PERGAMON

Vision Research 42 (2002) 2371–2393

Vision Research

www.elsevier.com/locate/visres

Additive noise, Weibull functions and the approximation of psychometric functions

U. Mortensen

Westfälische Wilhelms-Universität Münster, Fachbereich Psychologie und Sportwissenschaften, Institut III, Fliednerstr. 21, D-48149 Münster, Germany

Received 11 September 2001; received in revised form 2 July 2002

Abstract

The Weibull function is frequently chosen to define psychometric functions. Tyler and Chen (Vis. Res. 40 (2000) 3121) criticised the high-threshold postulate implied by the Weibull function and argued that this function implies the assumption of multiplicative noise. It will be shown in this paper that in fact the Weibull function is compatible with the assumption of additive noise, and that the Weibull function may be generalised to the case of detection not being high threshold. The derivations rest, however, on a representation of sensory activity lacking a satisfying degree of generality. Therefore, a more general representation of sensory activity in terms of stochastic processes will be suggested, with detection being defined as a level-crossing process, containing the original representation as a special case. Two classes of stochastic processes will be considered: one where the noise is assumed to be additive, stationary Gaussian, and another resulting from cascaded Poisson processes, representing a form of multiplicative noise. While Weibull functions turn out to approximate well psychometric functions generated by both types of stochastic processes, it also becomes obvious that there is no simple interpretation of the parameters of the fitted Weibull functions. Moreover, corresponding to Tyler and Chen's discussion of the role of multiplicative noise particular sources of this type of noise will be considered and shown to be compatible with the Weibull. It is indicated how multiplicative noise may be defined in general; however, it will be argued that in the light of certain empirical data the role of this type of noise may be negligible in most detection tasks.

1. Introduction

The assumption of additive noise that is independent of the stimulus generated activity is in most cases adopted without further discussion. However, Tyler and Chen (2000) argued that this assumption is incompatible with the Weibull function (or simply Weibull, for short)

$$\psi_{\mathrm{w}}(c) = 1 - \exp(-\alpha c^{\beta}), \quad \alpha > 0, \ \beta > 0, \ c \geqslant 0, \tag{1}$$

where $\psi_{\mathrm{w}}(c)$ is the probability of detecting the stimulus when the stimulus contrast or intensity equals c, and α and β are free parameters. Referring to a characteristic property of ψ_{w}, namely being of constant shape for different values of α but constant value of β when plotted the $\log c$-scale, the authors argue that the Weibull implies the noise to be multiplicative instead. The purpose of this paper is

- to show that the Weibull is compatible with the assumption of additive as well as with certain forms of multiplicative noise, and that the Weibull can be generalised to cater for detection processes that are not high threshold.
- to argue that the main problem connected with the Weibull, when fitted to detection data, refers to the interpretation of its parameters, since the Weibull approximates well psychometric functions defined with respect to a variety of detection mechanisms that are not necessarily high threshold.

Since Quick (1974), the Weibull has become almost the standard definition of a psychometric function when probability summation or nonlinear pooling (NP) effects are to be modelled. Quick [1] showed that, provided (i) the noise is representable by Gaussian random variables

E-mail address: mortens@psy.uni-muenster.de (U. Mortensen).

[1] Quick's approximations are numerical; he actually discussed the function $\psi_{\mathrm{q}}(c) = 1 - 2^{-\tilde{\alpha}c^{\beta}}$, which is equivalent to the Weibull if $\alpha = \tilde{\alpha}\log_{e} 2$ since then $2^{-\tilde{\alpha}c^{\beta}} = \mathrm{e}^{-\alpha c^{\beta}}$ for all $c \geqslant 0$. From a theoretical point of view (extreme value statistics) one would expect an approximation in terms of the double exponential $\exp(-\exp(-(ax+b)))$. Quick's finding rests on the fact that the Weibull yields, for all practical purposes, also good approximations to the double exponential.

PII: S0042-6989(02)00195-5

independent from the responses of the neural mechanisms ("channels") involved in the detection task, (ii) temporal stochastic effects are negligible, and (iii) detection is by probability summation among neural mechanisms, the psychometric function may be approximated by the Weibull function [2] (1). The free parameter α is interpreted as summarising the effect of the stimulus, i.e. the responses of neural mechanisms ("channels"), and the free parameter β is taken to reflect the effect of noise. Examples for the type of neural mechanisms that may be considered are spatial frequency channels (e.g. Graham, 1977, 1989), neurons with Gabor-type receptive fields (Daugman, 1980; du Buf, 1992, 1993, 1994; Marčelja, 1980), or neurons with difference-of-gaussian-type receptive fields (Wilson & Bergen, 1979). Even temporal probability summation effects have been modelled in terms of the Weibull (Blommaert & Roufs, 1987, Watson, 1979).

The Weibull is usually adopted because ψ_w is computationally convenient, in particular when probability summation effects are to be modelled, and because this function often fits data well. Quick's finding that this function serves as an excellent approximation to a psychometric function for the case of probability summation among channels with Gaussian distributed activities or NP of such activities provides some theoretical justification for the choice of ψ_w, provided one accepts the hypothesis that the sensory activity can be represented by the random variable $\eta = g + \xi$, with g a constant reflecting the effect of the stimulus and ξ a Gaussian random variable representing independent noise. So the Weibull serves as an approximation to some other, not precisely known function. Indeed, it seems that the Weibull has not yet been derived from neurophysiological or psychophysical results. Still, the Weibull is often taken as a kind of basis for particular deterministic models of stimulus processing, catering for the stochastic effects in detection and discrimination which are tacitly assumed to be independent of the properties of the deterministic model. It will be demonstrated in Section 4 that this assumption cannot necessarily been taken for granted.

1.1. Overview

In Section 2, first a general definition of the type of psychometric functions considered in this paper will be provided (Section 2.1); according to this definition, psychometric functions are specified in terms of distribution functions of random variables representing the neuronal activity underlying the detection process. Other conceptualisations of psychometric functions are possible, but will not be discussed in this paper. In particular, it will be assumed that the sensory activity can be represented by the random variable $\eta = g + \xi$ introduced above, and the distribution and the density function of η and consequently of ξ will be derived for the case that the psychometric function is given by ψ_w. The result will be applied to detection by probability summation on the one hand and by NP of channel activities on the other. An explicit specification of these two modes of detection will be given in Section 2.2.

In Section 2.3 the distribution function of the noise, corresponding to the Weibull, will be derived. From this the distribution function of the maximum $\eta = \max(\eta_1, \ldots, \eta_n)$ for the case of detection by probability summation among nonidentically activated channels will be derived. Also, the distribution function for the case of detection by NP of channel activities will be given. Tyler and Chen argue that the assumptions of detection by probability summation and additive noise in conjunction with the Weibull imply the need for negative values of c. Since negative intensities c do not exist the authors conclude that the Weibull is incompatible with the assumption of additive noise. On the basis of the results derived in Section 2.3 it will be shown that this conclusion is false.

In Section 3.1, the results concerning the distribution function of the noise will be employed to relax the high-threshold assumption that is implicit in the definition of ψ_w. In Section 3.2 it will be shown how a generalised version of ψ_w may be applied to 2AFC data. Unfortunately, only the case of identically activated channels can be treated analytically. Tyler and Chen concentrate, in their work, on this special case as well, aiming at a relation between parameters of the psychometric function and the number n of channels involved in a detection task; however, it may be doubted whether the assumption of identically activated channels is sufficiently realistic to allow conclusions concerning the value of n. The main purpose of this section is therefore to illustrate the principal applicability of the (generalised) Weibull to 2AFC data; to discuss real data, the case of nonidentically activated channels will have to be dealt with numerically.

In Section 2.3.5 psychometric functions that are parallel on $\log c$-scales will be considered; such functions will be called log-parallel for short. Given that detection is by probability summation, Tyler and Chen deduce from the log-parallelism of the Weibull that this function implies the noise to be multiplicative; it is easy to show that this conclusion is false.

The representation of sensory activity by a random variable $\eta = g + \xi$ is more often taken for granted than discussed. This is remarkable, since activity is a process in time, but η does not explicitly refer to time. Tyler and

[2] A more general definition of ψ would be $\psi(c) = 1 - (1 - \gamma)\exp(-\alpha c^{\beta})$, where γ is the probability of guessing that a stimulus was presented although none was shown. This correction for guessing will be neglected to keep the notation as simple as possible.

Chen call [3] η the "instantaneous internal response". The authors do not provide any definition of what they mean with this expression and leave it to the reader's intuition to find an interpretation. So, in Section 4, the relation between activity as a time-extended stochastic process and its representation by a random variable without reference to time will be discussed. Two types of stochastic processes will be considered, one assuming continuous sample paths, referring e.g. to spike rates, and one referring to counting processes, characterised by sample paths that are constant between discontinuous "jumps". It will be argued that the maximum of a sample path during a trial appears to be a reasonable choice for a random variable with respect to which a psychometric function may be defined; in Section 4.2 in particular the random variable $\eta = \max_{t\in(0,T]}[g(t) + \xi(t)]$ will be considered, where $(0, T]$ denotes a trial of duration T, g is the mean value function of the process and ξ is a sample path of the noise. Detection is then specified as a level-crossing process (the stimulus is detected if a sample path crosses a threshold level at least once in a trial). In general, the derivation of the distribution function for such a variable represents a formidable task, but for the case of continuous sample paths and stationary Gaussian noise a handsome approximation will be given, the validity of which has been explored before by Mortensen and Suhl (1991). It will be demonstrated that the Weibull fits well to data generated by such level-crossing processes, even if the high-threshold assumption was not made for these processes. Mortensen et al. (1991) also characterised the conditions under which the approximation

$$\eta \approx g_{\max} + \xi, \quad g_{\max} = \max_t g(t) = g(t_0), \quad \xi = \xi(t_0) \tag{2}$$

holds; t_0 is the time at which g assumes its maximum. This approximation will be referred to as "peak-detection" in the following. According to this assumption, the activity can be represented by the "peak", i.e. the maximal value $g_{\max}$ of g. So, the representation $\eta = g + \xi$ may be interpreted as reflecting the approximation (2), which may also be taken as an interpretation of Tyler and Chen's expression "instantaneous internal response", because η is then characterised by the response in the neighbourhood of t_0.

In Section 4.3 a counting process defined in particular as multiplicative stochastic processes, also known as branching or cascaded processes (Parzen, 1962, p. 58), will be considered; such processes arise when certain random events generate further random events. McGill (1967) and Teich, Prucnal, Vannucci, Breton and McGill (1982) assumed such processes in models of simple detection tasks. The distribution of spikes resulting from two cascaded Poisson processes is given by the Neyman-Type A (NTA) distribution (Neyman, 1939). [4] η will also be defined as the maximum of a sample path of such a process, and a psychometric function defined with respect to the NTA distribution will be compared to the Weibull. Again, the Weibull turns out to provide very good approximations to such psychometric functions.

In Section 5 a summarising discussion of the results will be given. Tyler and Chen explore the effects of a certain type of multiplicative noise, and so the general characterisation of multiplicative noise will briefly be considered; the cascaded Poisson process mentioned in the preceding paragraph is a special case of stochastic processes defining multiplicative noise. It turns out that while the most general characterisation of multiplicative noise results from a representation of sensory activity in terms of stochastic differential equations, a corresponding modelling of the activity in terms of stochastic differential equations is mathematically quite demanding, and in the light of certain data one may argue that the effects of multiplicative noise could be negligible; it seems that for sufficiently brief stimulus presentations (not much longer than 500 ms) experimental data may safely be interpreted on the basis of the simple postulate of peak detection, implying that the assumption additive noise is sufficient to capture the main features of stimulus processing.

2. Additive noise and the Weibull function

2.1. Psychometric functions

Generally, a psychometric function is a mapping [5] $\psi : M_c \rightarrow (0, 1)$, where $M_c \subset \mathbb{R}$ is a set of real numbers representing contrasts. For $c \in M_c$, $c \mapsto \psi(c) \in (0, 1)$, and $\psi(c)$ is the probability of detection if the contrast of the stimulus equals c. As usual, we assume $d\psi/dc \geqslant 0$ for all $c \in M_c$, i.e. the psychometric function is assumed to be a nondecreasing function of c.

For the purposes of this paper psychometric functions will be further defined in terms of a distribution function for η, where η may be defined as in (5) or in (8). The psychometric function, i.e. the probability of detection given the contrast c, is then given by

$$\psi(c) = 1 - (1 - p_{\text{fa}})P(\eta \leqslant \eta_s | c), \quad 0 < \psi(c) < 1, \quad c \geqslant 0, \; p_{\text{fa}} \geqslant 0, \tag{3}$$

[3] They write r instead of η, though.

[4] I am indebted to A. Reeves for pointing out to me the role of cascaded Poisson processes, as well as the work of Teich et al. (1982).

[5] $(0, 1)$ is the "open" interval defined by $\{x | 0 < x < 1\}$; the "closed" interval $[0, 1] = \{x | 0 \leqslant x \leqslant 1\}$ is not meaningful here since the maximal value of c for which $\psi(c) = 0$ or the minimal value of c for which $\psi(c) = 1$ may not have an empirical meaning; for instance for the Weibull function (9), $\psi(c) \rightarrow 1$ for $c \rightarrow \infty$.

where either $\eta = \eta_{ps}$ or $\eta = \eta_{np}$, and p_{fa} is the probability of guessing (false alarm) that a stimulus was presented if, in a particular trial, $\eta < \eta_s$. To simplify the expressions it will be assumed for the following that $p_{fa} = 0$, since the role of guessing is of no importance to the issues discussed here. The definition of ψ in (3) allows for "true" false alarms, i.e. for detection responses based on the event $\{\eta > \eta_s\}$ although no stimulus was presented. High-threshold models can be derived from (3) as special cases.

In the following, the density function corresponding to the Weibull function given the noise is additive will be derived. To avoid confusion with concepts introduced by Tyler and Chen, we employ a different notation. In particular, we stick to expressions that are generally used in the statistical literature: $F(x) = P(X \leqslant x)$ as a function of x will be called the distribution function, and the derivative $f(x) = \mathrm{d}F(x)/\mathrm{d}x$ will be called the (corresponding) density function. The expression probability distribution instead of density function will only be used with respect to discrete, e.g. Poisson, random variables.

2.2. Activation and detection

Let us assume that altogether n channels $C_1, C_2, \ldots, C_n$ are involved in the detection task. Let g_i be the response of the channel C_i to a stimulus. The channel is supposed to be linear so that $g_i = ch_i$, c contrast or intensity and h the "unit response", i.e. the response of the channel for $c = 1$; usually, h_i is given as convolution of the stimulus with the impulse response of the channel (more precisely, of the maximum value of the convolution within the experimental trial, see Section 4).

