Perception of categories: from coding efficiency to reaction times

Supplementary Information

Laurent Bonnasse-Gahot[1], Jean-Pierre Nadal[1,2,*]

1 Centre d'Analyse et de Mathématique Sociales (CAMS, UMR 8557 CNRS – EHESS), École des Hautes Études en Sciences Sociales, Paris, France
2 Laboratoire de Physique Statistique (LPS, UMR 8550 CNRS – ENS – UPMC – Univ. Paris Diderot), École Normale Supérieure, Paris, France
∗ E-mail: nadal@lps.ens.fr

1. Optimization of the decoding layer: supervised learning scheme

For the numerical simulations, we made used of a supervised learning scheme which we present here, proving that, in the asymptotic limit of a very large training set, the chosen cost function gives the cost $\overline{\mathcal{C}}$ considered in the theoretical analysis.

During learning, stimuli are presented sequentially, along with their category label. For a given stimulus x, the output $g(\mu|\mathbf{r},\mathbf{w})$ is compared with the desired binary output given by indicator function $t_\mu(x)$ (for ***t**eacher*), defined as follows:

$$t_\mu(x) = \begin{cases} 1 & \text{if } x \in \mu \\ 0 & \text{otherwise} \end{cases} \tag{S.1}$$

where $x \in \mu$ means that stimulus x belongs to the category labeled μ. The distance between the output $g(\mu|\mathbf{r},\mathbf{w})$ and the teacher value $t_\mu(x)$, is measured by the following training cost function:

$$\mathcal{C}_t(x,\mathbf{r}) \equiv \sum_{\mu=1}^{M} t_\mu(x) \ln \frac{t_\mu(x)}{g(\mu|\mathbf{r},\mathbf{w})} \tag{S.2}$$

Its average over all the realizations of the neural activity $\mathbf{r}$ is given by:

$$\mathcal{C}_t(x) = \int d^N\mathbf{r}\, P(\mathbf{r}|x) \sum_{\mu=1}^{M} t_\mu(x) \ln \frac{t_\mu(x)}{g(\mu|\mathbf{r},\mathbf{w})} \tag{S.3}$$

Let us now show that a large number of stimulus presentations during the learning phase leads to estimate posterior probabilities (in a way similar to the one presented in Duda et al., 2001).

After n stimulus presentations, the mean cost function becomes:

$$\begin{aligned}
\frac{1}{n}\sum_x \mathcal{C}_t(x) &= \frac{1}{n}\int d^N\mathbf{r} \sum_x P(\mathbf{r}|x) \sum_\mu t_\mu(x) \ln \frac{t_\mu(x)}{g(\mu|\mathbf{r},\mathbf{w})} && \text{(S.4)} \\
&= -\frac{1}{n}\int d^N\mathbf{r} \sum_\mu \sum_{x\in\mu} P(\mathbf{r}|x) \ln g(\mu|\mathbf{r},\mathbf{w}) && \text{(S.5)} \\
&= -\sum_\mu \int d^N\mathbf{r}\, \frac{n_\mu}{n} \frac{1}{n_\mu} \sum_{x\in\mu} P(\mathbf{r}|x) \ln g(\mu|\mathbf{r},\mathbf{w}) && \text{(S.6)}
\end{aligned}$$

where n_μ is the number of stimuli labeled μ among the n stimuli that were presented to the network.

For a very large number of stimuli, the mean cost $\overline{\mathcal{C}_t}$ then writes:

$$\overline{\mathcal{C}_t} \equiv \lim_{n\to\infty} \frac{1}{n}\sum_x \mathcal{C}_t(x) = -\sum_\mu \int d^N\mathbf{r}\, q_\mu \int dx\, P(x|\mu) P(\mathbf{r}|x) \ln g(\mu|\mathbf{r},\mathbf{w}) \tag{S.7}$$

hence, given that $\int dx\, P(x|\mu)P(\mathbf{r}|x) = P(\mathbf{r}|\mu)$, and that, according to Bayes rules $q_\mu P(\mathbf{r}|\mu) = P(\mathbf{r})P(\mu|\mathbf{r})$, we get

$$\overline{\mathcal{C}_t} = -\int d^N\mathbf{r}\, P(\mathbf{r}) \sum_\mu P(\mu|\mathbf{r}) \ln g(\mu|\mathbf{r},\mathbf{w}) \tag{S.8}$$

This is the same as $\overline{\mathcal{C}}$ except for a constant additive term (the entropy $\mathcal{H}(\mu|x)$), implying that minimization of the cost leads to estimate the posterior probabilities, as desired.

