ± 7.1 %, $p < 5e\text{-}15$, paired t-test). It is visible in Figure 5A that onsets on clicks also followed the same tendency. Finally, sustained firing rates during sounds were enhanced (+ 2.7Hz, ± 1.7Hz, +11% ± 7.6%, $p = 0.0025$, paired t-test) but suppressed once corrected by the baseline firing rate (− 4.9Hz, ± 1.6Hz, −52% ± 17%, $p < 5e\text{-}8$, paired t-test). Since evoked and sustained firing rates were estimated on the neuronal responses to the pre-click neutral stimulus, differences observed




**(A)** PSTH of Avocado's reference blocks in an example session (clicks: low rate = 6Hz)

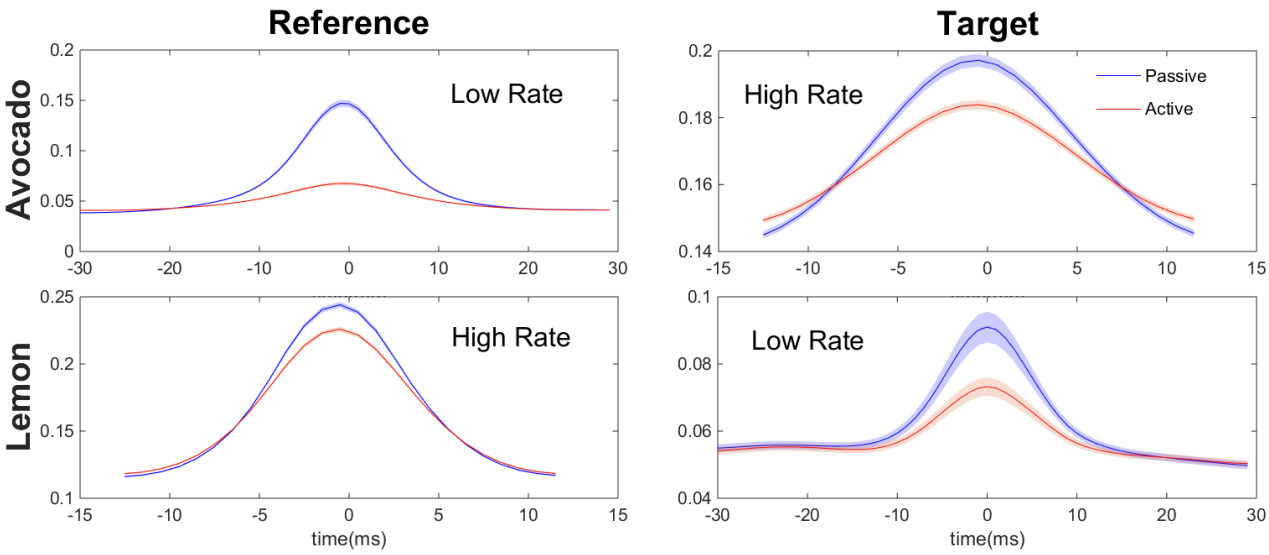



**(B)** Modulation of the neuronal signatures by task-engagement

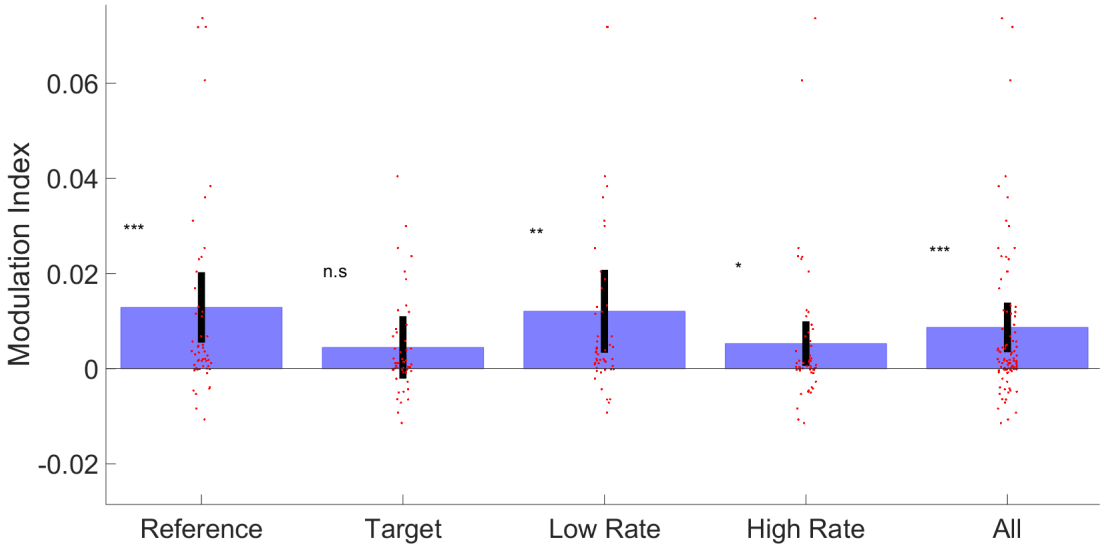
**Fig. 5 – Degradation of phase-locked signal during behaviour**. Engaging in the task is accompanied by an overall deterioration of phase-locked encoding of stimuli in A1. (A) PSTH of Avocado's neuronal response to the neutral stimulus (left) and reference - low rate - click train (right), in passive and active conditions (blue and yellow respectively), for an example session. Areas in which Evoked firing rate and sustained firing rate were measured have been highlighted in the figure. The baseline activity is higher in the engaged condition whereas evoked responses are lower, leading to a degraded apparent signal to noise ratio in the phase-locked encoding. (B) Modulation of the baseline firing rate, the evoked firing rate, the corrected evoked response (evoked - baseline) the sustained firing rate and the corrected sustained firing rate (sustained - baseline).

reflected a global change in the brain state, and therefore such a change was also expected to affect neuronal responses to clicks.