One may distinguish between at least two modes of interaction among the channels: probability summation and pooling. If detection is by probability summation, the stimulus is detected if at least one of the channels detects the stimuli; this assumption implies that there exists noise in each individual channel. If detection is by pooling, the responses of the channels are somehow, deterministically and usually nonlinearly summed, and only the pooled activity is contaminated by noise, so there is only a single noise variable. Clearly, both modes are idealisations. An also somewhat idealised representation of neural activity by random variables may be summarised as follows:

A1. *Probability summation*

1. *Representation of activity:* The activation in any channel C_i can be represented by a random variable
$$\eta_i = g_i + \xi_i, \tag{4}$$
where g_i represents the stimulus generated activity and is, for a given value of intensity or contrast c, a constant, and ξ_i is a random variable representing the noise in the channel; the distribution of ξ_i is independent of g_i.
2. *Detection:* Let, in a given trial, $\eta_{ps} = \max[\eta_1, \ldots, \eta_n]$ be the maximum of the random variables η_i, and let the η_i be stochastically independent. Suppose the stimulus is detected if the event
$$\{\eta_{ps} > \eta_s\} \tag{5}$$
occurs, i.e. if for at least one channel the event $\{\eta_i > \eta_s\}$ is observed, where η_s represents a critical activity representing a threshold. Then detection is said to be by probability summation (PS) among the channels. [6]

A2. *Nonlinear pooling*

1. *Representation of activity:* If detection is by NP, the pooled stimulus generated activity is given by
$$g = \left(\sum_{i=1}^{n} g_i^p\right)^{1/p} = c\left(\sum_{i=1}^{n} h_i^p\right)^{1/p}, \quad p > 0 \tag{6}$$
i.e. g is proportional to the intensity or contrast c. The total activity is defined as
$$\eta_{np} = g + \xi, \tag{7}$$
with ξ a random variable representing the noise; ξ is assumed to be independent of g.
2. *Detection* The stimulus is detected if the event
$$\{\eta_{np} > \eta_s\} \tag{8}$$
occurs, where η_s is again a critical activity representing a threshold.

Remarks

1. Eq. (4) in A1-1 and (7) in A2-1 represent specific forms of the assumption that the noise is additive. The assumption that ξ_i and ξ are independent of g_i and g, respectively, imply that the variances of η_i or η, $\sigma^2_{\eta_i} = \sigma^2_{\xi_i}$ or $\sigma^2_{\eta} = \sigma^2_{\xi}$, are independent of the g_i or of g. According to some neurophysiological findings $\sigma^2_{\eta_i}$ or σ^2_{η} may not be independent of g_i or g (see also Section 4). Tyler and Chen (p. 3129), declare that "if the noise distribution is

[6] Tyler and Chen connect the concepts of PS and high-threshold detection in a somewhat unconventional way: "The goal of High Threshold Theory is to define properties of summation over independent channels, which has come to be known as probability summation" (p. 3123); however, PS is not the "goal" of high threshold theory (HTT), which Tyler and Chen attribute to Quick (1974); HTT was discussed before Quick and without reference to PS (cf. Luce, 1963). Further, the—admittedly awkward and somewhat metaphorical expression "probability summation"—cannot reasonably be interpreted as representing a "property of summation over (·) channels", unless one explains what this property is meant to be.

Poisson rather than Gaussian the noise is no longer additive but varies with the mean level...". At this level of generality, this statement is wrong (the sum of two Poisson variables is again Poisson, and the fact that mean and variance are always identical for such variables can therefore not be counterindicative of additivity, see, e.g., Papoulis, 1965); however, the situation is different if the two processes are cascaded Poisson processes, as discussed by McGill (1967) and Teich et al. (1982). Here the variance turns out to be in excess of the mean, and this excess is indicative of the nonadditivity of the noise.

2. An alternative to additive noise is multiplicative noise. In the Discussion (Section 5) it will be shown that writing $g\xi$ is one possibility of many to represent multiplicative noise, and a distribution function for ξ will be given that implies the psychometric function to be the Weibull; this shows that the Weibull is compatible with the assumptions of either additive or multiplicative noise.

3. The assumption of stochastically independent channels in A1-2, equivalent to stochastic independence of the η_i and therefore of the noise variables ξ_i, is a simplification that can be relaxed for large values of n; then it is sufficient to postulate asymptotic independence, meaning that among neighbouring channels almost arbitrary dependencies may exist and only sufficiently far apart channels are independent (cf. Galambos, 1978).

4. Tyler and Chen introduce the expression "attentional summation" as being more meaningful than the term PS, which is common in visual psychophysics, but not standard in the statistical literature. Formally, however, there is no difference between attentional and PS and in this paper we will stick to the latter because we will be mainly concerned with the formal aspects of PS, and not with the conceptual differentiations between probability and attentional summation.

5. The particular form of NP defined in (6) has been chosen with respect to the discussion of the Weibull function; certainly, other forms of pooling the responses of different channels are possible. General neural network models (cf. Rolls & Deco, 2002) will provide theoretically more interesting interpretations of the notion of pooling; however, a simple relationship to the Weibull will be difficult to derive within the framework of such models, so a discussion of such an approach will be left to future research.

2.3. Additive noise and the Weibull function

2.3.1. The distribution function of the noise

In this section, the distribution of the random variable ξ, representing noise, will be derived if the psychometric function for a yes–no task is given by the Weibull function.

Theorem 1. *Suppose the activation and detection processes can be characterised as in assumptions A1 and A2 and detection is investigated in a yes–no task. Let* PS *stand for* PS *among channels with stochastically independent noise, represented by the random variables ξ_i, and* NP *for* NP. *The psychometric function is given by the Weibull function*

$$\psi(c) = \psi_{\mathrm{w}}(c) = 1 - \exp(-\alpha c^{\beta}), \quad \alpha > 0,\ \beta > 0 \tag{9}$$

with

$$\alpha = \begin{cases} \sum_{i=1}^{n} h_i^{\beta}, & \text{if detection is by } PS, \\ \left(\sum_{i=1}^{n} h_i^{p}\right)^{\beta/p}, & \text{if detection is by } NP, \end{cases} \tag{10}$$

and $h_i \geqslant 0$ for all i, if and only if the distribution function of either the ξ_i, $i = 1, 2, \ldots, n$, in case of detection by PS, or of ξ if the activity is by NP, is of the type defined by

$$F(x) = P(\xi \leqslant x) = \begin{cases} 1, & x > \eta_0 \\ \exp(-(\eta_0 - x)^{\beta}), & x \leqslant \eta_0. \end{cases}, \quad \beta > 0, \tag{11}$$

and

$$\eta_s = \eta_0. \tag{12}$$

Proof. See Appendix A.1.

Remarks

1. Note that $h_i \geqslant 0$ has been assumed as a condition instead of simply writing $|h_i|$, as is common when the Weibull is assumed. Writing $|h_i|$ allows for negative values of h_i; however, the theoretical background for doing this is unclear. One interpretation could be that tacitly a rectifier is postulated. With respect to the notion of peak detection (cf. (2) in Section 1) the assumption $h_i \geqslant 0$ appears to be more natural. Peak detection may be seen as an approximation of detection by level-crossing (see Section 4), but there the dependency upon $|h_i|$ does not follow.

2. The distribution function (11) is also known as Weibull-type distribution (Johnson, Kotz & Balakrishnan, 1994; Kotz & Nadarajah, 2000), defined as

$$F(x) = \exp(-((\eta_0 - x)/\sigma)^{\beta}), \quad x \leqslant \eta_0. \tag{13}$$

η_0 is obviously a location parameter; in Section 2.3.4, η_0 will be further commented upon with respect to the common assumption that the noise has a mean (i.e. expected) value equal to zero, making use of results concerning the expected value and the variance of η_{ps} on the one hand, and of ξ on the other. According to (13), there is another free parameter, σ. That is why we refer to the distribution of ξ as being "of the type" (11). The value of σ will influence the value of the variance of the random variables η_{ps}, η_{np} and ξ, but σ^2 is not equal to the variance of ξ. It turns out that neither σ nor η_0 can be estimated from psychometric functions. However, while η_0 plays a central role in the interpretation of the Weibull function, σ is just a scale parameter of almost no importance for this interpretation, so without loss of generality σ was set equal to 1 in (11).

3. If the psychometric function is given by the Weibull function and if detection is by PS, then the psychometric function is given by $\psi_w(c) = 1 - \exp(-\sum_i g_i^\beta)$, and if detection is by NP, it is given by $\psi_w(c) = 1 - \exp(-g^\beta)$, with $g = (\sum_i g_i^p)^{1/p} = c(\sum_i h_i^p)^{1/p}$. Thus the notation ψ_w does not distinguish between detection by PS or NP, although in case of NP one may have $p \neq \beta$; an example of this case was provided by Meinhardt (1999, 2000) (see Section 5). It may be noted that detection by NP does not require the Weibull-type distribution. However, in this paper, we concentrate on (11).

2.3.2. The density function for detection by probability summation

The distribution function of the maximum

$$\eta_{ps} = \max[\eta_1, \eta_2, \ldots, \eta_n], \quad \eta_i = \xi_i + g_i, \quad 1 \leqslant i \leqslant n$$

is given by

$$\begin{aligned} P(\eta_{ps} \leqslant y) &= G(y) \\ &= P(\xi_1 \leqslant y - g_1 \cap \cdots \cap \xi_n \leqslant y - g_n) \\ &= \prod_{i=1}^{n} F_\xi(y - g_i) \end{aligned} \tag{14}$$

and from (11) one has

$$G(y) = \exp\left(-\sum_i (\phi_i(y))^\beta\right), \quad -\infty < y < \infty, \tag{15}$$

where

$$\phi_i(y) = \begin{cases} \eta_s + g_i - y, & y \leqslant g_i + \eta_0, \\ 0, & y > g_i + \eta_0, \end{cases} \quad \eta_s = \eta_0. \tag{16}$$

The condition $\phi_i(y) = 0$ corresponds to the case $x > \eta_0$ in (11), implying $F(x) = 1$, which is equivalent to replacing $\eta_0 - x$ by 0. Note that because of (16) the random variable y is defined on $(-\infty, \infty)$.

The corresponding density function of $y = \eta_{ps}$ is then given by

$$f_{\eta_{ps}}(y) = \frac{dG(y)}{dy} = \beta G(y) \sum_{i=1}^{n} (\phi_i(y))^{\beta-1} \quad \text{for all } i. \tag{17}$$

Fig. 1 shows examples of densities and corresponding distributions for nonidentically channels; since this is just an illustration of how the density changes from $n = 30$ to 100, it was assumed that $h_i = \lambda_0 \exp(-\lambda i)$, $i = 1, \ldots, n$, where $\lambda_0 > 0$ and $\lambda > 0$ were chosen such that $\lambda_0 \exp(-\lambda n) \approx 0$.

For the special case of identically activated channels $g_1 = g_2 = \cdots = g_n$ one has

$$f_{\eta_{ps}}(y) = n\beta(\phi(y))^{\beta-1} \exp(-n(\phi(y))^\beta), \quad g_0 = g_i, \ 1 \leqslant i \leqslant n. \tag{18}$$

with

$$\phi(y) = \begin{cases} \eta_s - (y - g_0)^{\beta-1}, & y \leqslant g_0 + \eta_0, \\ 0, & y > g_0 + \eta_0. \end{cases} \tag{19}$$

The density function of the individual noise variables ξ_i follows immediately as

$$f_\xi(\xi) = \beta(\eta_0 - \xi)^{\beta-1} \exp(-(\eta_0 - \xi)^\beta), \quad -\infty < \xi \leqslant \eta_0. \tag{20}$$

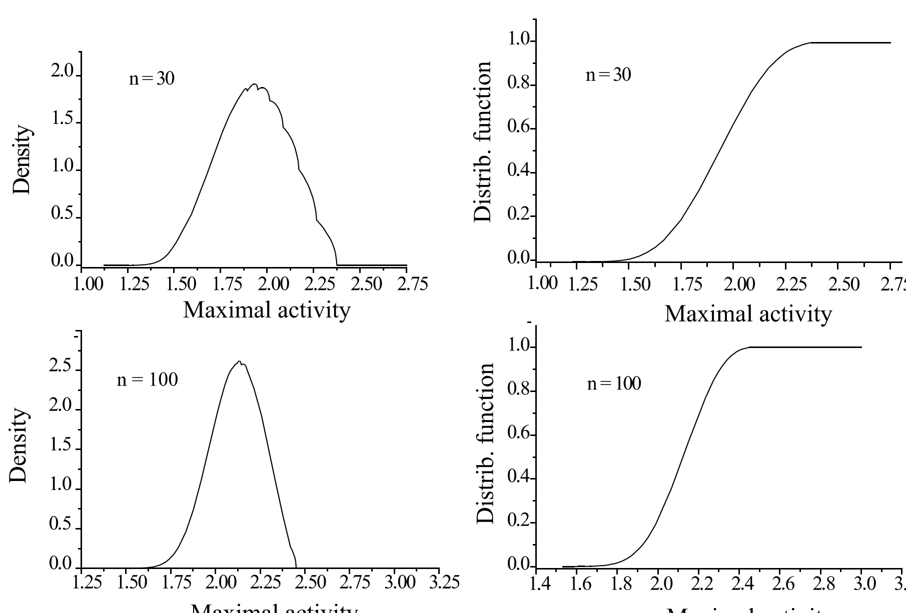

Fig. 1. Density and corresponding distribution functions for nonidentically activated channels.

Note that while y, representing a value of the activity the maximally activated channel, is defined on $(-\infty, \infty)$, the range of possible values of ξ is limited from above by the value of η_0, i.e. the noise can never exceed a certain value. From a physiological point of view this represents no restriction since the spike rate is limited from above; in this respect, the choice of a Weibull-distributed random variable is even more plausible than the choice of a Gaussian variable. On the other hand, a Weibull-distributed random variable has no finite lower limit. Therefore, the Weibull is necessarily an approximation.

For later reference, in particular in Section 2.3.4, the expressions for some expected values and variances will be given. For the case of identically activated channels one finds for the expected value and the variance of η_{ps}

$$E(\eta_{ps}) = g_0 + \eta_0 - (1/n)^{1/\beta}\Gamma(1 + 1/\beta), \tag{21}$$

$$\mathrm{Var}(\eta_{ps}) = (1/n)^{2/\beta}(\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta)), \tag{22}$$

where Γ denotes the gamma-function, i.e. $\Gamma(z) = \int_0^\infty x^{z-1}\mathrm{e}^{-x}\,\mathrm{d}x$. For the case of nonidentically activated channels no closed expression could be derived.

Proof. See Appendix A.2.

The expected value and the variance of the noise variables follow from (21) and (22) putting $g_i = 0$ for all i and $n = 1$:

$$E(\xi) = \eta_0 - \Gamma(1 + 1/\beta), \tag{23}$$

$$\mathrm{Var}(\xi) = \Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta). \tag{24}$$

The function $\Gamma(x)$ is an increasing function of x. It follows that the smaller the value of β, the larger the value of $\Gamma(1 + 1/\beta)$, i.e. the smaller the value of $E(\xi)$. For integers, Γ satisfies the relation $\Gamma(n + 1) = n!$, so that $\Gamma(1) = 0! = 1$. It follows that for an increasing value of β, $\Gamma(1 + 1/\beta) \to \Gamma(1) = 1$, so that $E(\xi) \to \eta_0 - 1$ and $\mathrm{Var}(\xi) \to 0$, i.e. the larger the value of β the smaller the variance of the noise; this corresponds to the well known fact that the psychometric function is the steeper the larger the value of β.

