Signal Detection Theory

Kim Boström

March 21, 2006

Contents

1 Psychophysical Tasks

In a psychophysical task, stimuli are presented to the subject and then the subject is asked for some formerly specified reaction. Psychophysical tasks can roughly be devided into three distinct types:

1. Detection
2. Discrimination
3. Identification

In a detection task, on each trial a stimulus is presented to the subject, and the subject is asked to report whether or not it detects the stimulus. The proportion of detected stimuli is called the *detect rate* and it constitutes the psychometric curve of the task.

In a discrimination task, on each trial one out of n distinct stimuli is presented to the subject, and the subject is asked to report which of the n stimuli it perceives. The proportion of correct answers is called the *proportion correct* and it constitutes the psychometric curve of the task.

In an identification task, on each trial n distinct stimuli are presented to the subject, and the subject is asked to identify them. This type of psychophysical task is also called an *n-alternative forced choice (nAFC)* task. Just like in the discrimination task, the proportion of correct answers is called the *proportion correct* and it constitutes the psychometric curve of the task.

In many discrimination and identification tasks there are $n = 2$ distinct stimuli, which are often called the *signal* and the *nonsignal stimulus*.

The signal detection model is an abstract model of perception that is able to cope with all three types of psychophysical tasks (see Fig. 1).

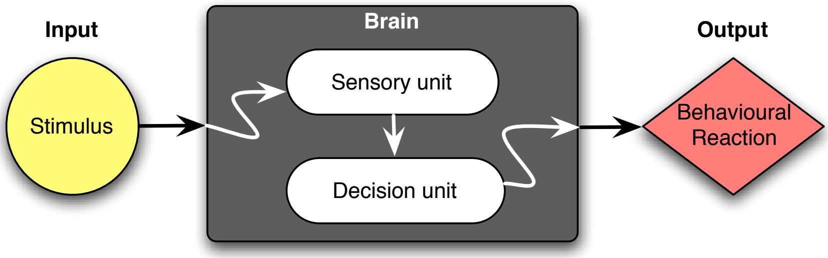

Figure 1: General scheme of the signal detection model.

2 Detection task

In a detection task, the subject is presented on each trial a stimulus and she must then decide whether or not he perceives a specific stimulus property x. For example, the parameter x might be the stimulus' contrast, its motion or its rotation. The stimulus property x is varied within a certain domain on each trial, either randomly or systematically.

Let us first assume that the detection process is perfect. In this case the stimulus property x is encoded by a *sensory unit* in the brain into a neuronal response y_x, which can be the firing rate of a single neuron or the overall activity of a specific neuron population. Let us call the function that maps the stimulus property x to its neuronal response y_x the *transducer function* τ:

$$x \overset{\tau}{\longmapsto} y_x. \tag{1}$$

The neuronal response $y_x = \tau(x)$ is sent to a *decision unit* and translated into a discrete response which is sent to some *motoric unit* to produce a behavioural reaction. We may leave out the motoric unit from our description if it is directly related to the output of the decision unit. In the following we will concentrate on *binary decisions*, so the behavioural response can take one out of two distinct values, for example Yes or No.

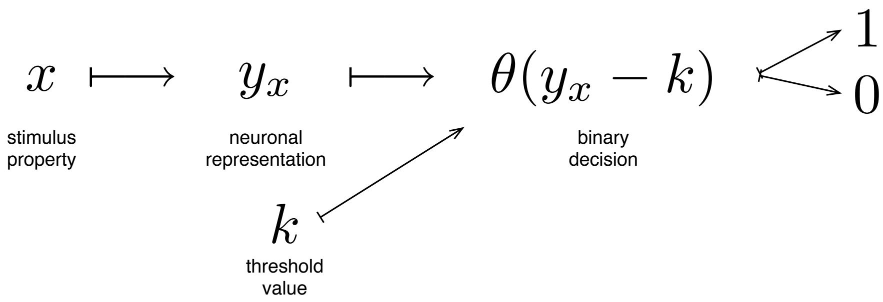

Figure 2: A perfect detection process.

Here is where the notion of a *threshold* enters in a natural way. Whenever the neuronal response y_x lies above a certain *internal threshold* k, the response is positive, otherwise negative. Let us assume that the threshold value is generated by some other internal unit of the brain, the *control unit*, and sent to the decision unit for comparison with the representation y_x. The decision unit maps it onto a behavioural response r which is either 1 or 0, dependent on the condition that the neuronal response is bigger that the threshold or not. Let us call this map the *decision function* ζ:

$$(y, k) \overset{\zeta}{\longmapsto} r. \tag{2}$$

The entire detection process is depicted in Fig. 2.