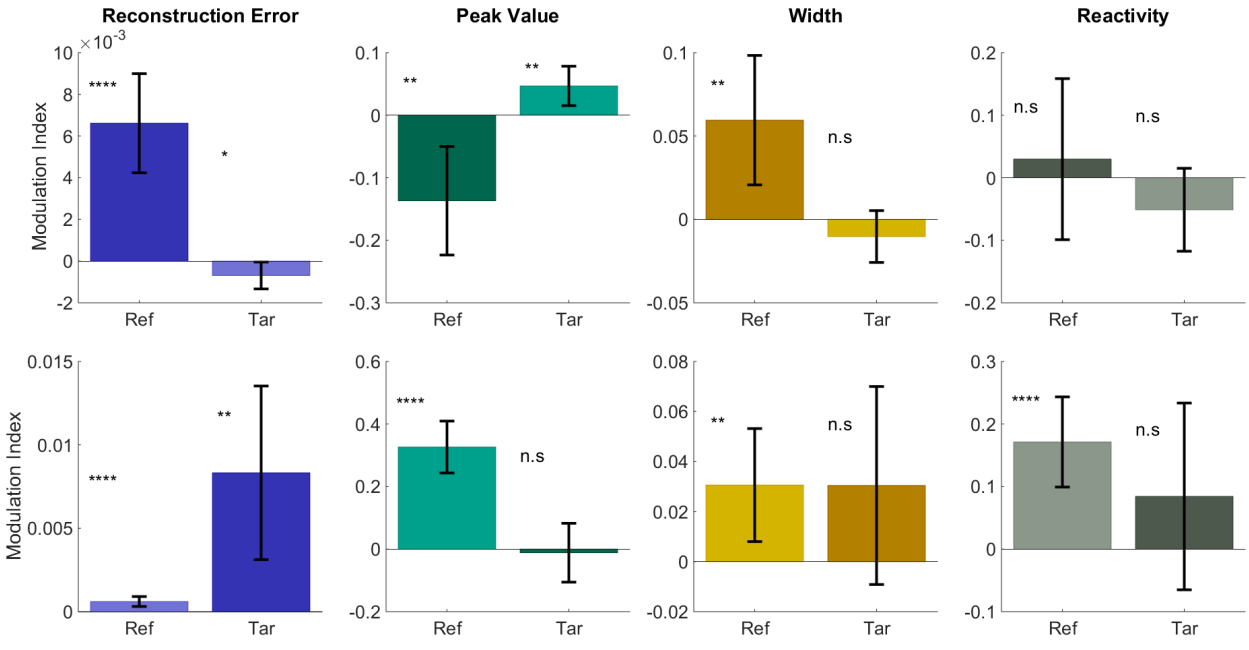
For each neuron, spontaneous firing rate was quantified by measuring the average firing rate during a 400 ms silence window before each click train, averaging over all blocks in all trials. Evoked firing rate was measured as the mean spike rate during a 15 ms window located 10 ms after the neutral stimulus onset. Sustained firing rate was measured as the mean firing rate during a 500 ms window




**(A)** Average click reconstructions




**(B)** Modulation of reconstruction error




**(C)** Modulation of reconstruction measures in passive encoding

**Fig. 6 – Decreased phase-locking and differential enhancement of target versus reference during behaviour** (A) Average reconstructed reference and target click for each animal, in passive (blue) and engaged (red) conditions, in the optimal encoding (passive filter for passive session, engaged filter for engaged session). The peak value is lower in the engaged condition revealing a degraded signal to noise ratio in the encoding. Reconstruction was performed with all neurons of each session and averaged over sessions. (B) Modulation of the reconstruction error for clicks regrouped by reference and target and by low and high rate. Reconstruction error is measured in the optimal encoding. The overall reconstruction error is significantly increased. The lower modulation for target click trains suggests an effect of task-engagement on reconstruction error, but this effect is not significant (N-way ANOVA, ref/tar: p=0.0633, low rate/high rate: p = 0.1324, interaction: p=0.3460). (C) Modulation of 4 measures on the reconstructed clicks (see *Methods*). Darker colors indicate low rate click trains. The top (resp. bottom) panel shows Avocado's (resp. Lemon's) data. Values were measured for each click position in click trains and averaged over click trains (yielding N measures per neuron in each condition where N is the number of clicks in a train of the corresponding condition). All measures have been computed in the passive encoding. Reconstruction error is especially increased for low rate click trains, as is the peak value, consistent with the lower evoked response in engaged sessions. Results on width and reactivity are discussed in the text.

located 500 ms after the neutral stimulus onset.

Considering the strong changes affecting neuronal responses we just described, we thought that neuronal phase-locked encoding of stimuli was degraded during behaviour.