The often encountered expression "steepness parameter" for β is, however, slightly misleading since β is the sole parameter determining the steepness of the psychometric function only if ψ_w is plotted against $z = \log c$; if plotted against c, the steepness of ψ_w increases with the value of α as well (this becomes obvious when one considers the derivative $\mathrm{d}\psi_w(c)/\mathrm{d}c$). From (10) it follows that the value of α depends upon n, the number of channels involved. For the special case of identically activated channels one sees from (21) and (22) that for increasing value of n, i.e. $n \to \infty$, $E(\eta_{ps}) \to g_0 + \eta_0$ and $\mathrm{Var}(\eta_{ps}) \to 0$, while $E(\xi)$ and $\mathrm{Var}(\xi)$ are of course invariant with respect to the value of n. So, for given value of β, the larger the value of n, the steeper the psychometric function will be, since then the variance of η_{ps} decreases.

2.3.3. The density function in case of detection by pooled activity

In this case one has directly from (11), $P(\eta_{np} \leqslant y) = P(\xi \leqslant y - g)$, so that with $g = (\sum_i g_i^p)^{1/p}$,

$$P(\eta_{np} \leqslant y) = \exp(-(\eta_0 - (y - g))^\beta), \quad y \leqslant \eta_0 + g. \tag{25}$$

and the density function follows immediately as

$$f_{\eta_{np}}(y) = \beta(\eta_0 + g - y)^{\beta-1}\exp(-(\eta_0 + g - y)^\beta), \quad y \leqslant \eta_0 + g. \tag{26}$$

For $g = 0$ this is the density of the noise. The expected value and the variance of η_{np} are given by

$$E(\eta_{np}) = g + \eta_0 - \Gamma(1 + 1/\beta), \tag{27}$$

$$\mathrm{Var}(\eta_{np}) = \Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta), \tag{28}$$

and g is given in (6). Note that the variance of η_{np} does not depend upon the value of n.

The relation between the results derived in the Sections 2.3.1–2.3.3 and those presented by Tyler and Chen is elaborated in the Appendix A.5.

2.3.4. The reduction of stimulus intensity in case of probability summation among channels

Central to Tyler and Chen's claim that the Weibull function implies multiplicative noise is their argument that if together with the Weibull function the additivity of noise is postulated, then, under the condition of detection by PS, the need for negative signal intensities follows. To illustrate, they consider the relation between the intensity of a stimulus that implies a probability of detection of $p_0 = 0.75$, if detection is by a single channel. Tyler and Chen observe correctly that if the same intensity is employed in case of detection by PS among $n = 100$ channels the probability of detection will be about equal to 1. So the intensity has to be reduced in order to get a probability of detection equal to $p_0 = 0.75$; this reduction implies that the distribution of η_{ps} will be shifted towards the left on the η-scale. The assumption that appears to be implicit in Tyler and Chen's argument is that if the density or distribution of η_{ps} is shifted by a certain amount, then the densities for the individual η_i have to be shifted by the same amount (this interpretation with respect to tacit assumptions at least complies with their Fig. 3): "... for a large enough number of channels, the mean signal needs to be set to a negative value to bring the signal + noise distribution down to threshold" (p. 3125).

This is not so. The relation between the distribution G of the maximum η_{ps} and the distribution F of the individual η_i is made explicit in extreme value statistics; let us, for simplicity's sake and like Tyler and Chen restrict

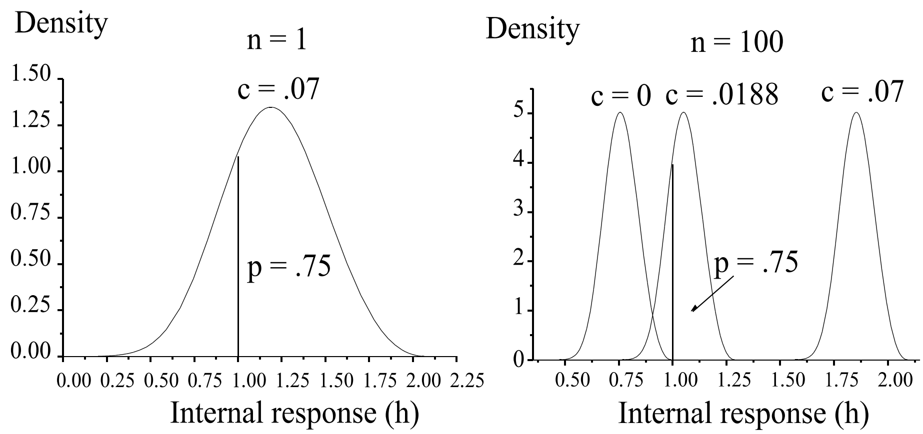

Fig. 2. Weibull densities and the adjustment of contrast (I).

ourselves to identically activated channels, which is no restriction of generality in this case. In extreme value statistics it is shown that $G(y) = P(\eta_{ps} \leqslant y) = F^n(b_n y + a_n)$, where the a_n and b_n are known as norming constants. They define, for any value of n, the shift of the distribution for each η_i. If Tyler and Chen's argument were correct it would be the value of the shift parameter a_n that implies the need for negative intensities. In this section a straightforward proof is given that the additive noise assumption does not imply, for any value of n, the need for negative intensities. A more explicit proof of this, making use of the properties of the transformation $b_n y + a_n$, is given in the Appendix A.3, providing a more explicit justification of the relation (29) below.

To begin with, recall that the Weibull function follows from the assumptions of (i) the distribution (11) and (ii) of additive noise. So properties implied by the Weibull function (like (29) below) cannot contradict the postulate of additive noise. Recall further that Tyler and Chen assume identically activated channels, and so one has $\alpha = \sum_i h_i^\beta = nh^\beta$. It follows then from (9) that $\exp(-c_1^\beta h^\beta) = \exp(-nc_n^\beta h^\beta)$, for all $n > 1$, implying $nc_n^\beta = c_1^\beta$, and one has

$$c_1 = n^{1/\beta} c_n, \quad c_n = (1/n)^{1/\beta} c_1. \tag{29}$$

So contrary to Tyler and Chen's claim one has:

1. If $c_n > 0$, then $c_1 > 0$, and $c_1 > 0$ implies $c_n > 0$. This holds for any value of n, small or large. All that can be said is that $n \to \infty$ implies $c_n \downarrow 0$.
2. The multiplicative re-scaling of the contrast c does not imply multiplicative re-scaling of the noise: the noise distribution depends on the parameters η_0 and β, which are not changed when the value of n is changed.
3. The expected values of the distributions do change with n and c. Let E_n denote an expected value if detection is by PS among $n > 1$ channels, and E_1 if detection by a single channel. For η_{ps}, one has, because of (29), the expected value

$$\begin{aligned} E_n(\eta_{ps}) &= \eta_0 + c_n h - (1/n)^{1/\beta}\Gamma(1+1/\beta) \\ &= \eta_0 + (1/n)^{1/\beta} c_1 (h - \Gamma(1+1/\beta)), \end{aligned}$$

and the expected value of an individual η_i is similarly given by

$$E_n(\eta_i) = \eta_0 + (1/n)^{1/\beta} c_1 h - \Gamma(1+1/\beta).$$

Note that here the term $\Gamma(1+1/\beta)$ is not scaled with $(1/n)^{1/\beta}$ as in the expression for $E_n(\eta_{ps})$, because apart from a shift by g, η_i is distributed like the noise ξ. So for increasing value of n, the threshold value c_n (i.e. the value of c for which the probability of the event $\{\eta > \eta_0\}$ equals p_0, say $p_0 = 0.75$, approaches 0 and $E_n(\eta_i) \to E(\xi) = \eta_0 - \Gamma(1+1/\beta)$, with $E(\xi)$ the expectation of the noise, while $E_n(\eta_{ps}) \to \eta_0$). Since $\Gamma(1+1/\beta) > 0$ for all $\beta > 0$ it follows that $E_n(\eta_{ps}) > E_n(\eta_i)$ for all $n > 0$.

Now $E_1(\eta_{ps}) = \eta_0 + c_1 h - \Gamma(1+1/\beta)$, and since $(1/n)^{1/\beta} < 1$ for $n > 1$, it follows that

$$E_1(\eta_{ps}) > E_n(\eta_{ps}) > E_n(\eta_i) \to E(\xi), \quad i = 1, 2, \ldots, n > 1, \tag{30}$$

where the contrasts $c_1 > 0$ and $c_n = (1/n)^{1/\beta} c_1 > 0$ are, for all $n > 1$, threshold intensities or contrasts.

The relation between c_1 and c_n may be illustrated graphically, [7] (see Fig. 2). Consider the case $\beta = 3.5$ and $\eta_0 = 1$; this is the threshold value chosen by Tyler and Chen, although they do not refer explicitly to R_Θ when they speak of the "threshold level equal to 1.0" (p. 3126). The value of β defines the standard deviation of the densities: for $n = 1$, $\sigma = \sigma_1 = 0.285$, and for $n = 100$, $\sigma = \sigma_{100} = 0.076$; a value of $\sigma = 0.67$ as given by Tyler and Chen is not possible, this value would require $\beta \approx 1.325$, incidentally, this value of β implies what Tyler

[7] According to Dr. Tyler (personal communication), only a graphical presentation of the argument is convincing.

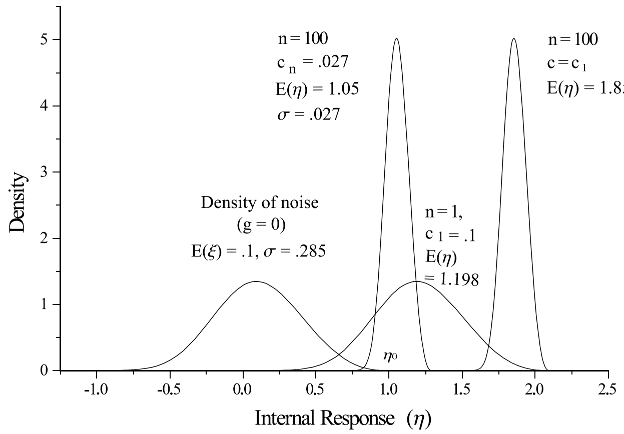

Fig. 3. Weibull densities and the adjustment of contrast (II).

and Chen call "bizarre" forms of the density. Note that the choice $\eta_0 = 1$ means that one can no longer choose the noise distribution to have an expected value equal to zero, because in that case [8] one would have $\eta_0 = \Gamma(1 + 1/\beta)$ (cf. (23)); for $\beta = 3.5$, one gets $\eta_0 = 0.8997$. The value of β implies that the density looks approximately normal, although this is totally irrelevant for the argument, which holds for any value of $\beta > 0$.

Detection occurs if $\eta > \eta_0$. Two cases will be considered: $n = 1$ and 100. For the case $n = 1$, g has to assume the value $g = 1.0976$ for $P(\eta > \eta_0) = 0.75$. Choosing (arbitrarily) a contrast $c = 0.07$ for this case, one finds $h = 15.68$; h will be assumed to be constant for different values of n, so the contrast c, characterising the stimulus, will have to be adjusted for different values of n. Fig. 2 illustrates what happens. The graph on the right ($n = 100$) shows the densities of η for different values of c_n. For $c = 0.07$ the density assumes nonzero values only for values of η larger than 1.5 or 1.6, so the probability of detection would always equal 1.

In Fig. 3 the densities for $n = 1$ and 100 for $\beta = 3.5$ and $\eta_0 = 1$ are presented again. The noise variable ξ has, correspondingly, an expected value $E(\xi) = 0.1$ and, of course, a standard deviation $\sigma = 0.285$. For a probability of detection $P(\eta > \eta_0 | n = 1) = 0.75$, a value of $g = 1.0976$ is required, as before. The detection contrast was this time set equal to $c_1 = 0.1$, implying $h = 10.978$. The expected values are $E_1(\eta_{ps}) = 1.198$, $E_n(\eta_{ps}) = 1.856$ if $c = c_1$ is considered, and $E_n(\eta_{ps}) = 1.05$ if $c_n = 0.027$ is chosen: for this value of c_n the probability of detection is again equal to 0.75. Note that for the appropriate threshold contrasts, $E_1(\eta_{ps}) > E_n(\eta_{ps})$, corresponding to (30).

2.3.5. Parallel psychometric functions on log-intensity scales

On p. 3126, Tyler and Chen argue that the Weibull function "has a constant form when plotted on log coordinates, i.e. is scaled in proportion to signal amplitude. This implies that the limiting noise is similarly scaled through the probability summation operator". From this, so the authors, it follows "that the noise is multiplicative".

The authors do not make explicit what they mean by the statement "the limiting noise is similarly scaled through the probability summation operator". [9] Let us therefore first look at the mathematical structure of the parallel shift on the log-c-scale: such a shift occurs by some manipulation either of the experimental conditions (e.g. cueing to focus on particular channels, or duration of stimulus presentation) or of the stimulus (e.g. of the spatial frequency parameter of Gabor patches). Since $\exp(-\alpha c^\beta) = \exp(-e^{\log\alpha + \beta\log c})$, formally the parallel shift on the log-c-axis is obviously due to a change of the value of α, not of β, and may occur when detection is by PS or by NP. α determines the location of the psychometric function on the scale and depends on g; the properties of the noise are defined by β and η_0 and are not changed by PS. What does change, with the number n of channels involved, is the distribution of the maximum η_{ps}. But the distribution of η_{ps} is not the distribution of the noise. Moreover, as stated before, the Weibull function (9), i.e. $\psi_w(c) = 1 - \exp(-\alpha c^\beta)$, is logically implied by the assumptions of (i) additive noise, and (ii) the Weibull distribution (11) of the noise. Consequently, the Weibull function cannot imply that the noise is multiplicative.

It may be shown that if (i) the assumption A1 (additive noise, peak detection and detection by PS) holds, and (ii) psychometric functions corresponding to different stimulus parameters are log-parallel, then the psychometric functions are implied to be Weibull functions. This property thus characterises the Weibull. Green and Luce (1975) provided a proof of this uniqueness property, assuming however, that (i) the number n of activated channels involved in the detection task varies with the experimental conditions that modify the value of α, and that (ii) the channels are identically activated. Mortensen (1988) provided a more general proof allowing for constant value of n and differently activated channels. [10] In principle, this result may be

[8] $E(\xi) = 0$ implies $\eta_0 - \Gamma(1 + 1/\beta) = 0$ and consequently $\eta_0 = \Gamma(1 + 1/\beta)$. More explicitly, let $\zeta = \eta - E(\eta_{ps})$; then

$$\begin{aligned} P(\zeta \leqslant z) = P(\eta \leqslant z + E(\eta_{ps})) &= \exp(-(\eta_0 - z - \eta_0 + \Gamma)^\beta) \\ &= \exp(-(\Gamma - z)^\beta), \end{aligned}$$

which means that the free parameter η_0 is replaced by the special choice $\eta_0 = \Gamma(1 + 1/\beta)$.