For fixed threshold k and given input x the response $r_x = \zeta(\tau(x), k)$ reads

$$r_x = \theta(y_x - k), \tag{3}$$

where

$$\theta(x) := \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{4}$$

is the Heaviside step function.

Now let us become more realistic and assume that the detection process is not perfect. We account for the imperfection by adding *noise* to the system. Let us first clarify what we understand by the term "noise" throughout this paper. Noise is not understood as a fluctuation over time or space or both. Instead, noise is represented by a random variable that takes a different value *on each trial*. The average over many trials approaches the expectation value of the random variable. In short: Noise, in the way we use it here, is *random fluctuations over trials*. The degree of these trial-by-trial fluctuations is termed the *variability*.

This being clarified, we now account for imperfections of the sensory unit and assume that the transducer function τ maps the stimulus property x onto a random variable Y_x,

$$x \overset{\tau}{\longmapsto} Y_x. \tag{5}$$

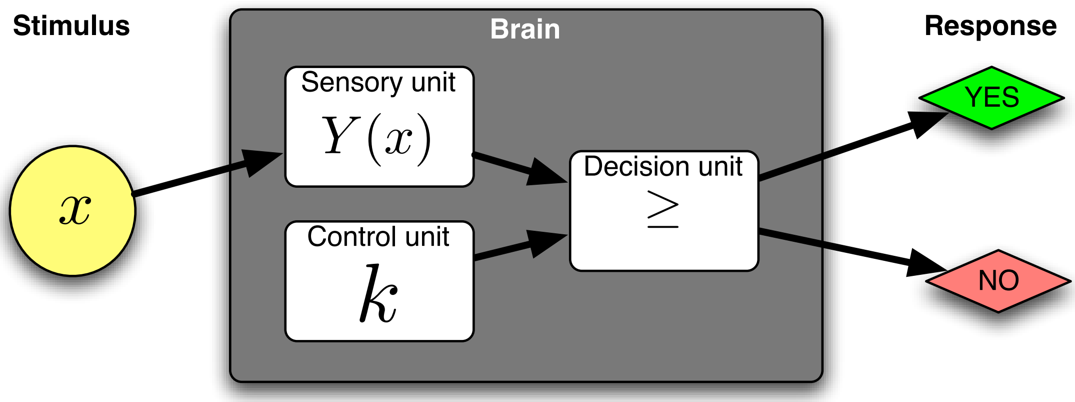

Figure 3: The detection process.

The random variable Y_x has realizations $y \in \mathbb{R}$ which are distributed by an *transducer probability* $q(y|x)$ so that for any $x \in \mathbb{R}$ we have

$$\int dy\, q(y|x) = 1. \tag{6}$$

The function $q(y|x)$ corresponds to the conditional probability of encoding the given input x into the neuronal response y. The above completeness relation means that an input x is encoded into *some* neuronal response. As usual, the expectation value of some function $f(Y_x)$ is obtained by

$$\langle f(Y_x)\rangle = \int dy\, q(y|x)\, f(y). \tag{7}$$

For the sake of simplicity, we will not account for imperfections of the control unit which would correspond to replacing the sharp threshold value k by a random variable K.

The output of the decision unit becomes a random variable

$$R_x = \theta(Y_x - k). \tag{8}$$

The expectation value of R_x equals the fraction of positive responses, and it is called the *detect rate*,

$$D(x) = \langle R_x \rangle, \tag{9}$$

which yields

$$D(x) = \int dy\, q(y|x)\, \theta(y-k) \tag{10}$$

$$= \int_k^\infty dy\, q(y|x). \tag{11}$$

The entire detection process is depicted in Fig. 4 and 3.

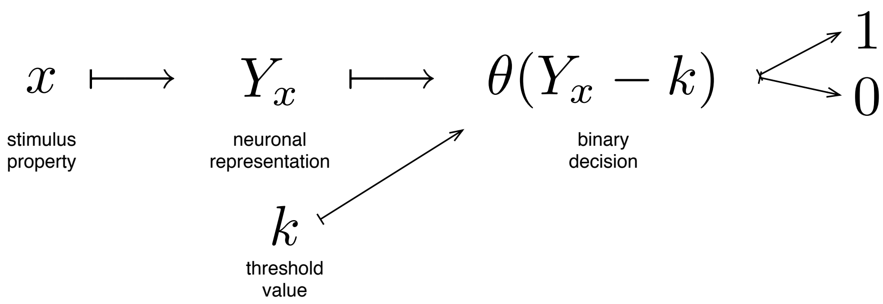

Figure 4: The detection process with an imperfect sensory unit.

