

Signal detection theory and vestibular thresholds: I. Basic theory and practical considerations

Daniel M. Merfeld

Received: 29 October 2010 / Accepted: 30 December 2010 / Published online: 26 February 2011
© Springer-Verlag 2011

Abstract Detection theory has been applied to the measurement of vestibular thresholds and vestibular sensory integration. Yet, a formal detection theory analysis of vestibular responses has not been published. Such a de novo analysis seems warranted because the vestibular system has characteristics that differ from other sensory systems, which impacts the application of detection theory. For example, the physical stimuli evoking vestibular responses are typically bidirectional (e.g., leftward/rightward); this bidirectional nature of vestibular responses leads to another characteristic—what is sometimes called vestibular bias—that must also be considered, since it can impact threshold measurements, including thresholds found using staircase procedures. This paper develops a basic model of vestibular noise and then analyzes this model for four standard paradigms—one-interval recognition, one-interval detection, two-interval detection, and two-interval recognition. While any of these paradigms might be justified for a specific application, it is concluded that one-interval recognition paradigms have advantages over other paradigms for many vestibular applications. One-interval recognition is favored over one-interval detection because it lends itself to a fixed detection boundary, is more efficient, and is less impacted by device vibration. One-interval recognition is generally favored over two-interval recognition because it assesses vestibular bias and can require substantially less time than two-interval tasks.

Keywords Detection theory · Spatial orientation · Vestibular · Perception · Threshold · Psychophysics

Introduction

“Signal detection theory” has long been used to guide the design and analysis of vestibular studies (e.g., Clark and Stewart 1968; Doty 1969; Ormsby 1974; Benson et al. 1986, 1989; Mah et al. 1989; Carpenter-Smith et al. 1995), but, after nearly a 20-year hiatus, there has been a recent resurgence of interest in the application of signal detection theory to vestibular responses (e.g., Gu et al. 2007; Sadeghi et al. 2007; De Vrijer et al. 2008; Grabherr et al. 2008; Zupan and Merfeld 2008; Barnett-Cowan and Harris 2009; MacNeilage et al. 2010; Mallery et al. 2010). In part, this resurgence has occurred because detection theory can help address some unique technical and practical challenges associated with vestibular psychophysics.

What is signal detection theory? In brief, detection theory is nothing more than the application of standard statistical hypothesis testing to the detection of a specific event (“signal”) despite the presence of noise. Therefore, those who understand the theoretical basis underlying Student’s *t*-test will readily understand the basis of detection theory. In other words, signal detection theory is a general statistical approach that helps make decisions about signals with noise. It does not address questions of how one might optimally filter a signal or how one might combine two or more noisy signals, both of which are the purview of estimation theory. Signal detection theory is often just called detection theory; other names include “hypothesis testing” and “decision theory”. Detection theory has been applied to a broad range of physiological responses, with an immense influence on psychophysics. For example,

D. M. Merfeld (✉)
Jenks Vestibular Research Laboratory, Massachusetts Eye and Ear Infirmary, Boston, MA, USA
e-mail: dan_merfeld@meei.harvard.edu

even when not explicitly noted, detection theory is implicitly invoked when thresholds are measured using tasks that require the subject to select one of two alternative answers, and the data are fit with some sort of cumulative distribution (e.g., cumulative Gaussian). In fact, it will be shown that detection theory directly relates thresholds to the standard deviation of the noise present.

What is “discrimination”? And how does discrimination relate to “detection” and “recognition”? According to Macmillan and Creelman (2005) and others (e.g., Treutwein 1995), discrimination is the ability to tell two stimuli apart. There are two types of discrimination. When one of the stimulus classes is a null stimulus, the task is called detection; the standard hearing test where a subject indicates whether they hear or do not hear a tone is a very common detection task. When neither stimulus class is null, the task is called recognition. For example, a subject discriminating leftward from rightward motion (or leftward from rightward orientation) is a common vestibular direction-recognition task (e.g., Benson et al. 1986, 1989; Carpenter-Smith et al. 1995; Gu et al. 2007; De Vrijer et al. 2008; Grabherr et al. 2008; Zupan and Merfeld 2008). Another vestibular recognition paradigm is exemplified by Mallery et al. (2010), where subjects compared two stimuli to determine whether the test stimulus was greater than a non-zero reference. Consistent with historical usage, we will refer to the general theory as detection theory. Otherwise, we will reserve the terms detect and detection to refer to paradigms, where one of the stimulus classes is the null stimulus (i.e., no motion).