[9] After all, PS is not a process or "operator", acting on the noise.

[10] The proof in Mortensen (1988) assumes high-threshold detection. An improved version of the proof, showing that under the condition of A1 log-parallelism also implies high-threshold detection, is available from the author on demand.

utilised to provide a test of the hypothesis that empirically determined psychometric functions are compatible with the assumption A1, i.e. with peak detection, PS among channels and the noise variable being independent of the mean, i.e. the deterministic part of the activity: the lack of log-parallelism signals that at least one of these assumptions does not represent a meaningful approximation. On the other hand, log-parallel psychometric functions like those presented by Nachmias (1967) uniquely point to the Weibull only if one can make sure that A1 holds. If, for instance, A2 holds instead, i.e. if detection is by NP and not by PS, then the psychometric functions are not necessarily defined as Weibulls; in fact, any distribution of the noise will yield log-parallel psychometric functions as long as the noise is independent of the deterministic part of the activity.

3. The Weibull function and 2AFC-experiments

3.1. Relaxing the high-threshold-assumption

According to assumption A1 detection occurs if η_{ps} assumes a value larger than η_0, and if η_0 equals the maximal value the noise variables ξ_i may assume, as in (11), then detection is high threshold. Let us now suppose that, as before, the distribution function of the noise is given by (11), but that the subject may choose the threshold value to be equal to some $\eta_s < \eta_0$; then detection is no longer high threshold. Consider now the event $\{\text{detection by the } i\text{th channel}\} = \{g_i + \xi_i > \eta_s\}$, with $\eta_s < \eta_0$. The probability of no detection is then given by (11) for $y = \eta_s$ and with $x = \eta_s - g$ one has

$$P(\eta_{\mathrm{ps}} \leqslant \eta_s) = \exp\left(-\sum_{i=1}^{n}(\eta_0 - \eta_s + g_i)^\beta\right) = \exp\left(-\sum_{i=1}^{n}(\delta + g_i)^\beta\right), \quad \delta = \eta_0 - \eta_s, \tag{31}$$

and the psychometric function is given by

$$\psi_{\mathrm{w}}(c) = 1 - \exp\left(-\sum_{i=1}^{n}(\delta + ch_i)^\beta\right). \tag{32}$$

The treatment of the case of detection by NP is analogous. Consider the condition $\{g + \xi > \eta_s < \eta_0\}$. With $x = \eta_s - g$ one has $P(\eta_{\mathrm{np}} \leqslant \eta_s) = \exp(-(\eta_0 - \eta_s + g)^\beta)$, and with $\delta = \eta_0 - \eta_s$ it follows that

$$\psi_{\mathrm{w}}(c) = 1 - \exp(-(\delta + g)^\beta), \quad \delta > 0. \tag{33}$$

To summarise, the Weibull distribution may be employed to define the psychometric function without necessarily implying a high-threshold model, although the form (32) does not seem to have been employed so far.

3.2. The Weibull function and the 2AFC-task

The experimental trial consists of two consecutive intervals I_1 and I_2, and with probability p_k the stimulus is presented within the interval I_k, $k = 1, 2$. Let $\eta^{(k)}$ represent the activity in I_k. The subject gives a correct response (CR) with probability $P(\mathrm{CR}) = P(\eta^{(2)} < \eta^{(1)}|I_1)p_1 + P(\eta^{(1)} < \eta^{(2)}|I_2)p_2$. Let $\eta^{\mathbf{n}}$ be the activity if, in an interval I_k, no stimulus was presented (so $\eta^{\mathbf{n}} = \xi$), and $\eta^{\mathbf{sn}}$ be the activity when a stimulus was presented, so $\eta^{\mathbf{n}} = \eta^{(2)}$ and $\eta^{\mathbf{sn}} = \eta^{(1)}$ if the stimulus was presented in I_1. Assuming $P(\eta^{(2)} < \eta^{(1)}|I_1) = P(\eta^{(1)} < \eta^{(2)}|I_2) = P(\eta^{\mathbf{n}} < \eta^{\mathbf{sn}})$, one has, with $p_1 = p_2$,

$$P(\mathrm{CR}|c) = \int_{-\infty}^{\eta_0} f_{\mathbf{n}}(\xi) \int_{\eta^{\mathbf{n}}}^{\eta_0 + ch} f_{\mathbf{sn}}(\eta^{\mathbf{sn}})\,\mathrm{d}\xi\,\mathrm{d}\eta^{\mathbf{sn}}, \tag{34}$$

with $f_{\mathbf{n}}$ the density of the noise, and $f_{\mathbf{sn}}$ the density of the activity when a stimulus was presented. (One may, equivalently, consider the difference $\eta_{\mathbf{sn}} - \xi$, but that would require the derivation of the convolution of $\eta_{\mathbf{sn}}$ and $-\xi$ etc., without changing the result.)

Theorem 2. *Suppose the noise is defined by the Weibull density* (17) *with identically activated channels. The psychometric functions for the 2AFC-task are given by*

$$P(D|c) = 1 - \int_0^\infty \mathrm{e}^{-y-(y^{1/\beta}+n^{1/\beta}g)^\beta}\,\mathrm{d}y, \quad g = ch. \tag{35}$$

Proof. See Appendix A.4.

The assumption of identically activated channels was also made by Tyler and Chen and by Pelli (1985); with respect to actual stimuli, the case of not identically activated channels will have to be mastered. Still, considering the case of identically activated channels one may get a certain intuition concerning the effect of different values of β. Fig. 4 illustrates this result (same parameters as in the numerical–graphical example). Generally, the curves become steeper, i.e. shift to the left, with increasing value of n; however, the rate at which they shift, differ for different β-values. The case of detection by pooling is formally covered by the case $n = 1$, and $g = c(\sum_i h_i^p)^{1/p}$.

The argument, that the number of (identically activated) channels involved in a detection task may be deduced from graphs of 2AFC-psychometric functions, is, however, not very convincing. Even if the psychometric functions could be determined free of experimental error their similarity is far too great to allow for a reasonable estimate of n, and the postulate of identically activated channels appears to be utterly unrealistic.

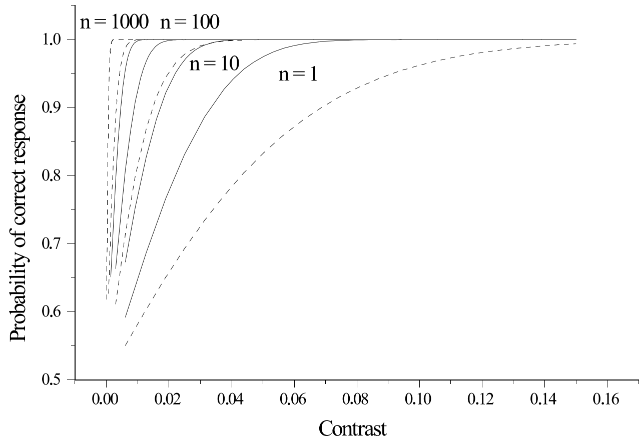

Fig. 4. 2AFC-Psychometric functions for different values of n.

Wilson (1980) presented a model of what he called the contrast transduction process, adopting the function originally introduced by Quick (1974); he proposed

$$\psi_Q(c) = 1 - 2^{-(1+(hc))^Q}, \tag{36}$$

with h denoting the sensitivity of the visual system to the stimulus, and c contrast. This may be re-written in the form

$$\psi_Q(c) = 1 - \exp(-\log 2(1 + (ch)^Q)). \tag{37}$$

ψ_Q is, as a function of c, S-shaped, with $\psi_Q(0) = 1/2$; in fact, ψ_q is meant to cater for 2AFC-experiments. At first glance, Wilson's adaptation of Quick's function looks similar to (33) with $g = ch$. However, in (33) the exponent in $\exp(\cdot)$ is $(\delta + g)^\beta$, whereas in (37) the exponent is $\delta + (ch\delta^{1/\beta})^\beta$, if one puts $Q = \beta$; moreover, δ is not a free parameter as in (33), but has the fixed value log(2). Wilson's formula was not derived assuming a particular density for the noise, but in an ad-hoc way to approximate a psychometric function, defined in terms of a Gaussian, to cater for 2AFC-experiments.

4. Random variables and neural activity

4.1. Neural activity as a stochastic process

The activity of a neuron is a stochastic process. This means that the time course of the activity during a trial can be represented by some function of time reflecting systematic aspects of the activity as well as random fluctuations. Such functions are known as random functions, sample paths or trajectories of the stochastic process. Generally, a stochastic process is defined as a family of random functions, defined as $X_t = \{x(t,\omega)|\omega \in \Omega, t \in (0,T]\}$, where $(0,T]$ denotes a trial, and Ω is a set of functions. In a given trial, a certain function $\omega \in \Omega$ is sampled. $x(t,\omega)$ is a numerical representation of ω at time t. For a more complete characterisation of stochastic processes, a probability measure P has to be introduced (see, for instance, Wong, 1971); however, there is no need to go into such details here.

There are two major classes of stochastic processes: those with everywhere continuous sample paths and those with sample paths that do not have this property. Counting processes belong to the latter class: the sample paths or trajectories of such a process are staircases. At each time t_i at which an event occurs the trajectory increases one unit and then remains constant until, at time t_{i+1}, the next event occurs.

Detection will be conceived as a level-crossing process, i.e. it will be assumed that a stimulus is detected if in at least one channel the activity reaches a certain level within the trial. Both types of processes will be considered.

4.2. Stochastic processes with continuous sample paths

Heller, Hertz, Kjaer and Richmond (1995) argued that all the relevant information about a presented stimulus is carried by what they call the effective time-varying firing rate, resulting from averaging the generation of spikes during a time window not smaller than 25 ms in primary cortex and not smaller than 50 ms in the inferior temporal cortex (IT). Gershon, Wiener, Latham and Richmond (1998) further investigated the information transmitted by visual neurons in V1 and IT and found that the distribution of the spike count is best represented by a Gaussian distribution truncated at 0, with the logarithms of the mean and the variance being linearly related, a finding also reported by Dean (1981), Bradley, Skottun, Ohzawa, Sclar and Freeman (1987), Snowden, Treue and Andersen (1992) and Britten, Shadlen, Newsome and Movshon (1993).

Let $\rho(t)$ be the spike rate at time t. According to Gershon et al.'s finding $\rho(t)$ can be approximated by a Gaussian variable, truncated at 0. The activation process is then defined by the process $\rho_t = \{\rho(t,\omega), t \in (0,T]\}$; one may symbolically write $\rho_t = \xi_t + g$, where g is the function (not just a single value) representing the noise free activation of the neural channel by the stimulus. Alternatively, one may write $E(\xi(t)) = 0$ for all t; then $E(\rho(t) = g(t))$, i.e. g is the mean value function of the activation process. [11]

In order to relate the representation of activity by stochastic processes to psychometric functions defined according to (3), i.e. as $\psi(c) = 1 - P(\eta \leqslant \eta_s|c)$, one has to define a random variable η that reflects the detection process. This means that one has to find some function of such a random function, i.e. a "functional". One possible functional is the likelihood ratio $\Lambda(\omega)$, relating the probability of an observed trajectory ω under the

[11] Note that the sets Ω_ρ and Ω_ξ of sample functions for the processes ρ_t and ξ_t are different: each $\omega \in \Omega_\rho$ is the sum of some $\omega \in \Omega_\xi$ plus g.

condition that a stimulus was shown to the probability of the same ω under the condition that no stimulus was shown. However, to assume that $\eta = \eta(\omega) = \Lambda(\omega)$ and to postulate that a detection response is given whenever $\eta > \eta_s$ for some suitably chosen η_s implies that the observer is ideal, which is not necessarily the case. Consequently, the likelihood ratio Λ may not be a good choice for the definition of η.

An alternative to the likelihood-ratio interpretation of η is to postulate

$$\eta = \max_{t\in(0,T]} \rho(t) = \max_{t\in(0,T]}[g(t) + \xi(t)] > \eta_s. \tag{38}$$

(38) is equivalent to saying that the stimulus is detected if the activity crosses the level η_s at some time within the interval $(0, T]$. Recall that in general

$$\max_{t\in(0,T]} \rho(t) \neq \max_{t\in(0,T]} g_{\max} + \xi(t_0), \quad g_{\max} = g(t_0),$$

where t_0 is the time at which g assumes its maximal value. The right hand side represents peak detection, which may serve as an approximation, though.

The random variable η is related to the waiting time τ required for a trajectory ρ to reach the threshold level η_s. If $\tau > T$, the trajectory has not reached η_s and the subject has not detected the stimulus. Therefore,

$$P(\eta > \eta_s) = P(\tau \leqslant T). \tag{39}$$

The distribution of waiting times can be defined in terms of a hazard function ϕ. One has

$$P(\tau \leqslant T) = 1 - \exp\left(-\int_0^T \phi(t)\,\mathrm{d}t\right) \tag{40}$$

and $\phi(t) = h(t)/(1 - H(t))$, h the density function of τ and $H(t) = \int_0^t h(\tau)\,\mathrm{d}\tau$ (cf. Papoulis, 1965). (39) allows to establish a relationship between detection probabilities and reaction times, and relates the detection process to temporal PS (see below).

One approximation is based on extreme value statistics and the assumptions of a "large" value of η_s and of a wide-sense stationary Gauss process ξ_t (implying the variance of $\xi(t)$ to be constant, i.e. independent of t and thus idealising Gershon et al.'s finding) and was discussed in Mortensen et al. (1991). Here, just the result is presented, namely

$$P(\eta \leqslant \eta_s) = \exp\left(-\frac{\sqrt{\lambda_2}}{2\pi}\int_0^T \exp\left(-\frac{(\eta_s - g(t))^2}{2}\right)\mathrm{d}t\right), \tag{41}$$

with λ_2 as an additional free parameter known as second spectral moment. λ_2 is a measure of the speed with which the noise process ξ_t fluctuates (Papoulis, 1965): for $\lambda_2 \to 0$ the trajectories $\xi(t,\omega)$ of ξ_t will become constants, and for $\lambda_2 \to \infty$ the $\xi(t,\omega)$ will fluctuate fiercely, i.e. ξ_t will resemble white noise. So λ_2 reflects a relevant aspect of the autocorrelation function of the noise without assuming a particular autocorrelation function.