**Modulation of the encoding quality:** To verify this hypothesis, we used optimal SR to compare the quality of click timings encoding in each condition. For a given time-varying stimulus and several responses to that stimulus, SR provides a means to decide which response is the most phase-locked to the stimulus. We observed that the optimal reconstruction quality was significantly lower during behaviour($p< 5e\text{-}4$, paired t-test), as can be visualized on Figure 6A. It was not clear what category of clicks (ref/target, high rate/low rate) was most affected by this drop in reconstruction quality (Figure 6B N-way ANOVA, ref/tar: p=0.0633, low rate/high rate: p = 0.1324, interaction: p=0.3460), although the general trend was that the drop in reconstruction quality preferentially affected reference click trains as compared to target. If confirmed, this effect of the stimulus identity on the reconstruction quality would suggest that A1 has a preliminary role in task-relevant representation of the stimuli (please note that reference and target are composed of the same number of low rate and high rate clicks so direct comparison between the two is possible).

**Modulation of neuronal tunings:** Using passive SR (Figure 6C) we were able to quantify changes in the neuronal tunings during behaviour (see *Methods* for the difference between passive and optimal SR). Reconstruction quality and peak value (first and second panels) were strongly lower in the engaged condition, consistent with modulations of evoked firing rates shown in Figure 5B. Low rate clicks appeared to be encoded as a train of separate onsets, and the response to each of them was suppressed. However, high rate clicks displayed opposite trends for the two animals, possibly due to the difference of their identity in the task. This would suggest a competing effect of task and rate in the modulation of the quality of encoding (low rate and reference being preferentially affected by the drop in encoding quality). Width (third panel), indicative of the spike dispersion after each click, is increased for reference click trains. The difference with targets is not significant, but supports an effect of the stimulus identity in the encoding. Reactivity (fourth panel) measures the difference of reaction times for each neuron between engaged and passive conditions. Surprisingly, only one ferret (Lemon) exhibits a strong modulation of reactivity, being more reactive in the engaged state for reference (neurons responded in average 2 ms earlier in the engaged state with respect to the passive state). The same result is not observed for target but this is probably due to the great difference between the sample sizes (high rate trains contain more clicks than low rate trains). Difference between the two animals are difficult to explain but could be due to the difference between their target stimuli.

Together, these results show that, during behaviour, A1 neuronal responses exhibit an overall drop in the encoding of the phase-locked component of the signal. This result might appear somewhat paradoxical since one could expect the brain to represent more accurately the stimulus when engaged

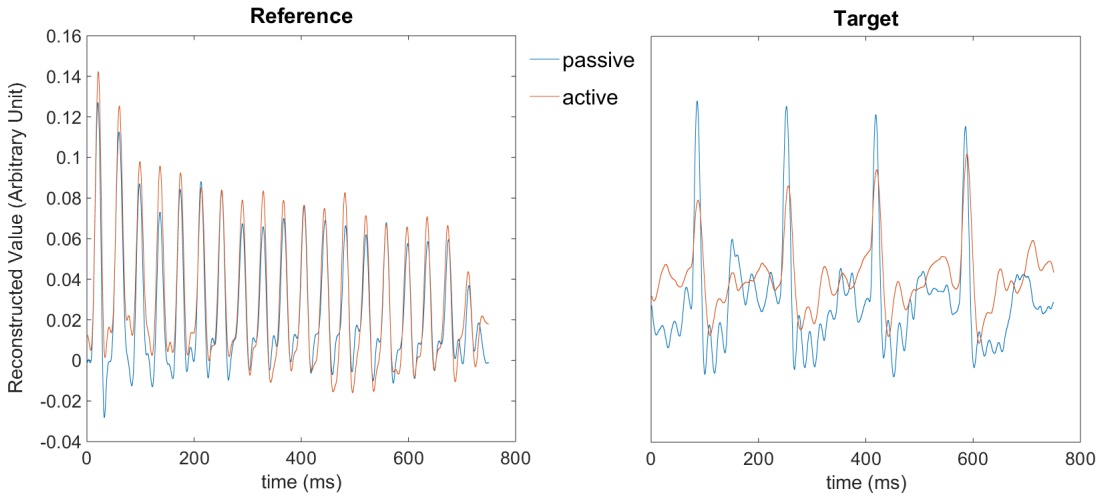
in a decision task based on the discrimination of these stimuli. However, this degradation of the signal seems to be task-relevant although only tendencies were observed at this stage. Moreover, the task only requires the ferrets to discriminate between the two stimuli, and not necessarily to focus on the exact features of the signal. Complementary analysis are currently done on the same data suggesting that the phase-locked encoding is partly replaced by a 'rate encoding' (neurons encode the stimulus by spiking with a different rate to each category along the time course of trials) explaining this global degradation of encoding we observe in the engaged state. More details on these analysis are discussed discussed later on.