In the numerical illustrations, learning is done through a gradient descent algorithm (Rumelhart et al., 1986) aiming at minimizing the cost function (S.2), with the presentation to the network of 30000 stimuli along with their category label.

2. Test on two categories: numerical details

In the numerical test of the ability of the decoding layer to achieve efficient decoding, Section 3.1.2, we assume the two categories to be equiprobable, and each one characterized by a Gaussian distribution, centered at $x^{\mu_1} = -2$ and $x^{\mu_2} = 2$, with a width $a^{\mu_1} = a^{\mu_2} = 1.5$. These numbers are arbitrary and chosen for illustrative purpose only. We consider a neuronal population with $N = 14$ coding cells. The activity r_i of each neuron is given by a Poisson statistics with mean firing rate $f_i(x)$, corresponding to a bell-shaped tuning curve:

$$f_i(x) = f_{\min} + (f_{\max} - f_{\min}) \exp\left(-\frac{(x - x_i)^2}{2a_i^2}\right) \tag{S.9}$$

The preferred stimuli of the cells are equidistributed over the domain $[-6, 6]$. The width and the minimal and maximal values of the tuning curves are the same for all the neurons $a_i = 1.38$, $f_{\min} = 0.001$ and $f_{\max} = 5$).

3. Temporal evolution of the output of the network

This section presents a qualitative comparison of the temporal evolution of the decision in our model and as found in the experimental data of McMurray and Spivey (2000). In this experiment, subjects are presented with a continuum of 9 stimuli, ranging from category /ba/ to category /pa/, and whose voice onset time (VOT) values vary from $x_1 = -50$ ms to $x_9 = 60$ ms. The task is to identify the category by clicking the corresponding button on a screen. Using an eye-tracking method, this behavioral study measures the time spent by subjects looking at the two buttons after hearing a given stimulus. Here we can consider the VOT as the relevant x-space.

This simulation makes use of the two category model of the previous section, with a rescaling of the parameters chosen to corresponds to this specific experiment. Here one unit of the x space in the simulation corresponds to a difference in VOT of 13.75ms (the spacing between two consecutive stimuli), with the categories centered at $x^{\mu_1} = -22.5$ms and $x^{\mu_2} = 32.5$ms, and a width $a^{\mu_1} = a^{\mu_2} \sim 20.6$ms. The stimulus domain corresponds to VOTs in the range

[−77.5ms, 87.5ms].

The temporal evolution of the output of the network reflects the accumulation of the categorical information extracted from the neuronal activity. The learning phase was performed on a time window $[0, \tau_a]$ so that $\tau_a f_{\max}$ represents the mean number of spikes emitted by cell i during this time interval when the stimulus corresponds to its preferred stimulus. One can then look at the output $g(\mu|\mathbf{r}, \mathbf{w})$ for different values of $\tau \in [0, \tau_a]$. Averaging over different realizations of this activity (1000 realizations in this numerical example), we finally get an estimate of the average value taken by the output $g(\mu|\mathbf{r}, \mathbf{w})$ for each interval $[0, \tau]$. Figure 1 (Left) shows the temporal evolution of the mean values of the output $g(\mu|\mathbf{r}, \mathbf{w})$ for different stimuli along the continuum $x_1 = -50\text{ms}, \dots, x_9 = 60\text{ms}$ (the curves getting redder and darker as τ increases). For comparison, Figure 1 (Right) shows the results from the above-mentioned experimental study of McMurray and Spivey (2000): one sees a gradual increase of categorical information, characterized by a sigmoid that expands over time, in qualitative compliance with our model.

4. Reaction times: numerical illustration

This numerical example involves two equiprobable Gaussian categories, centered in $x^{\mu_1} = -3$ and $x^{\mu_1} = 3$, with standard deviation $a^{\mu_1} = a^{\mu_2} = 1.5$. The neuronal population (coding layer) is made of $N = 10$ cells, with bell-shaped tuning curves (Eq. (S.9)). The preferred stimuli x_i of the neurons are initially equidistributed along the domain $[-6, 6]$. Before learning, each tuning curve has the same width ($a_i = 2$). Minimal and maximal values of the firing rates are respectively set to $f_{\min} = 0.001$ and $f_{\max} = 5$.

During the learning phase, 100000 stimuli are presented to the network, and both the weights $\mathbf{w}$ and the parameters of tuning curves (width and location) are optimized. The time window τ_a used during learning is equal to 1.