From (39)–(41) one derives that

$$\phi(t) = \frac{\sqrt{\lambda_2}}{2\pi}\exp\left(-\frac{(\eta_s - g(t))^2}{2}\right), \tag{42}$$

i.e. (41) provides an approximation for the hazard function characterising the waiting time distribution. Watson (1979) and Blommaert and Roufs (1987) discuss the role of temporal PS and start from the assumption that in a sufficiently small sub-interval $\Delta t \in (0, T)$, the probability of not detecting the stimulus is given by $\exp(-|g(t)|^\beta)$, $t \in \Delta t$, i.e. they define the hazard function in terms of the Weibull function. This assumption comes ex vacuo, i.e. no reason is given as to why this exponential should be the probability of detection in Δt. In particular, it is unclear why the probability depends upon $|g(t)|$ instead of $g(t)$. The immediate reason is, of course, that the power can only be taken for a positive number, and since g may assume negative values one has to turn them into positive ones. However, writing $|g(t)|$ is just an ad-hoc solution: formally, this notation signals the assumption of a rectifying process, but there is no discussion of such a process. Alternatively one may argue that this notation signals that detection means the crossing of an upper or lower bound of activity, but this assumption is not covered by writing $|g(t)|$. A clean solution for the problem of a crossing of an upper or lower boundary may be found in Buonocore, Nobile and Ricciardi (1987) for the special case that the noise can be defined by a stationary Gaussian (Ornstein–Uhlenbeck) process and turns out to be much more complicated. The approximation (41), on the other hand, refers only to the maximum of a sample path and is derived from rigorous results from the theory of extremes.

(41) defines a distribution function that is, mathematically, neither Weibull nor Gaussian. η is not some unspecified "instantaneous" activity but reflects a property of neural activity extended in time and can be related either to spike-rates or, because of (39), to waiting time distributions, allowing, in principle, to relate reaction times and detection processes. A discussion of this relation is beyond the scope of this paper. However, it may be pointed out that the Weibull function provides an excellent approximation also for psychometric functions defined in terms of (41). The approximation is illustrated in Fig. 5, where the impulse h has been taken from Blommaert and Roufs (1987) and is defined as

$$h(t) = \mu t^p \exp(-at), \quad t \geqslant 0, \tag{43}$$

with $\mu > 0$, $\alpha > 0$, $p > 0$. In particular $p = 7.5$; $a = 0.51$ was chosen and μ was determined such that $\max_t h(t) = 1$. The threshold was set equal to $S = 6.5$. The unit response is defined by the convolution $g_0 = h * s$, s a

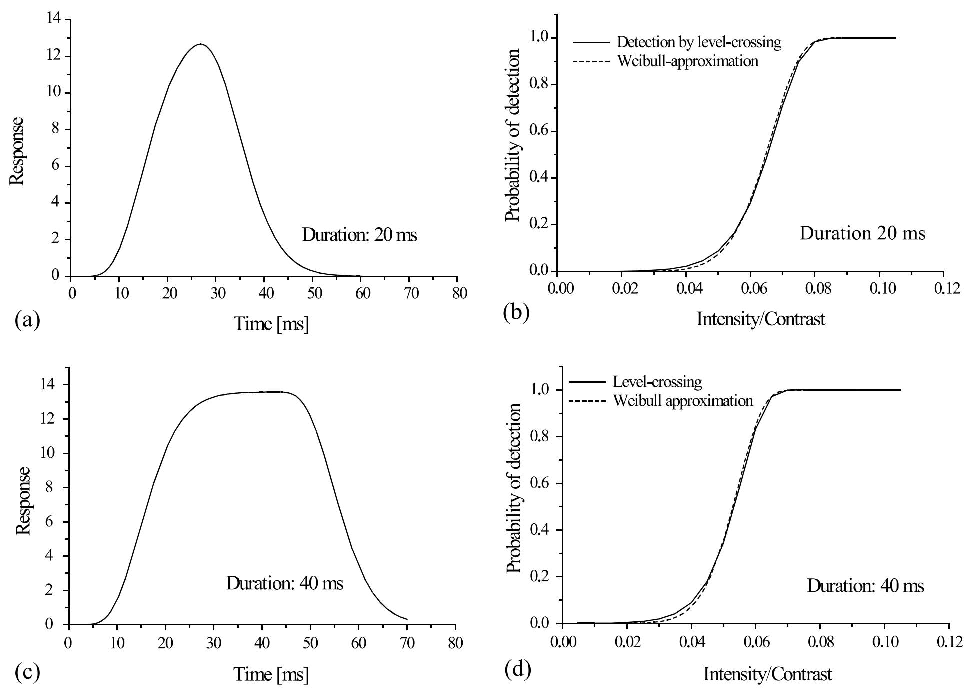

Fig. 5. Psychometric functions from level-crossings. (a) mean response for 20 ms stimulus duration, (b) corresponding psychometric function plus approximating Weibull function; (c) mean response for 40 ms stimulus duration, (d) the corresponding psychometric function.

rectangular pulse of duration either equal to 20 or 40 ms. The mean value function g is defined by $g = cg_0$.

The psychometric function was determined according to $\psi(c) = 1 - P(\eta \leqslant \eta_s|c)$, where $P(\eta \leqslant \eta_s)$ is defined in (41), assuming $\text{Var}(\xi(t)) = 1$ for all t, $\eta_s = 6.5$ and $\lambda_2 = 4000$; these values do not imply high-threshold detection in the strict, but in a practical sense ($p_0 = 0.00005$ for $T = 20$ ms, and $p_0 = 0.0002$ for $T = 40$ ms). Although these values do not imply peak detection in a strict sense, they allow an approximation by peak detection. In (Mortensen et al. (1991)), for certain data sets η_s- and λ_2-values were estimated that implied false alarm rates of 0.01 and also allowed for the peak-detection approximation.

For each psychometric function the corresponding Weibull function was determined, employing rough-and-ready estimates of the parameters of the Weibull function; for the stimulus duration equal to 20 ms, $\alpha = 8.76885 \times 10^9$ had to be chosen, and for the 40 ms duration one finds $\alpha = 4.63897 \times 10^{10}$; for both durations, the estimate $\beta = 8.5$ appeared to be optimal. It should be noted that the parameters α and β are not quite unique: values of β deviating slightly from the value 8.5 lead to an equally good fit, provided the value of α is correspondingly adjusted. In any case, the identical value of β for both Weibull approximations reflects the fact that the value of the second spectral moment λ_2 was the same for both durations.

The level-crossing probability (41) is also an approximation, and in the Discussion some alternatives will be indicated that let (41) appear to be rather simple. Still, (41) is in many respects a more plausible approximation than the one based on the unspecified notion of an instantaneous internal response. The density function corresponding to $F_\eta(y) = P(\eta \leqslant y)$, where η_s has been replaced by y to indicate that one now considers F_η as a function of y, is difficult to compute as long as g is defined as a convolution, not as a constant. It is clear, though, that there is no finite lower and upper limit for y; a good fit of the Weibull does, consequently, not mean that detection is indeed high threshold.

Fig. 5 demonstrates an excellent fit of the Weibull to a function that is mathematically defined quite differently. The fit implies that the parameters α and β of the Weibull can be determined such that

$$\frac{\sqrt{\lambda_2}}{2\pi} \int_0^T \exp(-\tfrac{1}{2}(\eta_s - cg_0(t))^2\,dt) \approx \alpha c^\beta. \quad (44)$$

Suppose the level-crossing interpretation of detection is correctly described by (41). The stochastic effects are then reflected by the second spectral moment λ_2, the decision criterion is given by η_s, and the "deterministic" effects of the stimulus are defined by $cg_0(t)$, where g_0 is defined by a convolution of the stimulus with the impulse response h. The parameters λ_2, η_s and those of the impulse response are all summarised by the Weibull parameters α and β. Thus the straightforward interpretation of β being the parameter that reflects "the noise". On the other hand, the psychometric functions in Fig. 5 are compatible with identical β-values for

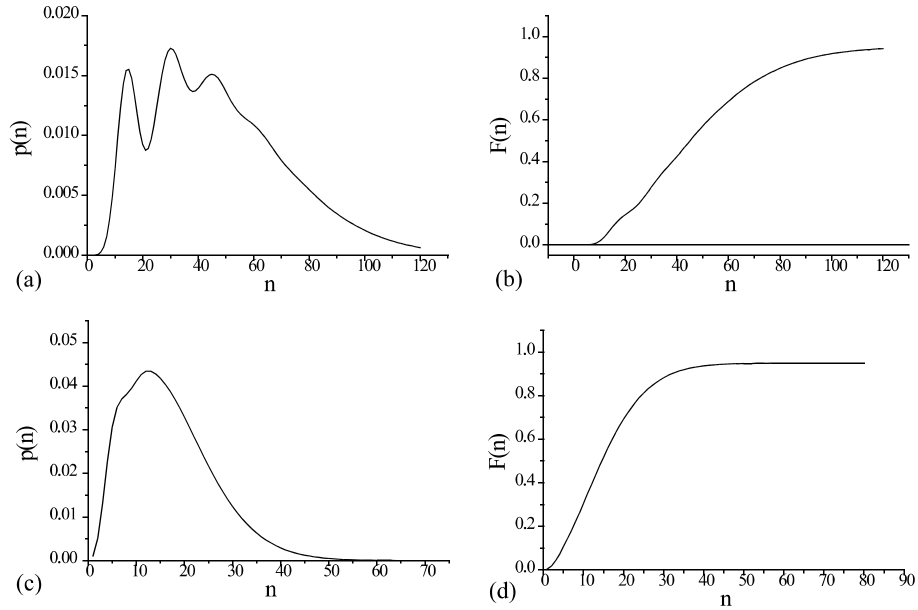

Fig. 6. NTA distributions, (a) $\mu = 3$, $\lambda_2 = 15$, (c) $\mu = 3$, $\lambda_2 = 5$, (b) and (d) the corresponding cumulative distribution functions.

both durations, implying that the corresponding psychometric functions can be considered as being log-parallel. One may argue that this finding results from the particular choice of the values of λ_2 and η_s: the value of η_s is "large", suggesting that the peak-detection approximation holds. Still, β reflects the particular combination of values of λ_2 and η_s; it is this combination of parameter values which determines what is meant when one speaks of "the effect of the noise", provided (42) is the correct hazard function.

4.3. Stochastic processes with stair-case sample paths

Finally, the approximation of a counting distribution, namely the Neyman-Tape-A (NTA) distribution, by the Weibull will be presented. For an application of this model to the discussion of thresholds for brief light flashes see e.g. Reeves, Wu and Schirillo (1998), who did not, however, consider psychometric functions but threshold versus intensity curves. The model assumes a form of multiplicative noise; further coments on multiplicative noise will be given in Section 5.

It may be recalled that already Brindley (1963) derived the Weibull from the assumption that detection is by PS among many independent detectors, where each of the detectors is sensitive to an m-quantum coincidence. Brindley made use of the fact that the number of photons emitted by a light source is Poisson distributed. However, Green and Luce (1975) pointed out that Brindley's derivations contain some inconsistencies. In the light of the work of of a model proposed by Teich et al. (1982) it may be interesting to see whether the NTA distribution can also be approximated by the Weibull, here without PS among neurons, though. The stimulus is assumed to be represented by a Poisson distribution $p(k|\mu)$, where μ defines the intensity of the signal, and

$$p(k|\mu) = \mathrm{e}^{-\mu}\frac{\mu^k}{k!}, \quad k = 0, 1, 2, \ldots \tag{45}$$

The response is defined by another Poisson process with intensity ak, with

$$p(n|ak) = \mathrm{e}^{-ak}\frac{(ak)^n}{n!}, \quad n = 0, 1, 2, \ldots, \ a > 0, \tag{46}$$

where a is known as multiplication parameter (Teich, 1981), representing the average number of events generated by an input event. The probability of observing n events, given the intensity μ, is then

$$\begin{aligned} p(n|\mu) &= \sum_{k=0}^{\infty} p(n|ak)p(k|\mu) \\ &= \mathrm{e}^{-\mu}\sum_{k=0}^{\infty}\mathrm{e}^{-ak}\frac{(ak)^n}{n!}\frac{\mu^k}{k!}. \end{aligned} \tag{47}$$

The expected value and the variance of n are given by

$$E(n) = a\mu, \quad \mathrm{Var}(n) = (1+a)a\mu. \tag{48}$$

In a Poisson process, expected value and variance of the number of events generated per unit of time are identical; according to (48), $\mathrm{Var}(n) > E(n)$ for the NTA distribution. One may construct a very simple model of detection postulating that the stimulus is detected whenever $n > k_0$, k_0 some "critical" number, within an interval of length $\Delta t \subseteq (0, T]$, where $(0, T]$ denotes a trial of duration T. For sufficiently small values of T one may put $\Delta t = (0, T]$ and calculate the probability of the event

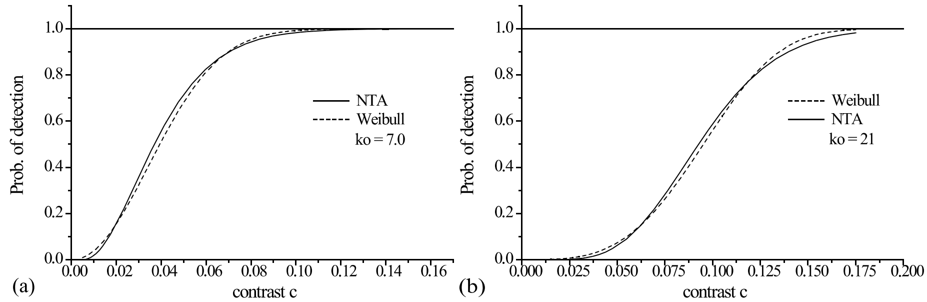

Fig. 7. Psychometric functions, defined in terms of the NTA distribution ($a = 2.5$), and approximating Weibull functions.

$\{n > k_0\}$, employing the NTA distribution. Fig. 6 shows an NTA-type probability distributions and the corresponding cumulative distributions.

Note that the NTA distribution is not continuous; it only looks like a continuous distribution because of the relatively large number of events that may occur within the unit time interval. The parameters of the distribution have been chosen such as to demonstrate the possibility of the probability distribution being multimodal. The multimodality can hardly be recognised in the cumulative distributions, which illustrates the smoothing effects of cumulation. Fig. 7 shows psychometric functions, defined in terms of the NTA distribution, with $\mu = 3$ in each case. In (a), the critical number of events for detection has been set equal to $k_0 = 7$, whereas in (b) k_0 was set equal to 22.

Psychometric functions defined in terms of the NTA distribution are shown in Fig. 7 together with approximating Weibull functions. Here, the intensity or contrast c has to be related to the parameter μ. In Fig. 5, values of c between $c = 0.001$ and 0.15 where considered. The psychometric functions defined in terms of the NTA distribution should be related to a similar range of c-values. To this end, μ was defined as a linear transformation of c, i.e. $\mu = u * \ c + v$, with u and v depending on the value of a. The interpretation of u and v depends on the specifics of the definition of the stimulus and will not be further discussed here. One may argue that v should be equal to 0, since the Poisson input and c should be related only by a change of unit; the case $v \neq 0$ may relate to the generalised form (33) of the Weibull. However, this detail does not seem to be of much inportance, see below.