### 2.2 A task-driven and behaviourally-driven adaptation mechanism

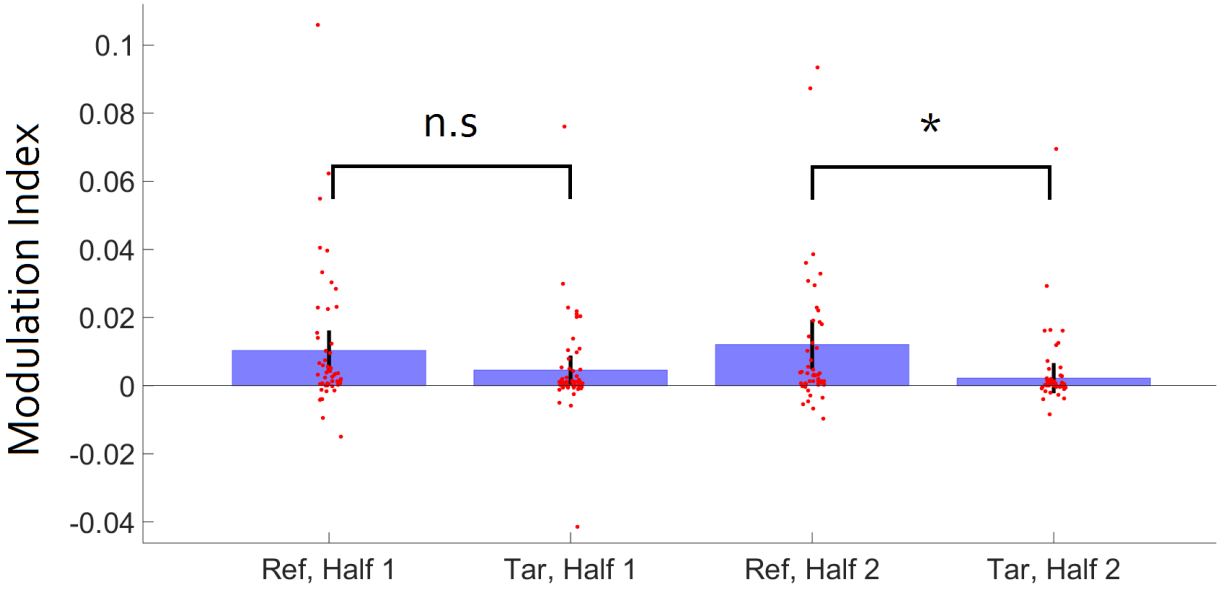
It is reported in Castro-Alamancos 2004, that adaptation is linked to animals behavioural state. One can observe on an example of reconstructed click trains Figure 7A, that during behavior, adaptation seemed to be reduced for target click trains during behaviour. This is consistent with the idea that in our paradigm, reference sounds are non-relevant stimuli. We thus expected to see a greater drop in reconstruction accuracy during the second half of reference click trains as compared to target in active sessions. We first looked at the modulation of the error in optimal SR between the two halves of click trains Figure 7B. We observed that modulation of reconstruction error was significantly affected by the stimulus identity during the second half of the clicks (2-way ANOVA, ref/target: p=0.0375, animal: p=0.322) whereas no difference was observed during the first half between reference and target (2-way ANOVA, ref/target: 0.3055, animal: 0.516). This suggested that neuronal responses in A1 could exhibit task-specific adaptation on the (safe) reference click train with respect to the (warning) target click train. To test this hypothesis, we computed the adaptation index (see *Methods*) for each click train in passive and engaged session. We defined the selectivity of adaptation as

$$S(r,t) = \frac{t-r}{r+t}$$

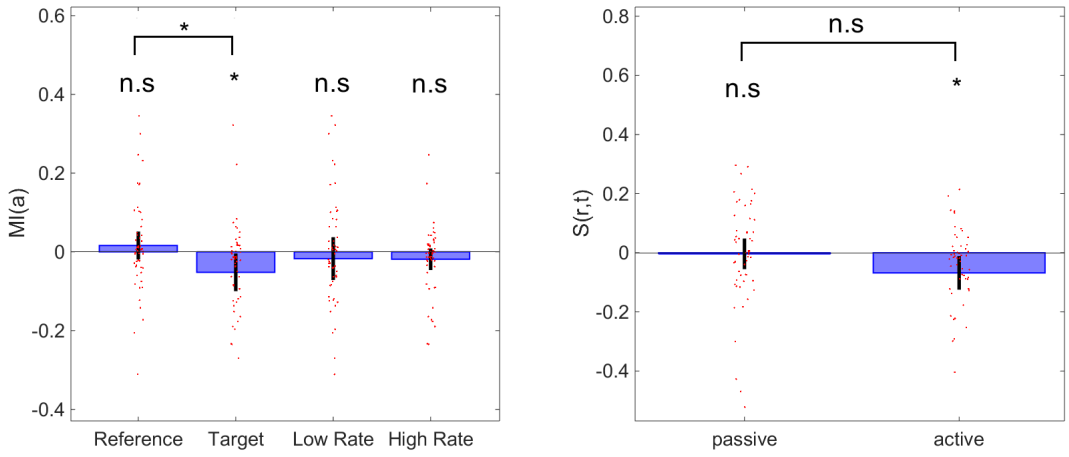
where $r$ and $t$ are the adaptation index of reference and target click trains. Negative values of $S(r,t)$ indicated a reference specific adaptation. In the engaged session, adaptation was observed to be significantly reference-specific (p=0.0112, paired t-test), thus supporting the idea of a task-specific adaptation. Adaptation was not observed to be significantly preferential for reference or target in passive sessions, suggesting that the mechanism we described could also be specifically activated by task-engagement: during passive sessions, animals do not adapt preferentially on any stimulus because there is no need to enhance one stimulus as compared to the other. However, no significant difference was observed in the adaptation selectivity between passive and engaged sessions (2-Way ANOVA, animal: p = 0.975, passive/engaged: p = 0.0699, see Figure 7C). A possible explanation of stimulus-specific adaptation could be due to the difference in the number of occurrences between reference and target (as shown in [**?**]), and thus be independent of behavioural state. However, when computing the modulation of adaptation index we observed a decrease in the adaptation of responses to target click trains (p = 0.0225, paired t-test) and this decrease preferentially affected target than reference (N-way ANOVA, ref/tar: p = 0.0135, low rate/high rate: p = 0.9524, interaction: 0.124). We could thus reject the hypothesis that a similarly selective adaptation was occurring in both state. We also tested the effect of rates, as a control that adaptation was not rate-specific. The difference of adaptation to target between passive and engaged states together with the fact that this difference is more pronounced for target than for reference confirms that the mechanism identified is task-relevant, driven by behaviour and is not simply due to the structure of the task itself (otherwise, similar modulations for reference and target should be observed). These results support the idea that in order to perform a temporal discrimination task, the difference of adaptation of neuronal responses to reference and target click trains is increased during behaviour, so that targets are less adapted than references. By this means, A1 neurons are therefore playing an early facilitative role in stimulus discriminability through the phase-locked encoding of signals, and thus contributing to plasticity in the temporal domain.




