



Stimulus reference frame and neural coding precision



Lubomir Kostal

Institute of Physiology of the Czech Academy of Sciences, Videnska 1083, 14220 Prague 4, Czech Republic

HIGHLIGHTS

- The stimulus reference frame affects the match between the coding precision and stimulus distribution.
- Stimulus coding efficiency should be quantified on the psychophysical scale.
- The psychophysical perspective yields the Jeffreys prior as the optimal stimulus distribution.

ARTICLE INFO

Article history:

Received 31 July 2015

Received in revised form

29 January 2016

Available online 22 March 2016

Keywords:

Fisher information

Coding accuracy

Measurement scale

Jeffreys prior

ABSTRACT

Any particular stimulus intensity, as a physical quantity, can be equivalently described in different unit systems. Researchers automatically expect the methodology and the inference obtained about the neural coding precision to be independent from such a subjective choice. We show, however, that the Fisher information, which is arguably the most popular measure of coding accuracy, may yield incompatible and in fact arbitrary results just by re-evaluating the identical stimulation scenario in transformed units. We consider only regular scale transformations given by strictly increasing and differentiable functions. On one hand, our results point to a potentially problematic aspect of the Fisher information application. On the other hand, we speculate that the unwanted transformation covariance may be removed by considering the psychophysical scale based on the ideal observer paradigm. We show that such scale implies constant Fisher information and that the matching stimulus distribution is given by the Jeffreys prior. The psychophysical perspective thus provides a novel justification for the special role of the Jeffreys prior in neural coding theory.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

The methods of signal estimation and detection theory play an important role in the study of the neuronal coding problem. In particular, the Fisher information and the Cramér–Rao bound are frequently employed to address the theoretical limits on the coding and discrimination precision (Bethge, Rotermund, & Pawelzik, 2002; Dayan & Abbott, 1999; Greenwood & Lansky, 2005a; Greenwood, Ward, Russell, Neiman, & Moss, 2000; Harper & McAlpine, 2004; Kostal, Lansky, & Pilarski, 2015; Paradiso, 1988; Pilarski & Pokora, 2015; Seriès, Latham, & Pouget, 2004; Stemmler, 1996; Zhang, Ginzburg, McNaughton, & Sejnowski, 1998; Zhang & Sejnowski, 1999). Comparing this precision for different types of responses (e.g., counts of action potentials or intervals in between them) then yields a potentially important piece of evidence about the principles of neural coding. In this regard, the influential *efficient coding hypothesis* (Barlow, 1961) states

that neuronal responses are adjusted, through evolutionary and adaptive processes, to optimally encode such stimulus statistics that matches the local sensory environment (Atick, 1992; Bialek & Owen, 1990; Hornstein, O'Carroll, Anderson, & Laughlin, 2000; Kostal, Lansky, & Rospars, 2008; Lewicki, 2002; Simoncelli & Olshausen, 2001; Smith & Lewicki, 2006). Consequentially, it is often expected that the high-probability stimulus regions match the peaks of the coding accuracy (Dayan & Abbott, 2001; Dean, Harper, & McAlpine, 2005; Laughlin, 1981; Wark, Lundstrom, & Fairhall, 2007; Watkins & Barbour, 2008; Wen, Wang, Dean, & Delgutte, 2009).

In this paper we investigate the general consequences of the measurement scale choice on the coding accuracy, as described by the Fisher information. We focus especially on the potential ambiguity of the conclusions that may be drawn from the shape of the Fisher information curve, viewed either as a function of the stimulus intensity (Section 3.1) or as a function of the probability of the stimulus intensity (Section 3.2). Finally, we speculate that the proper scale, on which the coding accuracy should be interpreted, results from the psychophysical function relating the decoding error to the just noticeable difference in perception (Section 3.3).

E-mail address: kostal@biomed.cas.cz.

We show, by following the efficient coding principle on the psychophysical scale, that the optimal stimulus distribution is given by the Jeffreys prior. The Jeffreys prior arises also as the capacity-achieving input distribution in the limit of vanishing noise during the information transmission (Bernardo, 1979; Brunel & Nadal, 1998; Kostal, Lansky, & McDonnell, 2013; McDonnell & Stocks, 2008; Wei & Stocker, 2015; Yarrow, Challis, & Seriès, 2012), hence our results provide yet another interpretation of its optimality.

2. Methods

The neural coding problem is often approached by virtue of the stimulus–response model, in which the neuronal response r (often given by the firing frequency) is related to the stimulus parameter θ . It is well known that the response varies randomly across trials (Dayan & Abbott, 2001; Shadlen & Newsome, 1998; Stein, Gossen, & Jones, 2005). The stimulus–response model, denoted also as the encoding model, is therefore fully described in terms of the response probability density (or mass) function $f(r; \theta)$.

The stimulus θ can be equivalently evaluated in different physical or measurement units $[\lambda]$, where $\lambda = \varphi(\theta)$, provided that the function φ is bijective. In this paper we restrict ourselves to *regular* transformations of the measurement scale, defined as follows.

Definition 1. Let the stimulus parameter θ take values in a closed interval of the real line $[\theta_{\min}, \theta_{\max}]$. The new stimulus parameter $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ is given by the regular transformation of θ ,

$$\lambda = \varphi(\theta), \quad (1)$$

if φ is strictly increasing and continuous on $[\theta_{\min}, \theta_{\max}]$ and differentiable on $(\theta_{\min}, \theta_{\max})$.