The psychometric functions differ with respect to the value of k_0; in (a), $k_0 = 7$, and in (b), $k_0 = 21$; the NTA parameter a was set equal to 2.5 for both functions. For the approximating Weibull, $\alpha = 600$, $\beta = 2.1$ was found, while for (b), $\alpha = 2313.3$ and $\beta = 3.45$ had to be chosen. If actual data had been generated by the NTA distribution, one would say that the fit of the Weibulls is satisfactory, i.e. the Weibull can be fitted to psychometric functions defined in terms of the NTA distribution. Note that the psychometric function for $k_0 = 21$ increases slower than the psychometric function for $k_0 = 7$, as it should. Still, a larger value of β had to be chosen to fit the Weibull. The point is that both parameters, α and β, have to be found, to fit the Weibull, and these parameters cannot be determined independent of each other.

It may be interesting to fit the generalised Weibull function to allow for genuine false alarms, in particular the version (33), since PS among any channels is not considered here. This requires the estimation of the parameters δ, β and h such that $g = ch$; for the $k_0 = 21$-function one finds $\beta = 3.15$, $\delta = -.045$, $h = 2313.28$. The fits for functions with parameters deviating slightly from these values are equally good, in particular the fit for the case $\delta = 0$, $\beta = 3.45$. In other words, not the generalised Weibull, but the Weibull itself appears to be the best function to fit to the NTA data (the extra parameter δ does not imply a really better fit). This demonstrates that the Weibull cannot only be fitted to data generated by additive noise, but also to data generated by a form of multiplicative noise.

The NTA model considered here is too simple to allow for a discussion of the effect of the variation of stimulus parameters, like the spatial frequency parameter of a Gabor patch. Depending on the form of a generalisation of the model one may predict different β-values for the approximating Weibull when such a parameter is changed, even if the "critical" number k_0 remains constant. Then psychometric functions would be predicted that are not log-parallel, and this would indicate that the assumptions of peak detection with invariant variance of the noise were inadequate.

5. Summary and discussion

5.1. Summary: the Weibull function and additive noise

The Weibull can be fitted to psychometric functions generated by mechanisms with additive or multiplicative noise, with either peak detection or level-crossing. The reason for this seems to be the fact that for freely chosen α and β the function $1 - \exp(-\alpha c^{\beta})$ can assume shapes that correspond to those generated by a multitude of functions. The may also be formulated the other way round: a multitude of mathematically differently defined functions assume very similar shapes.

Focusing more closely on a characterisation of the Weibull itself, it was shown that, provided the activity in a neural channel can be represented in the form $g + \xi$, a psychometric function is a Weibull function, if and only if (i) the noise ξ is Weibull distributed (i.e. has distribution (11)) with upper limit η_0 and (ii) the subject sets the internal threshold η_s equal to η_0. The representation $g + \xi$ corresponds to what may be called peak detection, which is a special case of detection by level-crossing. It is possible that the Weibull can also be derived from a level-crossing model analogous to that defined by (41). The noise process cannot be Gaussian then; this possibility has not been explored so far.

In any case, it follows that the Weibull function neither implies the assumption that noise is multiplicative with signal strength, nor the need for negative signal intensities. This holds for the case of detection by PS or by NP. When Tyler and Chen (p. 3126) speak of the "scaling of the limiting noise through the probability summation operator" without giving a formal definition of what they mean by "limiting noise", they presumably refer to the scaling of the expected value and variance of the maximum η_{ps} of the $\eta_1, \ldots, \eta_n$ with n, which may be seen in the expression for the expected value of the maximum and the corresponding standard deviation of η_{ps}. In case of identically activated channels this is the term $\Gamma(1 + 1/\beta)$, which is scaled with the factor $(1/n)^{1/\beta}$ (see (21) and (22)). However, the parameters η_0 and β of the individual noise variables are not scaled, and therefore there is no scaling of the noise that could be interpreted as pointing to the noise being multiplicative. In case of detection by NP there is certainly no rescaling of the noise if $\alpha = (\sum_i g_i^p)^{1/p}$ is changed, e.g. as a result of focusing attention on particular channels. To summarise, different Weibull functions are parallel on $\log c$-scales only when they do not differ with respect to the noise parameter β.

5.2. The Weibull function and high-threshold detection

A characteristic property of the Weibull distribution is that it implies the existence of a finite upper limit η_0 for the noise, i.e.

$$P(\xi \leqslant \eta_0) = 1, \quad \eta_0 < \infty. \tag{49}$$

This condition implies the high-threshold model if $\eta_s = \eta_0$; a possible, though not very convincing interpretation of the model would be that a subject knows (i.e. has learned) the maximal level of activity that may occur when no stimulus was presented and considers the possibility of a stimulus having been presented only if the activity is larger than this level.

A serious question associated with the high-threshold model refers, however, to the implications of (49). When a stimulus was shown, the activity is given by $\eta = g + \xi$, and η may assume values larger than η_0. Adopting the Weibull distribution for the noise therefore implies the postulate of the existence of a mechanism that stops the activity of becoming larger than a fixed value η_0, if no stimulus was presented, and allows for an activity larger than η_0, when a stimulus was shown. The claim that the Weibull distribution adequately represents the noise would then call for a characterisation of such a mechanism. One could think of some neural device limiting the noise; however, the fact that at the same time no definite upper limit for g is demanded is quite disturbing. Recall that a stimulus is detected if $P(\eta > \eta_s)$, with $\eta_s = \eta_0$, and $P(\eta > \eta_S) = 1 - P(\xi \leqslant \eta_s - g)$. Since ξ is defined on $(-\infty, \eta_0]$, there is no definite lower limit for ξ, implying that there is no definite upper limit for g. The combination of a demand for a definite upper limit for the noise and no definite upper limit for the stimulus generated activity is difficult to console with the neurophysiological truism that sensory activity can vary only on some finite interval. It follows that the Weibull can meaningfully be employed only as an approximation. This leads to the question how to interpret the estimated parameters.

5.3. Activation processes and the meaning of the Weibull parameters

The Weibull can be fitted to most empirically found psychometric functions. It does not seem to be necessary that the noise is Weibull distributed; this was pointed out in the last section, where it was demonstrated that the Weibull can be fitted to data generated by level-crossing processes in Gaussian noise. This may have to do with the fact that the shape of the Weibull depends upon its parameters in a more flexible way than for instance the Gauss distribution: for smaller values of c the Weibull may increase more slowly than for larger values of c. Note that when the Weibull is fitted to data generated by level-crossing processes the parameters α and β reflect the effect of the parameters λ_2 and η_s (see (41)) and of those parameters defining the temporal course of $g(\cdot)$, which may also influence the form of the psychometric function. The parameters of the Weibull should therefore be interpreted with great caution, even if the approximation

$$\eta = \max_{t \in J}[g(t) + \xi(t)] \approx g_{\max} + \xi(t_0), \quad g_{\max} \approx g(t_0)$$

holds (peak detection as defined in (2) in Section 1, t_0 the time at which $g(\cdot)$ assumes its maximal value), which may be the case for large values of λ_2 and large value of η_s: the relation between the parameters of the Weibull on the one hand and λ_2 and η_s on the other is not very clear from an analytical point of view and may require extensive numerical studies.

There are some more open questions. One of them refers to the nature of the postulated channels, another to the spike rate as representing the relevant channel activity. The activation of neurons even in V1 is known to depend on context, attention and learning (e.g. Gilbert, Ito, Kapadia & Westheimer, 2000). The activity of different neural units may vary considerably for different presentations of the same stimulus, due to the ongoing activity at the time of stimulus presentation (Arieli, Sterkin, Grinvald & Aertsen, 1996) and is known to be correlated (Aertsen, Erb & Palm, 1994). The correlated activity of different neurons even raises doubts concerning the hypothesis that the relevant variable carrying the information contained in neuronal activity is the spike rate. Gerstner, Ritz and van Hemmen (1993) argued that it is spike patterns of sets or assemblies of neurons, not spike rates of individual neurons that represent the activity generated by stimuli. Indeed, Vaadia et al. (1995) found that the activity of different neurons may become correlated without modification of the spike rate; upon stimulation different neurons may associate into functional groups and at the same time become dissociated from concurrently activated, competing groups. So it may not make sense to postulate fixed channels $C_1, \ldots, C_n$, and we may have to face up to the possibility that it is not an increased spike rate, but correlated activity in a functional group or cell assembly that characterises the detection process.

With regard to these possibilities, it could be interesting to find out in which way neuronal activity may be related to the Weibull distribution, possibly with respect to processes of NP of different neurons. du Buf (1992, 1994) e.g. proposed such a model for the interaction of simple cells to explain, for instance, Mach bands. Meinhardt (2000) reports superposition experiments, i.e. experiments where the stimulus were of the form $s = c_1s_1 + c_2s_2$ with c_1, c_2 contrasts and s_1, s_2 sinusoidal grating patches in one experiment and edge-type patterns in the other. [12] For each combination (s_1, s_2) a complete contrast interrelationship-function was determined, i.e. the relation between c_1 and c_2 for constant probability of detection. [13] Meinhardt assumed that for each component pattern s_i a filter with unit response h_i exists and that detection depends on the pooled response $(\sum_i h_i^p)^{1/p}$. The fits of the contrast-interrelationship functions on the basis of the NP model (6) were excellent, with $p \approx 3$ for the grating patches and $p \approx 2$ for the edge patterns. The value of p for the edge patterns may be taken as support for the hypothesis of detection by matched filters for each stimulus component (Logvinenko, 1995). However, this interpretation does not hold for the grating patterns. Most important is the finding that the β-values estimated for the various psychometric functions, defined as Weibull functions, assumed values in the neighbourhood of $\beta = 6$. The different estimates for p and β supports the pooling model, and certainly not a model of PS, defined with respect to the Weibull distribution with equal value of β for all possible channels. Any attempt to establish a relationship between Meinhardt's results and the type of results reported by Vaadia et al. will be a very difficult (recall that the temporal aspects of the activation process have to be taken into account as well, which Meinhardt did not do) and are far beyond the scope of this paper. To summarise, all these findings together with the complexities of the effects of attentional focusing (see Braun, Koch & Davis (2001) for a review of recent findings) challenge Tyler and Chen's and Pelli's (1985) idea that one can infer from psychometric functions the number of channels involved, whatever they are, and additionally assuming that they are identically activated. Tyler and Chen's claim to have provided a "rigorous and neurophysiologically plausible approach to the universe of integrative mechanisms underlying psychophysical measures" (p. 3141) appears to be somewhat far fetched.

5.4. The role of multiplicative noise

The general discussion of activation processes leads to that of the role of mutiplicative noise. Tyler and Chen describe the effect of such noise as "dramatic", although they do not deduce these effects from data. The role of multiplicative noise in visual detection was pointed out before by e.g. McGill (1967), Lillywhite (1981), Tolhurst, Movshon and Thompson (1981), and Teich et al. (1982); however, at least in psychophysical work concerned with the coding of patterns the standard assumption appears to be that of additive noise. Since the characterisation of multiplicative noise by Tyler and Chen is somewhat unclear [14] a brief reflection on what

[12] For the grating stimuli, s_1 was defined by $f_1 = 5c/\deg$, while the spatial frequency f_2 of s_2 was one of the set $\{2, 2.5, 3, 3.5, 4\}$. The width of the grating patches was 0.4^0, and the width of the edge patterns was either 0.2^0 or 0.4^0.

[13] Details concerning the normalisation of contrasts are omitted.

[14] Tyler and Chen illustrate the possible effect of multiplicative noise by stating that the noise is represented by $\sigma_R \propto kR^q + \sigma_N$, where σ_R is the noise of the response (called "root-multiplicative" by Tyler and Chen, whatever that means), R is the "the strength of the mean signal", and σ_N represents noise that is present even if no stimulus is presented (p. 3138); this is what they call the irreducible part of the noise. It is not clear in which sense σ_R represents the noise. If the "irreducible" part of the noise and the stimulus generated part of the noise are independent, one would expect the variance to be of the form $\sigma_R^2 \propto kR^q + \sigma_N^2$, but this is not what Tyler and Chen write in their Eq. (18). If, on the other hand, σ_R is meant to represent a standard deviation (after all, from their Eq. (10) one may take that σ_N is a standard deviation), σ_R may also be a standard deviation. But then σ_R should be proportional to $(kR^q + \sigma_N^2)^{1/2}$. Note also that σ_R, R and σ_N are independent of time, and consequently they reflect some property of the activity as it develops in time, but is not clear which property.

the notion of multiplicative noise could mean may be in order.

Quite generally, a neural system can be conceived as a dynamical system, which is never noise free in a strict sense. The most general description of such systems is in terms of stochastic differential equations [15] catering also for multiplicative noise (Honerkamp, 1990); systems with only additive noise are special cases. The details of the definition of multiplicative noise within the framework of stochastic differential equations are beyond the scope of this paper, but it may be noted that this type of noise is conceived as resulting from stochastic fluctuations that are external to the system under consideration (Horsthemke & Lefever, 1984). To illustrate, consider the (extremely simplified) case of a neuron characterised by the impulse response [16] $\mu_0 \exp(-\lambda t)$. Due to noise from the environment of the neuron the parameter λ may fluctuate; one may put $\lambda = \lambda_0 + \zeta(t)$, $\lambda_0 > 0$ a constant and $\zeta(t)$ a randomly fluctuating function of time. A similar argument can be applied to the constant μ_0. In any case, the response of the neuron to a sufficiently brief pulse is then $g(t) = \mu_0 \exp(-\lambda t) = \mu_0 \exp(-\lambda_0 t)\xi(t)$ with $\xi(t) = \exp(-\zeta(t)t)$ representing multiplicative noise; this may be used to find the distribution function for the noise such that the psychometric function is again the Weibull. [17]

The notion of multiplicative noise will also be encountered when considering multiplicative (i.e. branching or cascaded) stochastic processes. In particular, the relation between the stimulus and the corresponding neuronal activity may be represented by two cascaded Poisson processes. The basic elements of the model have been presented in Section 4.3. Other than for a simple superposition of Poisson processes the variance of the counts at the detector is not equal to the mean, but in excess of the mean; a similar result was also derived by Tuckwell (1977), who modelled the membrane potential of a single neuron, responding to synaptic inputs with Poisson rate parameters, in terms of a stochastic differential equation. Tyler and Chen just assume that the standard deviation of the random variable representing activation is a power function of the mean activity, without specifying the processes behind this relation.

So at least from a theoretical point of view there is ample evidence that multiplicative noise is a component of the activity in the visual system, or generally in any sensory system. Taking multiplicative noise explicitly into account will complicate the modelling of neural processes with respect to psychophysical data considerably, either because one has to deal with stochastic differential equations, or because one has to generalise the model of Teich et al. so that it can cope with stimulus patterns encountered in experiments aiming, for instance, at pattern coding processes and therefore dealing with populations of neurons having particular forms of receptive fields or even receptive fields adapting to the input. So the question is whether the multiplicative noise component can be neglected.