**(A)** Example click train reconstruction.




**(B)** Modulation of reconstruction error during click trains first and last halves.




**(C)** Adaptation of click trains is task relevant.

**Fig. 7 – Adaptation of neuronal response:** (A) Two reconstructed click trains from an example session among Lemon's recordings. On the left, the reconstruction of reference click trains exhibits similar adaptation in both passive (blue curve) and engaged (red curve) conditions. On the right, target click trains reconstruction exhibits adaptation in the passive condition whereas facilitation occurs instead in the engaged condition. (B) Reconstruction error for reference and target during the first half (left) and last half (right) of click trains. During the first half, the difference of encoding between reference and target is not significant, but becomes significant in the second half, suggesting a selective adaptation. (C) Left panel: MI($a$) Modulation from passive to engaged conditions of the adaptation index (see *Methods*), from left to right: for reference, target, low rate and high rate click trains. Each red data point shows the modulation for one neuron and one session. Adaptation is consistently decreased for target click trains (p = 0.0225, paired t-test) independent of the rate, whereas it does not change for reference click trains. Neuronal adaptation is influenced by the task (p=0.0135) but not by the rate (p = 0.1291), (N-way ANOVA). Right panel: S($r, t$) is the selectivity of the adaptation to the task, as measured by $\frac{t-r}{r+t}$ where $r$ and $t$ are the adaptation index for target and reference click trains. Adaptation is selective in the active state, and not in the passive state (Although difference between the two is not significant: p = 0.699)

## 2.3 Non-linear stimulus reconstruction increases single-trial reconstruction reliability

We wanted to assess whether the classical SR method could be improved by a non-linear model (see *Methods*). Among the arguments in favour of such an idea, is the fact that neurons themselves are known to respond non-linearly to their inputs. Here we propose an argument supporting the fact that non-linear SR achieves better decoding than linear SR.

Comparing the two models is a delicate question when no ground truth is known about what the neuronal responses are really encoding. Reconstruction error was not our preferred choice for comparing the two methods: results in favour of non-linear SR could be interpreted as the fact that it was a more powerful model to fit data on a given stimulus. Although methods exist in order to remove this kind of bias (Deviance Information Criterion), a ground truth must be known which is not the case here or would need to make use of neuronal models. Instead we rather wanted to quantify to what extent a model was able to robustly separate clicks from silence in the neuronal response. Comparison between reconstructed clicks for the two models, as shown for an example session in Figure 8A, seemed to indicate that the non-linear model would indeed outperform the linear model on this criterion. In this aim, we computed the inter-trial variability at each time bin in the click train. Lower inter-trial variability near click positions indicated lower variability of the timings and peak values of the reconstructed clicks. Lower inter-trial variability during silence implied that fewer clicks were falsely decoded during silence or that the overall reconstructions were less noisy.

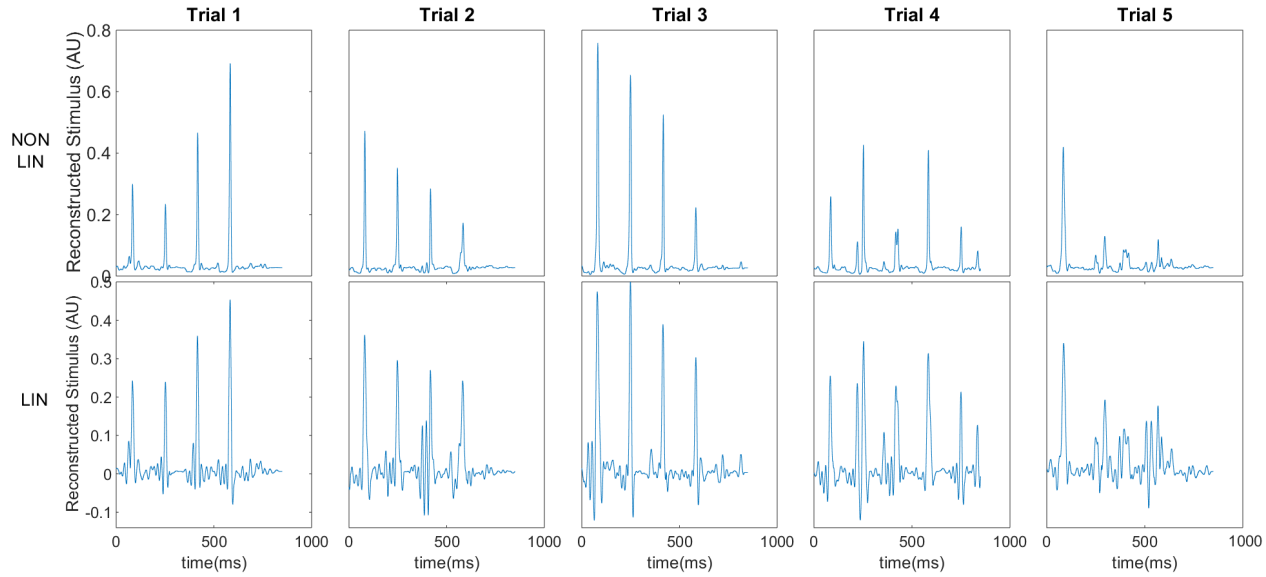
Inter-trial variability was quantified at each time bin as the estimated coefficient of variation (CV)