After learning, we look at the response of the network following the presenta-

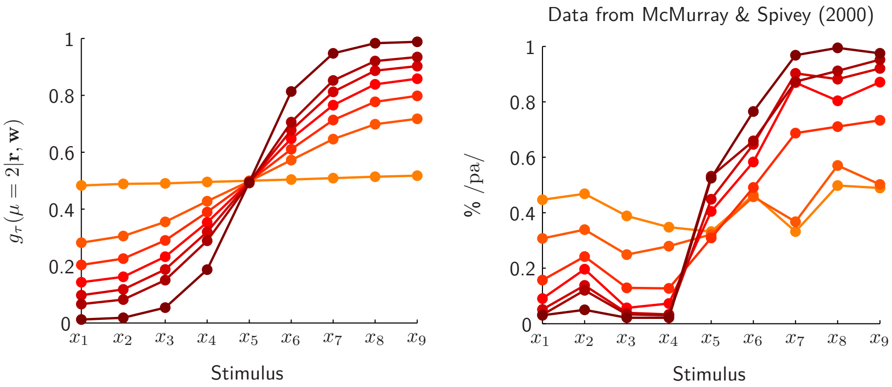

Figure 1: Qualitative comparison of the temporal evolution of the decision between the model and as found in experimental data. (Left) Averaged temporal evolution of $g(\mu = 2|\mathbf{r}, \mathbf{w})$ along the continuum $x_1, \ldots, x_9$. The increase in the length of the time window $[0, \tau]$ is indicated by a color gradient ranging from orange to dark red. (Right) Evolution of the proportion of looking time to the category /pa/ vs the category /ba/ for different stimuli whose voice onset time (VOT) values vary from $x_1 = -50$ ms to $x_9 = 60$ ms (data extracted from McMurray and Spivey, 2000)

tion of a stimulus, according to the diffusion model presented in section 2.2.3. The simulation of this diffusion process is done as follows. We first generate a Poisson process by dividing the time interval $[0, 3\tau_a]$ into 3000 bins. For a neuron i, each interval, of width $d\tau = \tau_a/1000$, receives a spike according to a Bernoulli law of parameter $f_i^0(x)\, d\tau$ ($d\tau$ being small, we thus get a Poisson process associated with each neuron). We then compute the temporal evolution of the output α_τ as well as the time τ_d for which this quantity reaches one of the two bounds for the first time. In this numerical example, the bound γ is set equal to 0.3. For each stimulus x, this process is run 10000 times, which makes it possible to have an estimate of the mean reaction time $\overline{\tau_d}(x)$. In the end, this operation is done for 20 stimuli equidistributed along a continuum ranging from -4 to 4.

Following learning, the behavior of the neural population, with respect to discrimination sensitivity and reaction times, qualitatively reproduces a classic situation of categorical perception, as summarized in Figure 2. Identification curves are characterized by an S-shape; mean reaction times are longer at the boundary between categories than within category (see e.g. Pisoni and Tash, 1974; Studdert-Kennedy et al., 1963); discrimination accuracy (as quantified by Fisher information $F^0_{\mathrm{code}}(x)$) is higher at the boundary between categories than within (e.g. Liberman et al., 1957; Repp, 1984; Bornstein and Korda, 1984; Goldstone, 1994; Kuhl and Padden, 1983), which captures the so-called categorical perception phenomenon.

Figure 3 (Left) compares the mean reaction times obtained in the numerical simulation with the ones predicted from formula (3.28) and (B.3).We can first emphasize the remarkable correspondence (up to a scaling factor) between the simulated data and the data predicted by our equation, despite the fact that there is only 10 cells in the coding layer. Using parameters of the linear regression extracted from Fig. 3, we can then reconstruct the mean reaction time for the whole continuum. This reconstructed mean reaction time is shown on Figure 3 (Right, red line), together with the values obtained in the simulation (open circles). Here again, one can note the remarkable correspondence between the simulated and predicted values. Note though that the values given by our formula (see the x-axis in Fig. 3 (Left)) are smaller than the true values, hence the need in each case of rescaling the data in order to reconstruct the simulated reaction times. We attribute this bias to finite size and discretization effects.