A possible answer to the question may be derived from cross-validation studies where the results of one experiment are employed to predict the results of another. For instance, Roufs and Blommaert (1981) estimated the impulse and the step responses of sustained and transient channels and predicted the step response data from the impulse response data, assuming peak detection, i.e. additive noise, as defined above. Mortensen et al. (1991) discussed the same data, assuming detection by level crossing in additive Gaussian noise; the results agreed (not surprisingly, since detection by level-crossing could be approximated by the peak detection) with those of Roufs et al. Meinhardt and Mortensen (1998) predicted threshold data for a stimulus defined as a rectangular bar by data from experiments employing sawtooth patterns, again assuming peak detection. While these findings by no means preclude the existence of multiplicative noise, they suggest that in psychophysical models of detection the role of multiplicative noise may indeed be negligible.

5.5. Criteria for choosing or rejecting the Weibull

It seems that psychophysical data or general principles pointing uniquely to the Weibull as the "true" psychometric function are not yet known. One may invoke a principle like the weakest-link principle from the theory of reliability of materials, equating for instance the breaking of a link of a chain with the supra-threshold ($\eta_i > \eta_s$) activity of a neuron, which is known to imply the Weibull function. Galambos (1978) gave a stimulating presentation of this principle with respect to asymptotically independent random variables allowing for arbi-

[15] Let $X_t = X(t)$ be a vector which represents the state of a system at time t; for instance, the ith component could represent the activity of the ith neuron in a network of neurons. A stochastic differential equation for X has the form $dX_t/dt = a(X_t, t) + b(X_t, t)\xi_t$, where a and b a functions of X_t and t and ξ_t represents noise. If b equals a constant, ξ_t is additive, otherwise ξ_t is multiplicative. The characterisation of detection as a level-crossing problem may also be approached in terms of SDEs: Buonocore et al. (1987) provide a solution in closed form for the special case that "noise" can be represented as an Ornstein-Uhlenbeck process, which is stationary and Gaussian. A discussion of the results of Buonocore et al. with respect to detection processes is beyond the scope of this paper and will be left to future work.

[16] It is usually assumed that the system can be linearised around a time-independent reference state so that the response of a channel to a stimulus can be characterised by a convolution of the stimulus with the impulse response of the linearised system.

[17] Let $\eta = g\xi$, ξ noise. Then $P(\eta \leqslant \eta_s) = \exp(-g^\beta)$ with $g = ch$ if the distribution function of the noise is given by $P(\xi \leqslant y/g) = \exp(-(\eta_s g/y)^\beta)$, $\xi > 0$. Thus there exists a distribution function for multiplicative noise implying the Weibull. Whether the assumption of such a noise distribution makes psychophysical sense is open to discussion, though.

trary dependencies among neighbouring units (corresponding here to neural units) that are not too far apart from each other. Unfortunately, the principle does not apply: the distribution for the complete network of neurons is not necessarily equal to that for a single neural unit, since more than one neuron may signal the stimulus.

Since the implication $P(\xi \leqslant \eta_0) = 1$ for finite η_0 is no reason to reject the Weibull (see the remarks on high-threshold detection), the question is which criteria should be invoked to decide against the Weibull distribution as an approximation to be chosen in a given experimental context.

Tyler and Chen argue that because the Weibull density of the noise assumes "bizarre" forms for sufficiently small values of β, the Weibull distribution should be discarded as a possible noise distribution. Bizarreness is here defined as sufficiently strong deviation from the shape of a Gaussian density function, which is considered to be more plausible than the Weibull distribution since the Gaussian can be motivated by the central limit theorem (CLT). This is, however, not a strong argument unless one has good reasons to believe that the CLT holds without restriction. Recall that Gershon et al. (1998) suggested, on the basis of empirical findings, a Gaussian distribution truncated at zero. Depending on the location of the mode of this distribution one may find strong deviations from a nontruncated Gaussian, suggesting that the CLT cannot be invoked in a general, sort of sweeping way.

On the other hand, there is a clear and simple criterion that indicates when the Weibull can be rejected: when even the generalised version allowing for true false alarms cannot be fitted to the data.

Acknowledgements

I am greatly indebted to Yvonne Grabbe for many clarifying discussions.

Appendix A

A.1. Proof of Theorem 1

A.1.1. Detection by PS

Necessity: Suppose now that (9) holds. The stimulus is detected if $\{\eta_{\mathrm{ps}} > \eta_0\}$ occurs, and $P(\eta_{\mathrm{ps}} > \eta_0) = 1 - P(\eta_{\mathrm{ps}} \leqslant \eta_0)$. For stochastic independent η_i, $P(\eta_{\mathrm{ps}} \leqslant \eta_0) = P(\eta_1 \leqslant \eta_0 \cap \cdots \cap \eta_n \leqslant \eta_0) = \prod_i F(\eta_0 - g_i)$, and

$$1 - \psi_{\mathrm{w}}(c) = \exp(-\alpha c^\beta) = P(\eta \leqslant \eta_0). \tag{A.1}$$

It follows that [18]

$$\mathrm{e}^{-g_i^\beta} = F(\eta_0 - g_i) \tag{A.2}$$

for all i, so we may drop the index i for simplicity. Let $F(x) = P(\xi \leqslant x)$, and let $x = \eta_0 - g$, $g \geqslant 0$. From (9) it follows that $g = 0$ implies $P(\xi \leqslant \eta_0) = 1$, and for $g \to \infty$ it follows that $P(\xi \leqslant \eta_0 - g) \to 0$, so that ξ may assume values on $(-\infty, \eta_0]$. Then

$$\begin{aligned} P(\xi \leqslant x) = P(\xi \leqslant \eta_0 - g) &= \exp(-g^\beta) \\ &= \exp(-(\eta_0 - \eta_0 + g)^\beta) \\ &= \exp(-(\eta_0 - (\eta_0 - g)^\beta) = \exp(-(\eta_0 - x)^\beta). \end{aligned}$$

If $x \to \eta_0$, then $P(\xi \leqslant x) \to 1$, and for $x \to -\infty$ one has $P(\xi \leqslant x) \to 0$, so $F(x)$ defines indeed a distribution function on $(-\infty, \eta_0]$.

Sufficiency: Suppose the distribution of ξ is given by (11). For $g_i \geqslant 0$ for all $c > 0$,

$$\begin{aligned} P(g_i + \xi > \eta_0) = P(\xi > \eta_0 - g_i) &= 1 - P(\xi \leqslant \eta_0 - g) \\ &= 1 - \exp(-(\eta_0 - (\eta_0 - g_i))^\beta) \\ &= 1 - \exp(-g_i^\beta), \end{aligned}$$

so that $\psi_{\mathrm{w}}(c) = 1 - \prod_i P(\xi \leqslant \eta_0 - g_i) = 1 - \exp(-\sum_i g_i^\beta)$, i.e. (11) is indeed sufficient. □

Detection by NP: The proof for the case of detection by PS transfers to detection by NP, for the case $n = 1$, with $g = c(\sum_i h_i^p)^{1/p}$. □

A.2. Derivation of expected values

The expected value of η is given by

$$\begin{aligned} E(\eta_{\mathrm{ps}}) &= \int_{-\infty}^{\infty} \eta f(\eta)\,\mathrm{d}\eta = \int_{-\infty}^{\eta_0+g} \eta f(\eta)\,\mathrm{d}\eta \\ &= n\beta \int_0^{\eta_0+g} \eta(\eta_0 + g - \eta)^{\beta-1} \\ &\quad \times \exp(-n(\eta_0 + g - \eta)^\beta)\,\mathrm{d}\eta. \end{aligned}$$

Let $y = n(\eta_0 + g - \eta)^\beta$. Then

$$\eta = \eta_0 + g - \left(\frac{y}{n}\right)^{1/\beta}$$
$$\frac{\mathrm{d}\eta}{\mathrm{d}y} = -\frac{1}{n\beta}\left(\frac{y}{n}\right)^{1/\beta-1}.$$

Further, $y \to \infty$ for $\eta \to -\infty$, $y \to 0$ for $\eta \to \eta_0 + g$. It follows that

$$\begin{aligned} E(\eta_{\mathrm{ps}}) &= \int_0^\infty \left(\eta_0 + g - \left(\frac{y}{n}\right)^{1/\beta}\right)\left(\frac{y}{n}\right)^{1/\beta-1}\left(\frac{y}{n}\right)^{1-1/\beta}\mathrm{e}^{-y}\,\mathrm{d}y \\ &= \int_0^\infty \left(\eta_0 + g - \left(\frac{1}{n}y\right)^{1/\beta}\right)\mathrm{e}^{-y}\,\mathrm{d}y \\ &= (\eta_0 + g) - \left(\frac{1}{n}\right)^{1/\beta}\int_0^\infty y^{1/\beta}\mathrm{e}^{-y}\,\mathrm{d}y \\ &= \eta_0 + g - (1/n)^{1/\beta}\Gamma(1 + 1/\beta), \end{aligned} \tag{A.3}$$

where $\Gamma(p) = \int_0^\infty y^{p-1}\mathrm{e}^{-y}\,\mathrm{d}y$.

[18] This can be shown explicitly: (A.1) implies $-\sum_i g_i^\beta = \sum_i \log F(\eta_0 - g_i)$. Differentiation with respect to g_i leads to (A.2).

The second moment of η is given by

$$E(\eta^2) = \int_{-\infty}^{\eta_0+g} \eta^2 f(\eta)\,d\eta$$
$$= \int_0^{\infty} ((\eta_0+g)^2 - (1/n)^{2/\beta} - 2(\eta_0+g)(1/n)^{1/\beta})e^{-y}\,dy,$$
$$= (\eta_0+g)^2 + (1/n)^{2/\beta} - 2(\eta_0+g)(1/n)^{1/\beta}\int_0^{\infty} y^{1/\beta}e^{-y}\,dy$$
$$= (\eta_0+g)^2 + (1/n)^{2/\beta}\Gamma(1+2/\beta) - 2(\eta_0+g)(1/n)^{1/\beta}\Gamma(1+1/\beta).$$

The variance of η is given by $\mathrm{Var}(\eta) = E(\eta^2) - E^2(\eta)$. After a little algebra one finds

$$\mathrm{Var}(\eta) = (1/n)^{2/\beta}(\Gamma(1+2/\beta) - \Gamma^2(1+1/\beta)). \qquad \square \tag{A.4}$$

A.3. Negative intensities and extreme value statistics

A.3.1. Norming constants and limiting distributions

Let $G_n = F^n$, and $\omega_F = \sup\{x|F(x) < 1\}$, i.e. ω_F is the maximal value x may assume. Let x be such that $0 < F(x) < 1$; then $G_n \to 0$ for increasing value of n, i.e. G_n will be degenerate. For G_n not to become degenerate, x has to assume values in the neighbourhood of ω_F. As shown in the theory of extreme values there exist *norming constants* a_n and b_n such that for any x for which $G_n(x)$ is not degenerate, the condition $F^n(b_n x + a_n) = F^n(x_n) = G_n(x)$ is satisfied, with $x_n = b_n x + a_n$, and $F(x_n) \to 1$, i.e. $x_n \to \omega_F$. For $n \to \infty$, $G_n \to G_\infty$, and G_∞ is called limiting distribution; G_∞ may be used to approximate $F^n(x_n)$ by $G_\infty(x) = G_\infty((x_n - a_n)/b_n)$. G_∞ is unique only up to a linear transformation, i.e. if $G_\infty(x)$ is a limiting distribution with corresponding norming constants a_n, b_n, then $G^*_\infty(x) = G_\infty(Bx + A)$ is also a limiting distribution, and the corresponding norming constants a^*_n, b^*_n, satisfy the condition

$$a^*_n = b_n A + a_n, \quad b^*_n = b_n B, \tag{A.5}$$

(Galambos, 1978, p. 61).

If F is of Weibull type, then G_n and G_∞ or G^*_∞ are also of Weibull type and the approximations become exact. In order to allow for arbitrary transformations $Bx + A$ we employ $G^*_\infty(x) = G_\infty(Bx + A)$, so one has

$$F^n(x^*_n) = G_\infty(Bx + A), \tag{A.6}$$

where, $G_\infty = \exp(-(-x)^\beta)$, $x < 0$ is the standard form of the limiting distribution (cf. Galambos, 1978; p. 51). The corresponding norming constants are known to be $a_n = \omega_F$, $b_n = \omega_F - \inf\{x|1 - F(x) \leqslant 1/n\}$, and from (11) it follows that $a_n = \eta_0$, $b_n = (1/n)^{1/\beta}$, (cf. Galambos, 1978, p. 52), where in the determination of b_n the fact that G_∞ is also of Weibull type has been exploited (cf. Leadbetter, Lindgren, & Rootzén, 1983, p. 25). For any value p_0 of the probability of detection, the corresponding contrasts c_1 and c_n are such that $1 - p_0 = F^n(x^*_n) = F(x)$, and from (A.6) one has $F(x) = \exp(-(\eta_0 - x)^\beta) = G_\infty(Bx + A)$. The definition of G_∞ implies then the choice $B = 1$, $A = -\eta_0$, and $x^*_n = b^*_n x + a^*_n$ implies, according to (A.5), $b^*_n = b_n$ and $a^*_n = a_n - b_n\eta_0$.

A.3.2. Shift of location and the sign of intensities

Generally, the transformation $x_n = b_n x + a_n$ or $x^*_n = b^*_n x + a^*_n$ means that location and the spread of the distribution is changed. It has to be shown that the shift of location does not imply the need for negative intensities. We consider the transformation x^*_n, containing the case x_n as a special case ($\eta_0 = 0$).