$$CV(t) = \frac{\hat{\sigma}(t)}{\hat{\mu}(t)}$$

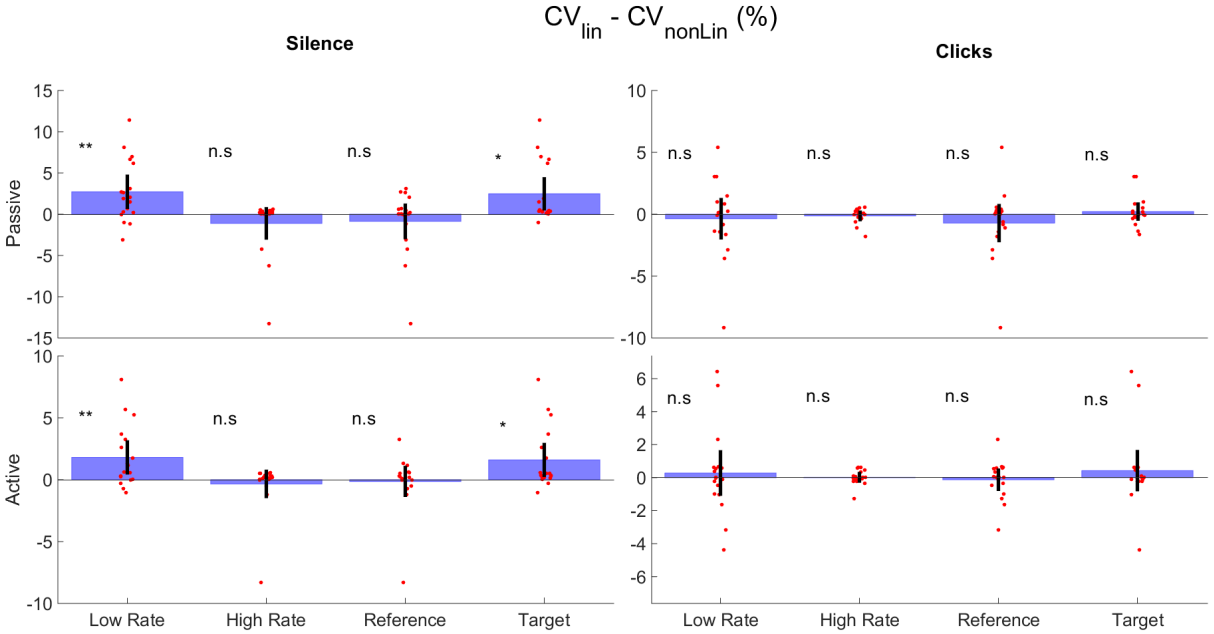
where $\hat{\mu}(t)$ and $\hat{\sigma}(t)$ are the measured sample mean and standard deviation across all trials of the reconstructed stimuli at bin $t$. CV is a standard measure of dispersion of a probability or frequency distribution. We show here the comparison of the two methods using a fixed set of SR parameters defined in *Methods*.

We observed (Figure 8B) that non-linear SR yielded lower inter-trial variability than linear SR during inter-click gaps (left panel) for low rate click trains (p < 5e-4, paired t-test) and target clicks (p < 5e-4, paired t-test) once data from both animal were pooled. No differences between the two models were observed during inter-click gaps for high rate clicks and reference clicks nor at click positions (right panel) for all clicks. Overall variability in inter-click gaps was not significantly different between the two models yet a strong tendency (p = 0.0563, paired t-test) was observed in favour of the non-linear model.

To explain the fact that no difference was observed near click positions, we propose that variability could already be present in the neuronal encoding of the clicks, since neuron could well be spiking at a different rate each time a click occurs. The fact that no difference was observed for high rate clicks could be due to the fact that neurons are less phase-locked at high rates [**?**]. The difference observed on target click trains seems to be due to the difference in training set size between reference and




**(A)** Single-trial reconstructions from a sample session




**(B)** Difference in Coefficient of variation CV between linear and non linear conditions.

target. This would suggest that the non-linear algorithm might be more efficient in using examples and generalize them to unknown data: for a given training set size, the non-linear method converges more rapidly to the optimal filter than the linear one. It will be easy to test this hypothesis later, by taking equal training and test set sizes for reference and target.

Non-linear SR thus appeared to be a possible improvement of linear stimulus SR for neuronal response analysis. Further work needs to be done in order to compare the two methods and determine precisely in what context each of them achieves better decoding.