Detection theory uses signals provided by brain “estimation” processes. Estimation theory, which is not the focus of this paper, describes the application of statistical signal processing to extract information from noisy signals but not decision-making per se. Specifically, estimation theory helps estimate variables in the presence of noise. Standard approaches include minimum variance unbiased (MVU) estimation, simple linear weighting maximum likelihood (ML) estimation, Wiener filters, Bayesian *maxima* a posteriori (MAP) estimation, and Kalman filters. In this paper, we will assume that a noisy signal has been estimated using one of the above optimal techniques, by simple filtering, or by one of a myriad of suboptimal estimation approaches. Given such a noisy signal, detection theory then guides the decision-making process. (For recognition, did I move to the right or left? For detection, did I move or not move?) Obviously, how the signals are estimated and sampled is critical to signal detection, but this is a separate topic that requires separate coverage and cannot be summarized in a few pithy paragraphs.

In a short paper like this, we cannot be comprehensive, so some issues are left partially explored. To maintain focus, this paper concentrates primarily on vestibular

psychophysical responses, but much of the information relates to the use of detection theory for other behavioral responses (e.g., VOR thresholds, etc.).

To keep the length reasonable, we assume a rudimentary knowledge of detection theory. For those interested in pursuing these topics to a deeper understanding, several books and papers are recommended. The book by Macmillan and Creelman (2005) provides an excellent introduction to the application of signal detection theory to psychophysics, and the book by Green and Swets (1966) is considered a classic. For those interested in a more general, more theoretical, and more mathematical coverage, two books provide a very good introduction to estimation (Kay 1993) and detection (Kay 1998). Those interested in Kalman filtering—a classic advanced estimation approach—might consider Gelb (1974) and/or Brown and Hwang (1992). Three reviews are also recommended—one focused on the use of adaptive procedures for psychophysical studies (Leek 2001) and two that focus on fitting psychometric functions (Wichmann and Hill 2001a, b), where a psychometric function describes the sigmoid-like shape (e.g., Fig. 1d) that typically occurs when the percentage of correct responses (or related parameter) is plotted versus stimulus amplitude (or other physical stimulus parameter).

Before proceeding, it is important to note that no general psychophysical model of perception for any sensory modality has ever been built exclusively on the incremental knowledge gleaned from detection theory. Therefore, while detection theory is powerful when applied correctly, recall that it only evaluates the ability to tell things apart and does not estimate their magnitude. Hence, the application of detection theory complements—and does not replace—the use of other standard psychophysical techniques (e.g., Guedry 1974) like magnitude estimation.

In this paper, we apply detection theory to vestibular responses. Because vestibular responses have some unique characteristics (e.g., bidirectional, vestibular “bias”, linear, etc.), we do not begin by using earlier psychophysical applications of detection theory (e.g., Green and Swets 1966). Instead, while cognizant of these earlier works, we begin *de novo* with basic signal detection theory (Kay 1998).

More specifically, we will present a model underlying “one-interval recognition” tasks and then will highlight comparisons to “one-interval detection”, “two-interval detection”, and “two-interval recognition”. Following are some of the specific questions that will be answered. What are the differences between the models underlying recognition and detection paradigms? What are the differences between the models underlying one-interval versus two-interval paradigms? For all paradigms, a 3-down/1-up staircase would target a percent correct of 79.4%, but what