The regular transformations are appealing since they preserve the relative ordering of the stimulus intensities on various scales, i.e., $\theta_1 > \theta_2$ implies $\varphi(\theta_1) > \varphi(\theta_2)$. An example of such φ is given by the sound loudness parameter: one may use the units of sound intensity, pressure or level, which are mutually related by quadratic and logarithmic transformations (Raichel, 2006). The probability distributions $f(r; \theta)$ and $f(r; \lambda)$ thus describe *the same* encoding model, and there is no persuasive preference for the θ -parameterization over the λ -parameterization, perhaps besides convenience.

The problem of the stimulus coding precision, i.e., the exact evaluation of the smallest achievable error, is generally non-trivial (Kostal et al., 2015; van Trees & Bell, 2013). It is often more practical to evaluate the Cramér–Rao bound on the decoding mean square error (MSE) instead. If there is no bias in the decoding and the model satisfies certain regularity conditions (Ibragimov & Has'minskii, 1981; Pitman, 1979), the Cramér–Rao bound states that the MSE for given stimulus θ satisfies

$$\text{MSE}(\theta) \geq \frac{1}{J_\theta(\theta)}. \quad (2)$$

The function $J_\theta(\theta)$ is the *Fisher information*,

$$J_\theta(\theta) = \int \left(\frac{\partial \log f(r; \theta)}{\partial \theta} \right)^2 f(r; \theta) \, d r, \quad (3)$$

which is strictly positive and continuous on $[\theta_{\min}, \theta_{\max}]$, see, e.g., Brown and Gajek (1990) and Ibragimov and Has'minskii (1981, Chapter 7) for details.

The Fisher information $J_\lambda(\lambda)$ in the regularly transformed units is related to the original function $J_\theta(\theta)$ as

$$J_\lambda(\lambda) = \left(\frac{d \varphi^{-1}(\lambda)}{d \lambda} \right)^2 J_\theta(\varphi^{-1}(\lambda)), \quad (4)$$

which can be proven by employing the chain rule for derivatives (Lehmann & Casella, 1998), with φ^{-1} denoting the inverse to Eq. (1). In this paper we examine the potential effect of regular transformations on the inference about neural coding precision (given by the Fisher information), hence Eq. (4) is of key importance.

Furthermore, the stimulus parameter itself may be a random variable, distributed according to the probability density function $p_\theta(\theta)$. The corresponding probability density function $p_\lambda(\lambda)$ in the regularly transformed units is then given by Papoulis (1991)

$$p_\lambda(\lambda) = p_\theta(\varphi^{-1}(\lambda)) \left| \frac{d \varphi^{-1}(\lambda)}{d \lambda} \right|. \quad (5)$$

From Eq. (4) follows that the square root of the Fisher information transforms similarly to Eq. (5). Therefore, if the proportionality

$$p_\theta(\theta) \propto \sqrt{J_\theta(\theta)}, \quad (6)$$

holds in the stimulus units of θ , it holds *invariantly* as $p_\lambda(\lambda) \propto \sqrt{J_\lambda(\lambda)}$ in any regularly transformed units. The particular stimulus distribution given by Eq. (6) is known as the *Jeffreys prior*, an important concept in the Bayesian inference due to the invariance property (Jeffreys, 1946; Kass & Wasserman, 1996).

3. Results and discussion

3.1. Fisher information under regular transformations of the stimulus measurement units

The profile of the Fisher information as a function of the stimulus parameter has been investigated extensively in the computational neuroscience literature. For example, Greenwood, Ward, and Wefelmeyer (1999) employed $J_\theta(\theta)$ to describe the stochastic resonance effect in a generalized McCulloch–Pitts neuronal model. Wilke and Eurich (2002) analyzed the accuracy with which a neural population encodes a number of stimulus features, and found that the stimulus value yielding maximal Fisher information is below the mode of the mean response curve. More generally, the mode of $J_\theta(\theta)$ was employed to identify the optimal stimulation intensity in a number of stochastic (Greenwood & Lansky, 2005a,b) and empirical (Lansky, Pokora, & Rospars, 2008) neuronal models. Optimal coding accuracy in terms of latency (time to first spike after the stimulus onset) was investigated in a similar way (Levakova, 2016; Wainrib, Thieullen, & Pakdaman, 2010). Jenison and Reale (2003) used the Fisher information to quantify the estimation precision of multiple sound localization parameters in the auditory system. And finally, profiles of $J_\theta(\theta)$ and certain information-theoretic measures were compared in Kostal and Lansky (2013) and Yarrow et al. (2012). See also Pilarski and Pokora (2015) for additional references.

In a recent study Kostal and Lansky (2015) demonstrated that, paradoxically, the Fisher information cannot be used to decide whether loud or quiet tones are encoded more precisely in a classical model of the auditory nerve fiber (Winslow & Sachs, 1988). The Fisher information turned out to be an increasing function of the sound intensity evaluated in decibels (dB SPL), and a decreasing function of the same intensity in Pascals.

We now prove that the Fisher information of a given encoding model may in fact follow the shape of any desired function, just by changing the stimulus measurement scale.

Theorem 1. Let $g(x) > 0$, $x \in [0, 1]$ be a Lipschitz continuous function. For any given $J_\theta(\theta)$ there exists a unique regular transformation φ in Eq. (1) and a constant $a > 0$ such that

$$J_\lambda(\lambda) = ag(\lambda), \quad (7)$$

Download English Version:

<https://daneshyari.com/en/article/326323>

Download Persian Version:

<https://daneshyari.com/article/326323>

[Daneshyari.com](https://daneshyari.com)