**Fig. 8** – Comparison between linear and non-linear stimulus reconstruction. (A) Non-linear (top) and linear (bottom) reconstructions of 5 example blocks (low rate click trains, 6Hz) from one of Avocado's sessions. One can see that decoding is very accurate with the non-linear model, whereas reconstructions with the linear model are noisy. (B) Comparison of the coefficient of variation (CV) for linear and non-linear reconstructions, during inter-click silences (left) and at click positions (right). Blue bars indicate the mean difference $CV_{lin} - CV_{nonLin}$ across each condition (low rate, high rate, reference and target), and each red data point represents the value for a single session (n=18). Comparisons have been done separately for passive (top) and engaged (bottom) sessions.

# 3 Discussion

## 3.1 Biological mechanisms of task engagement

**Thalamo-cortical regulation of sensory inputs** We described the radical changes induced by task-engagement in the neuronal responses. Among them, is a robust and consistent increase in the spontaneous firing rate during all experiments. This effect has also been reported in other studies in A1 (for example [**?**] for macaques) although discrepancies exist in literature. In [**?**], spontaneous firing rates in rats were not reported to increase in a set-up comparable to ours. However, evoked firing rates were reported to decrease, consistent with our results. Modulations in the evoked and spontaneous firing rates could be due to acetylcholine (aCh) release in the neocortex in consequence of to task-engagement.

Indeed, it has been found that injections of aCh in the barrel neocortex resulted in changes in evoked and spontaneous firing rate [**?**], similar to our observations. Depending on the dose of aCh, spontaneous firing rates are decreased or increased while evoked firing rates are suppressed. The same type of results are observed when stimulating the basal forebrain (BF) with opto-genetic methods [**?**]. Cholinergic neurons from the BF project in the auditory cortex [**?**] and the auditory thalamus [**?**]. Presence of aCh in the cortex and the thalamus results in desynchronized neuronal activity [**?**] and switches the thalamic firing in a tonic mode. This has the effect of depolarizing thalamo-cortical neurons [**?**], leading to suppressed responses in the cortex. As the BF is one of the most important cholinergic centre in the brain, a likely interpretation of these experiments is that stimulation of the BF induces aCh release in the thalamus and the neocortex, responsible for modulations in the neuronal firing rates like those we reported.

Release of aCh by the BF is strongly correlated with task-engagement. Lesions of the BF strongly decrease task performance, by impeding modulation of attention on relevant stimuli [**?**], decreasing vigilance and inducing perseveration [**?**]. Moreover, aCh releases are correlated with active behavioural states [**?**] and are predictive of stimulus detection performance[**?**]. This strongly suggests that, when animals engage in a task, their BF increases its aCh release in the neocortex and the thalamus to increase performance.

Regulation of aCh is governed by the pre-frontal cortex (pFC). Indeed, lesions of pre-frontal cortex abolish cortical aCh release [**?**]. In the light of those results, a possible mechanism explaining the neuronal modulations we observed is the following: sensory cortices stimulation (incoming stimuli) activate the pFC which induces the BF to release aCh in the neocortex depending on task demands. The released aCh activates the thalamus and desynchronises the neocortex, resulting in suppressed responses and enhanced or suppressed spontaneous firing rate. This explanation is also supported by ongoing study in our lab showing that stimulation of ferrets pFC with opto-genetic methods while they are listening to sounds, results in enhanced spontaneous firing rate and suppressed responses in A1.

Why would response suppression be useful for task-performance ? A possible explanation could be the following: in passive state, no task is defined, and sensory inputs propagate through the cortex to be processed in background. In the engaged condition, a task is well defined and sensory inputs must be focused to precise parts of the brain. Suppression of sensory inputs in the neocortex by the thalamus might allow a more efficient representation of stimulus features in the brain [**?**].

**Stimulus adaptation during task engagement:** We investigated the effects of task-engagement in stimulus adaptation and found task-specific and behaviourally-driven adaptation. We have shown

that in the engaged condition, target click trains are less adapted than in the passive condition, and that a similar decrease in adaptation did not affect reference click trains. Our data suggest that in the passive state, adaptation for the two stimuli is equivalent, even though the target click trains were presented less often. This contrasts with the results exposed in [**?**] in cats primary auditory cortex, where rare stimuli tend to be less adapted than common stimuli (adaptation in this study was defined as the exponential decrease in the sustained response to pure tones). Adaptation could be a mechanism useful to enhance neuronal code efficiency [**?**] and might be linked to the mechanisms of task-engagement involving aCh regulation we just described [**?**]: adaptation is reduced and responses are suppressed during active states [**?**]. Here we projected neuronal responses to the stimulus domain where adaptation was easier to measure using SR. We found that stimulus adaptation was task-relevant and enhanced the contrast between targets and references, as was demonstrated in other studies looking at receptive field plasticity [**?**, **?**].