3 Discrimation task

In a binary discrimination task, the subject is on each trial presented randomly one out of two distinct stimuli and she must then decide which of the stimuli she perceives. Often, one of the stimuli is called the *signal* and the other one the *nonsignal* stimulus. Let us follow this convention and denote the signal and nonsignal stimulus by s and n, respectively, and the property of the signal stimulus by x_s and that of the nonsignal stimulus by x_n. The perceptual system of the subject will then adapt its internal threshold in such a way that both stimuli can optimally be distinguished. The subject's response is either "s" or "n" corresponding to whether she perceives the signal or the nonsignal stimulus.

3.1 Possible situations

In contrast to the detection task, where there is either a detect or nondetect situation, we now have four different situations:

$$(s|s) = \text{hit} \tag{12}$$
$$(n|s) = \text{miss} \tag{13}$$
$$(s|n) = \text{false alarm} \tag{14}$$
$$(n|n) = \text{correct rejection.} \tag{15}$$

Let us look for the probabilities of these events.

The signal and nonsignal stimuli x_s and x_n are translated by the sensory unit into the neuronal responses Y_s and Y_n, respectively,

$$x_i \mapsto Y_i \quad i = s, n. \tag{16}$$

Let us constitute without loss of generality that the signal stimulus produces the higher average neuronal response,

$$\langle Y_s \rangle \geq \langle Y_n \rangle. \tag{17}$$

Since x_s produces the higher neuronal response Y_s, the subject will answer "s" if the neuronal repsonse happens to lie above the adapted internal threshold k, and otherwise she will answer "n".

Thus, the probability of a hit depends on the internal threshold parameter k and is given by

$$\begin{aligned} P_k(s|s) &= P(Y_s \geq k) && (18)\\ &= \int dy\, q(y|x_s)\, \theta(y-k) && (19)\\ &= \int_k^\infty dy\, q(y|x_s). && (20) \end{aligned}$$

The miss probability is given by

$$\begin{aligned} P_k(n|s) &= P(Y_s < k) = 1 - P(Y_s \geq k) && (21)\\ &= 1 - P_k(s|s), && (22) \end{aligned}$$

so that hit and miss probability add up to unity,

$$P_k(s|s) + P_k(n|s) = 1. \tag{23}$$

The probability of a false alarm reads

$$\begin{aligned} P_k(s|n) &= P_k(Y_n \geq k) && (24)\\ &= \int dy\, q(y|x_n)\, \theta(y-k) && (25)\\ &= \int_k^\infty dy\, q(y|x_n), && (26) \end{aligned}$$

and it is easy to see that the probabilities for false alarm and correct rejection also add up to unity,

$$P_k(s|n) + P_k(n|n) = 1. \tag{27}$$

In plain words:

$$\text{hit rate} + \text{miss rate} = 1 \tag{28}$$

$$\text{false alarm rate} + \text{correct rejection rate} = 1. \tag{29}$$

As we can see there are exactly *two* degrees of freedom. Conventionally, the *hit rate*

$$\mathrm{HR}_k := P_k(s|s) \tag{30}$$

and the *false alarm rate*

$$\mathrm{FAR}_k := P_k(s|n) \tag{31}$$

are chosen as the two parameters that uniquely describe the performance of the subject within a binary discrimination task. The probabilities are estimated by a large sample of trials of either signal or nonsignal stimuli presentation. If the number of an event x is denoted by $N(x)$, then the hit rate and the false alarm rate are estimated by

$$P_k(s|s) \approx \frac{N(s|s)}{N(s)} \tag{32}$$

$$P_k(s|n) \approx \frac{N(s|n)}{N(n)}. \tag{33}$$

3.2 Success probability

The overall performance of the subject is measured by the *success probability*, which is also referred to as the *proportion correct* or *percent correct*. The success probability is obtained in the following way. While $(y|x)$ is the conditional event of y occurring under the condition that x has already occurred, the *joint event* (y,x) denotes the joint occurrence of x and y, and of course we have $(y,x) = (x,y)$. The joint probability $P_k(y,x)$ is related to the conditional probability $P_k(y|x)$ and the *apriori probability* $P(x)$ of x via