It is $x^*_n = b^*_n x + a^*_n = b_n x + a^*_n$, with $x = \eta_0 - g_1 = \eta_0 - c_1 h$ and $x^*_n = \eta_0 - g_n = \eta_0 - c_n h$. It follows that $x = (x^*_n - a^*_n)/b_n$, so that x and consequently c_1 can be determined from x^*_n. We have to show that $x = \eta_0 - c_1 h$ is such that $c_1 > 0$. Obviously, from the definition of x^*_n and a^*_n above, $b_n x = x^*_n - a^*_n = b_n(\eta_0 - c_1 h) = \eta_0 - c_n h - \eta_0(1 - b_n) = -c_n h + b_n\eta_0$, from which $b_n c_1 h = c_n h$ and consequently

$$c_1 = c_n/b_n = n^{1/\beta}c_n \tag{A.7}$$

follows. So, regardless of the value of n, the shift by a^*_n does not imply the need for negative intensities. $\square$

A.4. Proof of Theorem 2

One has

$$\int_\xi^{\eta_0+ch} f_{\mathbf{sn}}(\eta)\,d\eta = n\beta\int_\xi^{\eta_0+ch}(\eta_0+g-\eta)^{\beta-1} \times \exp(-n(\xi_o+g-\eta)^\beta)\,d\eta.$$

Let $y = n(\eta_0+g-\eta)^\beta$; then

$$\eta = \eta_0 + g - (1/n)^{1/\beta}, \quad (\eta_0+g-\eta)^{\beta-1} = (y/n)^{(\beta-1)/\beta}$$

so that $d\eta = -(1/n\beta)(y/n)^{1-1/\beta}\,dy$. For $\eta \to \eta_0 + g$ follows $y \to 0$, for $\eta \to \xi$ it follows that $y \to n(\eta_0+g-\xi)^\beta$

Then

$$\int_\xi^{\eta_0+ch} f_{\mathbf{sn}}(\eta)\,d\eta = \int_{n(\eta_0+g-\xi)^\beta}^{\infty}(y/n)^{(\beta-1)/\beta}(y/n)^{1-1/\beta}e^{-y}\,dy$$
$$= \int_{n(\eta_0+g-\xi)^\beta}^{\infty} e^{-y}\,dy$$
$$= 1 - e^{-n(\eta_0+g-\xi)^\beta}.r \tag{A.8}$$

It follows that

$$P(\text{“yes”}|c) = \int_{-\infty}^{\eta_0} f_{\mathbf{n}}(\xi)(1 - \mathrm{e}^{-n(\eta_0+g-\xi)^\beta})\,\mathrm{d}\xi = 1 - \int_{-\infty}^{\eta_0} f_{\mathbf{n}}(\xi)\mathrm{e}^{-n(\eta_0+g-\xi)^\beta}\,\mathrm{d}\xi.$$

Now

$$I \stackrel{\text{def}}{=} \int_{-\infty}^{\eta_0} f_{\mathbf{n}}(\xi)\mathrm{e}^{-n(\eta_0+g-\xi)^\beta}\,\mathrm{d}\xi = n\beta \int_{-\infty}^{\eta_0} (\eta_0 - \xi)^{\beta-1}\mathrm{e}^{-n(\eta_0-\xi)^\beta - n(\eta_0+g-\xi)^\beta}\,\mathrm{d}\xi$$

Let $y = n(\eta_0 - \xi)^\beta$; then $\xi = \eta_0 - (y/n)^{1/\beta}$ and $\mathrm{d}\xi = -(1/n\beta)(y/n)^{1/\beta-1}$. Then

$$I = \int_0^\infty (y/n)^{1\beta-1}(y/n)^{1-1/\beta}\mathrm{e}^{-y-n((y/n)^{1/\beta}+g)^\beta}\,\mathrm{d}y = \int_0^\infty \mathrm{e}^{-y-n((y/n)^{1/\beta}+g)^\beta}\,\mathrm{d}y,$$

which may be rewritten as

$$I = \int_0^\infty \mathrm{e}^{-y-(y^{1/\beta}+n^{1/\beta}g)^\beta}\,\mathrm{d}y. \tag{A.9}$$

So we have

$$P(\text{“yes”}|c) = 1 - \int_0^\infty \mathrm{e}^{-y-(y^{1/\beta}+n^{1/\beta}g)^\beta}\,\mathrm{d}y; \tag{A.10}$$

it does not seem possible to find a closed expression for the integral I. However, one sees that for $c \to 0$, i.e. $g \to 0$, $I \to \int_0^\infty \mathrm{e}^{-2y}\,\mathrm{d}y = 1/2$, as it should be. □

A.5. Tyler and Chen's derivations

According to Tyler and Chen, the density of the noise is given by

$$D_\beta(r) = \beta\rho^{\beta-1}\mathrm{e}^{-\rho^\beta}, \quad \rho = R - R_\Theta - r, \tag{A.11}$$

with R the "mean signal" (also called the "effective mean response over the set of channels", cf. p. 3124), R_Θ the threshold and r the random variable representing the "noise around the mean". The claim that (A.11) defines the density of the noise, calls for some comments:

(1) Obviously, $R_i = g_i$ as defined in this paper, and $R = (\sum_i R_i^\beta)^{1/\beta}$, which may be deduced from their Eq. (2), so $R = g = (\sum_i g_i^\beta)^{1/\beta}$; Tyler and Chen's hint to Robson and Graham (1979) is somewhat misleading because Robson et al. refer to a model not really corresponding to that of Tyler and Chen.

(2) (A.11) corresponds to the density (26) for detection by pooled activity, with $R_\Theta = \eta_s = \eta_0$ and $r = \eta_{\text{np}}$, although the signs of R_Θ and η_0 do not agree, due to a slight error when Tyler and Chen differentiate their Eq. (3) in order to arrive at (A.11) (see (A.12) below). The density (A.11) does not correspond to the density (17) for η_{ps} for detection by PS among not identically activated channels, although this type of detection seems to be what the authors want to consider, since in the derivation of (A.11) they do not refer to detection by NP.

Note that confounding detection by PS with detection by NP implies that the pooling parameter p is implicitly set equal to β.

(3) Although the authors refer to R as the "effective mean", R is not the mean of η_{ps}; as mentioned in Section 2.3.2, no closed expression for the mean (i.e. expected value) seems to exists for the case of not identically distributed η_i, and in any case this mean will depend somehow upon $\Gamma(1 + 1/\beta)$. R is not the expected value of r in case of detection by NP either, which would be given by $\eta_0 + R - \Gamma(1 + 1/\beta)$ (cf. (27)).

Note that in all expressions for expected values the term $\Gamma(1 + 1/\beta)$ appears. One may, of course, transform the variable η_{np} such that its expected value equals $R = g$: let $Z = \eta_{\text{ps}} - \eta_0 + \Gamma$, $\Gamma = \Gamma(1 + 1/\beta)$; then $E(Z) = \eta_{\text{ps}} - \eta_0 + \Gamma = \eta_0 + g - \Gamma - \eta_0 + \Gamma = g$. The distribution function of Z is then given by $P(Z \leqslant z) = P(\xi \leqslant z + g + \eta_0 - \Gamma) = \exp(-(g + \Gamma - z)^\beta)$, assuming $p = \beta$, and the density is $f_Z(z) = \beta(g + \Gamma - z)^{\beta-1}\exp(-(g + \Gamma - z)^\beta)$, which is not equivalent to (A.11). Getting rid of the term Γ in the expectation means to get Γ back in the expression for the density function.

Tyler and Chen derive D_β from their Eq. (3), p. 3124, that is from

$$\psi = 1 - \mathrm{e}^{-R^\beta} = \int_{R_\Theta}^\infty D_\beta(r - R)\,\mathrm{d}r. \tag{A.12}$$

It seems that the authors' idea was to arrive at D_β by differentiating both sides with respect to R, which is strange since R is a parameter in D_β, i.e. of the distribution of the "noise around the mean" R. The differentiation of the right hand side of the above equation with respect to R does not make sense, since the variable of interest is r (the differentiation should be with respect to the upper and lower limits of r).

References

Aertsen, A., Erb, M., & Palm, G. (1994). Dynamics of functional coupling in the cerebral cortex: an attempt at a model-based interpretation. *Physica D, 75*, 103–128.

Arieli, A., Sterkin, A., Grinvald, A., & Aertsen, A. (1996). Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science, 273*, 1868–1871.

Blommaert, F. J. H., & Roufs, J. A. J. (1987). Prediction of thresholds and latency on the basis of experimentally determined impulse responses. *Biological Cybernetics, 36*, 329–344.

Bradley, A., Skottun, B., Ohzawa, I., Sclar, G., & Freeman, A. D. (1987). Visual orientation and spatial frequency discrimination: A comparison of single neurons and behavior. *Journal of Neurophysiology, 57*, 755–772.

Braun, J., Koch, C., & Davis, J. L. (2001). *Visual attention nd cortical circuits*. Cambridge, MA: The MIT Press.

Brindley, G. S. (1963). The relation of frequency of detection to intensity of stimulus for a system of many independent detectors each of which is stimulated by a *m*-quantum coincidence. *Journal of Physiology, 169*, 412–415.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Visual Neuroscience, 10*, 1157–1169.

Buonocore, A., Nobile, A. G., & Ricciardi, L. M. (1987). A new integral equation for the evaluation of first-passage-time probability densities. *Advances in Applied Probability, 19*, 764–800.

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research, 20*, 847–856.

Dean, A. F. (1981). The variability of discharge of simple cells of cat striate cortex. *Experimental Brain Research, 44*, 437–440.

du Buf, J. M. H. (1992). Modeling spatial vision at the threshold level. *Spatial Vision, 6*, 25–60.

du Buf, J. M. H. (1993). Responses of simple cells: events, interferences, and ambiguities. *Biological Cybernetics, 68*, 321–333.

du Buf, J. M. H. (1994). Ramp edges, Mach bands, and the functional significance of the simple cell assembly. *Biological Cybernetics, 70*, 449–461.

Galambos, J. (1978). *The asymptotic theory if extreme order statistics*. NY: John Wiley & sons.

Gershon, E. D., Wiener, M. C., Latham, P. E., & Richmond, B. J. (1998). Coding strategies in monkey V1 and inferior temporal cortices. *Journal of Neurophysiology, 79*, 1135–1144.

Gerstner, W., Ritz, R., & van Hemmen, J. L. (1993). Why spikes? Hebbian learning and retrieval of time-resolved excitation patterns. *Biological Cybernetics, 69*, 503–515.

Gilbert, C., Ito, M., Kapadia, M., & Westheimer, G. (2000). Interactions between attention, context and learning in primary visual cortex. *Vision Research, 40*, 1217–1226.

Graham, N. (1977). Visual detection of a periodic spatial stimuli by probability summation among narrow band channels. *Vision Research, 17*, 637–652.

Graham, N. (1989). *Visual Pattern Analysers*. Oxford: Oxford University Press.

Green, D. M., & Luce, R. D. (1975). Parallel psychometric functions from a set of independent detectors. *Psychological Review, 82*, 483–486.

Heller, J., Hertz, J. A., Kjaer, T. W., & Richmond, B. J. (1995). Information flow and temporal coding in primate pattern vision. *Journal of Computational Neuroscience, 2*, 175–193.

Horsthemke, W., & Lefever, R. (1984). *Noise-induced transitions*. Berlin, Heidelberg, New York, Tokyo: Springer-Verlag.

Honerkamp, J. (1990). *Stochastische Dynamische Systeme*. Weinheim: VHC Verlagsgesellschaft mbH.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions (Vol. 1)*. NY: John Wiley & Sons.

Kotz, S., & Nadarajah, S. (2000). *Extreme value distributions*. London: Imperial College Press.

Leadbetter, M. R., Lindgren, G., & Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Berlin: Springer-Verlag.

Lillywhite, P. G. (1981). Multiplicative intrinsic noise and the limits to visual performance. *Vision Research, 21*, 291–296.

Logvinenko, A. D. (1993). Lack of convexity of threshold curves for compound grating: implications for modeling visual pattern detection. *Biological Cybernetics, 70*, 50–64.

Logvinenko, A. D. (1995). On deriving analyser characteristics from summation-at-threshold data. *Biological Cybernetics, 73*, 547–552.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, E. Galanter (Eds.), *Handbook of mathematical psychology (Vol. I)*. NY: John Wiley & Sons.

Marčelja, S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America, 70*, 1297–1300.

McGill, W. (1967). Neural counting mechanisms and energy detection in audition. *Journal of Mathematical Psychology, 4*, 351–376.

Meinhardt, G. (1999). Evidence for different nonlinear summation schemes for lines and gratings at threshold. *Biological Cybernetics, 81*, 263–277.

Meinhardt, G. (2000). Detection of compound spatial patterns: further evidence for different channel interactions. *Biological Cybernetics, 82*, 269–282.

Meinhardt, G., & Mortensen, U. (1998). Detection of aperiodic test patterns by pattern specific detectors revealed by subthreshold summation. *Biological Cybernetics, 79*, 413–425.

Mortensen, U. (1988). Visual contrast detection by a single channel versus probability summation among channels. *Biological Cybernetics, 59*, 137–147.

Mortensen, U., & Suhl, U. (1991). An evaluation of sensory noise in the humna visual system. *Biological Cybernetics, 66*, 37–47.

Nachmias, J. (1967). Effect of exposure duration on visual contrast sensitivity with square wave gratings. *Journal of the Optical Society of America, 57*, 421–427.

Neyman, J. (1939). A new class of "contagious" distributions, applicable in entomology and bacteriology. *Annals of Mathematical Statistics, 10*, 35–57.

Papoulis, A. (1965). *Probability, random variables, and stochastic processes*. Tokyo: McGraw-Hill Kogakusha, Ltd.

Parzen, E. (1962). *Stochastic Processes*. San Francisco: Holden-Day, Inc.

Pelli, D. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A, 2*, 1508–1531.

Quick, R. F. (1974). A vector-magnitude model of contrast detection. *Kybernetik, 16*, 65–67.

Reeves, A., Wu, S., & Schirillo, J. (1998). The effect of photon noise on the detection of white flashes. *Vision Research, 38*, 691–703.

Robson, J. G., & Graham, N. (1979). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research, 1979*, 409–418.

Rolls, E. T., & Deco, G. (2002). *Computational neuroscience of vision*. Oxford: Oxford University Press.

Roufs, J. A. J., & Blommaert, F. J. J. (1981). Temporal impulses and step responses of the human eye obtained psychophysically by means of a drift-correcting perturbation technique. *Vision Research, 21*, 1203–1221.

Snowden, R. J., Treue, S., & Andersen, R. A. (1992). The response of neurons in areas V1 and MT of the rhesus monkey to moving random dot patterns. *Experimental Brain Research, 88*, 3898–4000.

Teich, M. C. (1981). Role of the doubly stochastic Neyman type-A and Thomas counting distributions in photon detection. *Applied Optics, 20*, 2457–2467.

Teich, M. C., Prucnal, P. R., Vannucci, G., Breton, M. E., & McGill, W. J. (1982). Nultiplication noise in the human visual system at threshold: 1. Quantum fluctuations and minimum detectable energy. *Journal of the Optical Society of America, 72*, 419–431.

Tolhurst, D. J., Movshon, J. A., & Thompson, I. D. (1981). The dependence of response amplitude and and variance of cat visual cortical neurons in the behaving monkey. *Experimental Brain Research, 41*, 414–419.

Tuckwell, H. (1977). In stochastic models of the activity of single neurons. *Journal of Theorotical Biology, 65*, 783–785.

Tyler, C. W., & Chen, C. C. (2000). Signal detection theory in the 2AFC paradigm: attention, channel uncertainty and probability summation. *Vision Research, 40*, 3121–3144.

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H., & Aertsen, A. (1995). Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature, 373*, 515–518.

Watson, A. B. (1979). Probability summation over time. *Vision Research, 19*, 515–522.
Wilson, H. R. (1980). A transducer function for threshold and suprathreshold human vision. *Biological Cybernetics, 38*, 171–178.
Wilson, H. R., & Bergen, J. R. (1979). A four mechanism model of threshold spatial vision. *Vision Research, 19*, 515–522.
Wong, E. (1971). *Stochastic processes in information and dynamical systems*. NY: McGraw-Hill Book Company.