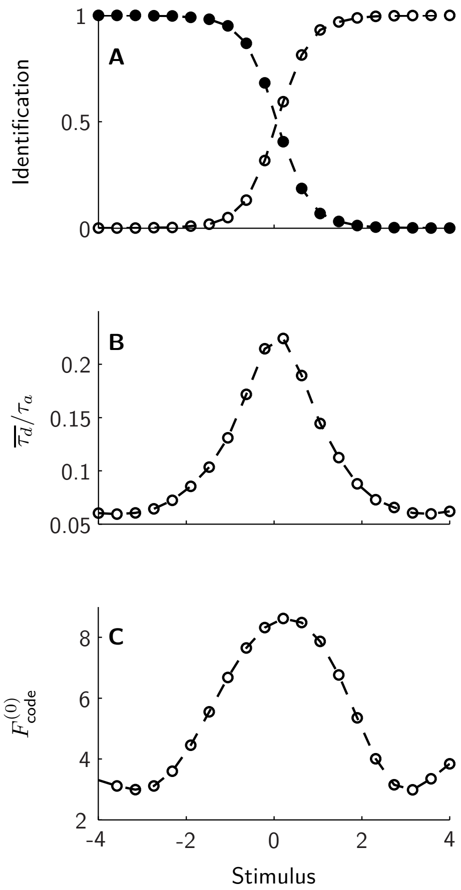

Figure 2: Perceptual consequences of category learning: results of the numerical simulation. (A) Mean identification function. (B) Mean reaction times. (C) Fisher information rate of the neuronal population (measure of perceptual sensitivity). These results qualitatively reproduce a classic situation of category learning, in particular in the case of phonemic perception (see e.g. Pisoni and Tash, 1974, Fig. 3). Identification curves are characterized by an S-shape; mean reaction times are longer at the boundary between categories than within category; discrimination accuracy (as quantified by Fisher information $F^0_{\mathrm{code}}(x)$) is higher at the boundary between categories than within, *ie* the neural population exhibits categorical perception.

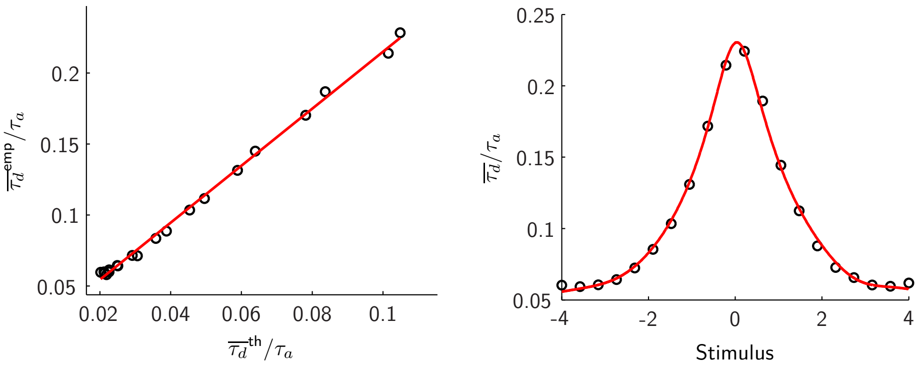

Figure 3: **Reaction times: comparison between simulated data and theoretical prediction.** (Left) Mean reaction times τ_d^{emp} obtained by numerical simulation for the 20 stimuli spanning the considered continuum, as a function of the mean reaction times given by Eq. (3.28). The red line corresponds to the linear regression (correlation coefficient r=0.9986, p=1.7e-24). (Right) Mean reaction times as a function of the stimulus presented. The open circles indicates the mean reaction times obtained by numerical stimulation, whereas the red line corresponds to the results derived from Eqs. (3.28), (B.3).

References

Bornstein, M., Korda, N., 1984. Discrimination and matching within and between hues measured by reaction times: some implications for categorical perception and levels of information processing. Psychol. Res. 46, 207–222.

Duda, R., Hart, P., Stork, D., 2001. Pattern Classification. Wiley, New York, USA, 2nd Edition.

Goldstone, R., 1994. Influences of categorization on perceptual discrimination. J. Exp. Psychol. Gen. 123 (2), 178–200.

Kuhl, P., Padden, D., 1983. Enhanced discriminability at the phonetic boundaries for the place feature in macaques. J. Acoust. Soc. Am. 73 (3), 1003–1010.

Liberman, A., Harris, K., Hoffman, H., Griffith, B., 1957. The discrimination of speech sounds within and across phoneme boundaries. J. Exp. Psychol. 54, 358–369.

McMurray, B., Spivey, M., 2000. The categorical perception of consonants: The interaction of learning and processing. Proc of the Chicago Linguistics Society 34 (2), 205–220.

Pisoni, D., Tash, J., 1974. Reaction times to comparisons within and across phonetic categories. Percept. Psychophys. 15 (2), 285–290.

Repp, B., 1984. Categorical perception: issues, methods, findings. In: Speech and Language: Advances in Basic Research and Practice.

Rumelhart, D., Hinton, G., Williams, R., 1986. Learning representations by back-propagating errors. Nature 323, 533–536.

Studdert-Kennedy, M., Liberman, A., Stevens, K., 1963. Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. J. Acoust. Soc. Am. 35 (11), 1900.