$$P_k(y,x) = P_k(y|x)P(x). \tag{34}$$

The success probability is now given by

$$\mathrm{PC}_k = P_k(s,s) + P_k(n,n), \tag{35}$$

which can be decomposed as

$$\begin{aligned}
\mathrm{PC}_k &= P_k(s|s)P(s) + P_k(n|n)P(n) &(36)\\
&= P_k(s|s)P(s) + P_k(n|n)(1-P(s)) &(37)\\
&= P_k(s|s)P(s) + (1-P_k(s|n))(1-P(s)), &(38)
\end{aligned}$$

because $P(s)+P(n) = 1$. Commonly, the experimenter will chose $P(s) = 1/2$, so that

$$\mathrm{PC}_k = \frac{P_k(s|s) + (1-P_k(s|n)}{2}. \tag{39}$$

Substituting the notations $\mathrm{HR}_k = P_k(s|s)$ for the hit rate and $\mathrm{FAR}_k = P_k(s|n)$ for the false alarm rate, we can write the success probability as

$$\mathrm{PC}_k = \frac{\mathrm{HR}_k + 1 - \mathrm{FAR}_k}{2}. \tag{40}$$

3.3 ROC curve

We see that almost all the probabilities depend on the internal threshold k. For a fixed k there will be a fixed pair of values $(\mathrm{HR}_k, \mathrm{FAR}_k)$ that uniquely describe the performance of the subject in the given discrimination task. If we now urge the subject towards a positive or negative response, we will obtain a *response bias* which has its origin in a shift of the internal threshold k. If we then plot the points $(\mathrm{HR}_k, \mathrm{FAR}_k)$ for different response biases, and therefore for different (but generally unknown) values of k, then we obtain the *Receiver Operating Characteristic (ROC) curve*:

$$\mathrm{ROC} := \{(\mathrm{HR}_k, \mathrm{FAR}_k) \mid k\}. \tag{41}$$

A typical ROC curve is shown in Fig. 5.

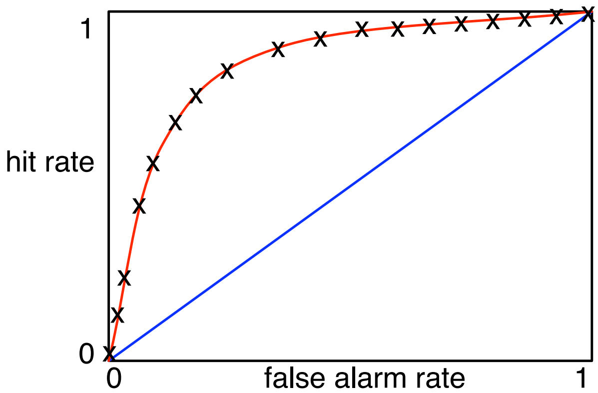

Figure 5: Typical ROC curve in a binary discrimination task.

Assume that the subject is completely blind. Whatever the stimulus, the response of the subject will be totally random. Given the subject is initially unbiased, it will respond "s" and "n" with equal probability $1/2$. Thus, the hit rate and the false alarm rate will both be $1/2$. If we now urge the subject towards a positive response, she will hit the signal stimulus by chance more often, so that $\mathrm{HR} = 0.7$, say. But the false alarm will rise in exactly the same way, hence $\mathrm{FAR} = 0.7$. So whatever the response bias, the hit rate and the false alarm rate will be of equal magnitude, because the blind subject cannot discriminate between signal and nonsignal stimulus. The ROC curve corresponding to a blind subject will be a diagonal (blue line in Fig. 5). The *area under the curve (AUC)* of this diagonal reads 0.5. If the subject is not totally blind, the performance will be a little better, so the hit rate will lie slightly above the false alarm rate. As a consequence, the area under the curve will be bigger than 0.5. Is the subject an ideal observer, then the hit rate will constantly be unity for all false alarm rates, and the corresponding area will also become unity. Thus, the area under the ROC curve is a good measure for the overall performance of the subject *independently* from the internal threshold k:

$$\mathrm{AUC} := \int \mathrm{ROC}. \tag{42}$$

A family of ROC lines for different performances is depicted in Fig. 6.