**Decreased quality of phase-locked encoding and possible transition to a rate encoding:** Our analysis showed that phase-locked components of the neural encoding of click trains were substantially degraded during behaviour. This result was unexpected and hard to explain for us. One could expect both the phase-locked and the rate component of the neural responses to be enhanced during behaviour [**?**] (in this study, both component are found to be enhanced yet the rate encoding is more robustly enhanced than the phase-locked). We propose that in our task, low and high rate click trains might be easily distinguished in the spectral domain, given the level of difference between low rates and high rates (Most of low rates are under 9Hz, close to the limit of ferrets audible range, while all high rates have frequencies in the audible range [**?**]). Therefore, a phase-locked encoding might be less efficient than a rate-based spectral encoding. It is readily known that neurons along the primary auditory pathway progressively convert a temporal coding into a rate coding [**?**]. An ongoing study in our lab on the same data shows that in the active state, discriminability of responses to target and reference click trains based on the rate is enhanced in engaged state during the post-click silence period, also consistent with [**?**]. Thus we propose that in our discrimination task, discrimination of the stimuli is easier in a rate encoding, while phase-locked encoding might be inefficient and too costly explaining why engaged responses are less phase-locked than passive responses.

### 3.2 Stimulus Reconstruction

**Reconstructed stimuli and PSTH:** In most neuroscience studies, neuronal activity is studied as PSTH requiring to average responses across trials. Directly comparing or summing PSTH can be problematic. Some neurons increase their rate in response to events (excitatory neurons) while others decrease their firing (inhibitory). Moreover, lags in the neuronal response are not constant over the population, so spikes evoked by stimulus events cannot always be aligned easily. Using SR, one can find the optimal lag and weight to give to every small portions of neural responses in order to project the response in the stimulus domain, where they are easier to interpret. SR is particularly suitable for click trains and more generally to reconstruct temporal envelopes of auditory stimuli [**?**]. Here we showed that SR also has the advantage to allow single trial exploration of the neuronal response, which is usually forbidden with average PSTH. SR can be used in a similar way to decode attentional selection from single-trial EEG recording [**?**].

**Dealing with neuronal non-linearities in stimulus reconstruction** Linear SR assumes that neurons respond linearly to the sound envelope of stimuli. Such a linear correspondence is rather unlikely, since neurons themselves respond non-linearly to their inputs. A first approach is to take non-linearity into account during the decoding phase [**?**]. An alternative is to improve the stimulus representation by finely modelling the non-linearities that occur downwards the recorded area. This way, several representations can be tested and compared based on the quality of the fit [**?**].

**Aspects of non-linear reconstruction:** Non-linear SR has a few advantages over linear SR that we expose here. First of all, non-linear SR being equivalent to a simple feed-forward neural network,

it can be thought as a sketch of how the brain integrates neuronal data to extract stimulus features, such as click timings here. Moreover, non-linear SR yields bounded reconstructed stimuli (between 0 and 1) while the linear model can yield negative values, sometimes difficult to interpret.

Another important aspect of the difference between linear SR and our non-linear model is the possibility to saturate in information. We argue that linear SR is not suited to decode a neuronal population in which stimulus features are variably encoded. Here is an example of this idea: suppose the stimulus to be reconstructed is defined as

$$\begin{cases} s(t) & = 1 \quad \text{if a click occurs at time t} \\ s(t) & = 0 \quad \text{otherwise} \end{cases}$$

In a linear framework, it occurs that some click positions are encoded "too" precisely by the network, leading to $\hat{s}(t) > 1$. Similarly, during silence periods, it can occur that when adding evidence from all neurons, $\hat{s}(t) < 0$ (this occurs for example for a neuron that reliably respond to one aspect of the stimulus at a given rate, but which in a minority of cases, enhance its response, due to variability and uncertainty of the neuronal encoding). Intuitively such situations should be considered as better than $s(t) = 1$ at click positions and $s(t) = 0$ during silence since the contrast between clicks and silence would be higher and thus click timings could be decoded with better accuracy. However, this "better" situation is actually penalized by linear SR, possibly preventing some neurons to be used by the reconstruction filter at the full range of their informative potential. The sigmoid transformation in the non-linear model can be viewed as a transformation of the stimulus envelope to be reconstructed with a mapping from $[0, 1]$ to $[-\infty, +\infty]$. Additional evidence in the non-linear SR cannot be penalized since the stimulus bounds can never be reached.

We showed that these differences resulted in less inter-trial variability of the reconstructed stimuli. Therefore, non-linear SR might out-perform linear SR in applications such as decoding attention [**?**].

**Decoding linearly or non-linearly neuronal response:** Restricting the optimal backwards mapping to a linear mapping does not mean that one assumes the brain only performs a linear filtering on the primary sensory cortex responses to decode stimuli. Linear mapping can actually be thought of as an unbiased way of quantifying which stimulus features are readily encoded in the neuronal responses. It is very interesting to know, at each point of a multi-layer neural network, which features have already been extracted, which have not, and which have been discarded. More sophisticated decoding such as deep neural networks might extract additional features from this code and thus fail to provide an objective view of the state of stimulus feature extraction in the area of interest.

That said, our use of SR here was slightly different. We did not want to obtain an absolute and unbiased view of the stimulus features encoded in A1, but rather wanted to compare this encoding in passive and engaged condition. Decoding the responses with better accuracy might allow us to estimate more precisely the stimulus adaptation and other measures we defined. We thus proposed a simple non-linear addition to the classical SR and found that it could also be efficiently used for comparing reconstructed stimuli envelopes. This addition was not made to model a physical aspect of biological neurons, but rather to improve the efficiency with which the neural responses were used to reconstruct the stimulus, and to allow the fact that evidence encoded in the neuronal population could saturate. We believe that as long as the decoding performed on the neuronal data stays "close" to a linear decoding, differences observed in the decoded stimuli really reflect differences in the neuronal encodings.