In the next section we will relate the AUC to the success probability of a corresponding identification task, namely the two-alternative forced choice task.

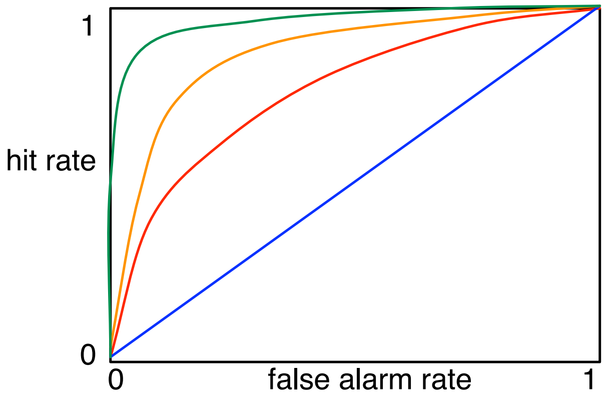

Figure 6: Typical ROC curves for different levels of performance.

4 Identification task

In a binary identification task the subject is on each trial presented two stimuli and she must then indicate which stimulus is which. This kind of task is also referred to as a *Two-alternative forced choice (2AFC)* task. As in the previous section, let us call one stimulus the *signal* stimulus and the other the *nonsignal* stimulus.

In contrast to the discrimination task, where there are four different situations, in the 2AFC there are only 2 situations: A *correct* and a *false* choice.

As in the discirmination task, the signal and nonsignal stimuli x_s and x_n are translated by the sensory unit into the neuronal responses Y_s and Y_n, respectively,

$$x_i \mapsto Y_i \quad i = s, n. \tag{43}$$

Let us again constitute without loss of generality that the signal stimulus produces the higher average neuronal response,

$$\langle Y_s \rangle \geq \langle Y_n \rangle. \tag{44}$$

Consequently, the subject will choose the stimulus with the highest neuronal representation as the signal stimulus, so that the probability of a correct choice reads

$$\mathrm{PC} = P(Y_s \geq Y_n) \tag{45}$$

$$= \int dy \int dy' \, \theta(y - y') q(y|x_s) q(y'|x_n). \tag{46}$$

Now it is fairly easy to formally combine the above expression with the *hit rate* of a binary discrimination task. Along (20) the hit rate $h(k) := \mathrm{HR}_k$ is given by

$$h(k) = \int dy \, \theta(y - k) q(y|x_s). \tag{47}$$

Substituting this expression into (46) gives

$$\mathrm{PC} = \int dk\, q(k|x_n)h(k), \tag{48}$$

where we have replaced y' by k. Now because the false alarm rate $f(k) := \mathrm{FAR}_k$ is given by

$$f(k) = \int_k^\infty dy\, q(y|x_n) \tag{49}$$

$$= 1 - \int_{-\infty}^k dy\, q(y|x_n) \tag{50}$$

we have

$$\frac{df}{dk} = -q(k|x_n), \tag{51}$$

and therefore

$$df = -q(k|x_n)\, dk. \tag{52}$$

The limits are given by

$$f(-\infty) = 1, \quad f(\infty) = 0, \tag{53}$$

so that a substitution of (52) and (53) into (48) yields

$$\mathrm{PC} = \int_0^1 df\, h(f), \tag{54}$$

which is nothing but the area under the ROC curve! Altogether we find that the success probability PC in a two-alternative forced choice paradigm corresponds to the area under the ROC curve of the corresponding discrimination task,

$$\mathrm{PC} = \mathrm{AUC}. \tag{55}$$

This is a well-known important result of signal detection theory. The significance of the result is given by the fact that one can now *compare* the performance of a subject in a discrimination task with its performance in a corresponding 2AFC task. If the subject had an AUC of, say, 0.7 in a given discrimination task, then one knows that her performance would be $\mathrm{PC} = 0.7$ in the corresponding identification task, where the signal and nonsignal stimulus are both presented in each trial. Moreover, it shows that it is *sufficient* to perform a 2AFC instead of the more time-consuming discrimination task, if one wants to determine the overall performance of the subject. A discrimination task is more time-consuming and more difficult, because it demands a bunch of trials to be performed for several values of the internal threshold k which is not directly accessible, in order to get the ROC curve. On the other hand, a 2AFC is approximately as time-consuming as the discrimination task for one fixed threshold value k. Thus, a 2AFC is much more *efficient* than a discrimination task.