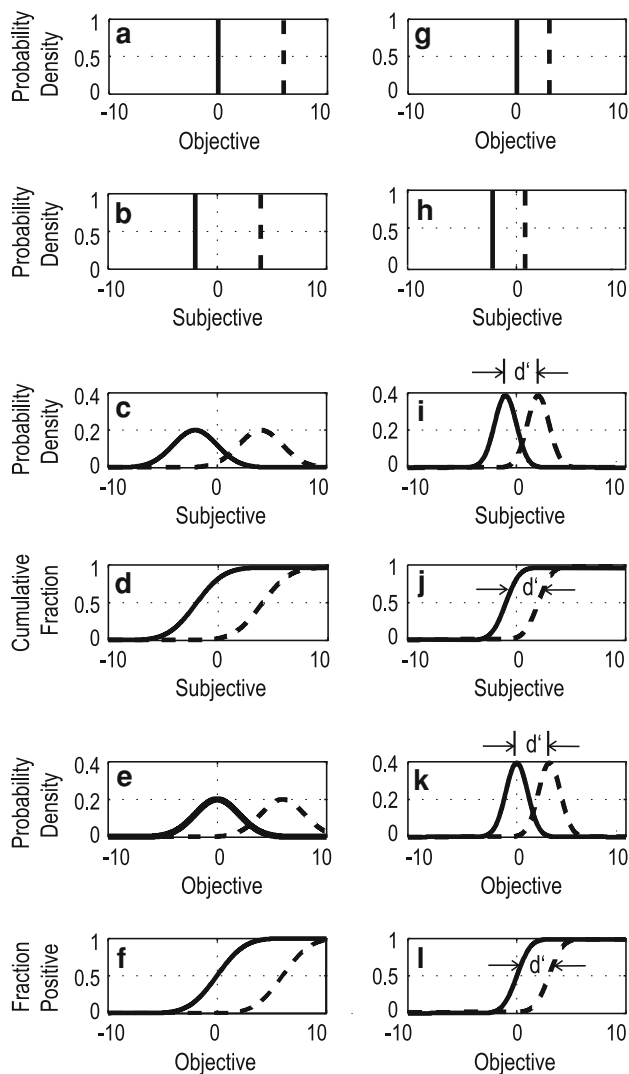


Fig. 1 **a** shows objective stimuli having amplitudes of 0 (the null stimulus) and 6. **b** shows that a vestibular bias of -2 yields the sensed stimuli of -2 and $+4$. **c** A probability density function (PDF) represents that the perceived amplitude will have some variation ($\sigma = 2$) due to noise. **d** A cumulative distribution function (CDF) is calculated as the integral of the PDF. **e**, **f** show the PDF and CDF in objective coordinates. The *right* column (**g** through **l**) represents the same quantities but normalized by 2—the standard deviation from the *left* column. In these normalized units, **g** the objective stimuli are 0 and 3, **h** the vestibular bias is -0.5 yielding sensed stimuli of -0.5 and 2.5, (**i** through **l**) the noise has a standard deviation of 1

does this mean? Since we often are interested in the underlying noise characteristics that determine the threshold, how can we relate experimental thresholds to the standard deviation of the noise?

Methods and background

The physical stimuli measured by the vestibular system are bidirectional. Therefore, vestibular responses are

bidirectional. This characteristic fundamentally influences the application of detection theory to vestibular responses. Specifically, subjects can rotate to the right or left, translate up or down, or tilt forward or backward and sense these different directions of motion. In contrast, photons provided to a subject during a light detection task are unidirectional; perceptions of light opposite to those evoked by photons do not exist as common experience includes nothing “on the other side” of complete darkness. Similarly, the standard hearing test, which is a common clinical application of detection theory, is unidirectional.

Because large differences exist between psychophysical functions for unidirectional and bidirectional stimuli, a brief comparison is warranted. The log of the stimulus amplitude is typically used for unidirectional stimuli; the log is not typically used for bidirectional stimuli because the log of a negative number is imaginary. The theoretical psychophysical function for detection (Yes/No) of unidirectional stimuli tasks ranges between 0% yes for very small magnitudes and 100% yes for large magnitudes. In comparison, because standard detection paradigms for bidirectional stimuli require that all stimuli be either all positive or all negative, the theoretical psychophysical function for detection (Yes/No) of bidirectional stimuli ranges between 50% yes for very small magnitudes and 100% yes for large magnitudes. Furthermore, both unidirectional and bidirectional cumulative distribution functions have two free parameters, but very different, and even somewhat contradictory, terminologies are used. For a direction-recognition task using bidirectional stimuli, we refer to “bias” (e.g., vestibular bias) as the stimulus level that yields the percentage correct midway between the lower and upper bounds of the psychometric function, and we used “threshold”, which is linearly proportional to the standard deviation of the noise (c.f. Table 1), to refer to the width of the transition (e.g., the standard deviation of a Gaussian probability density function underlying the psychometric function). For unidirectional stimuli, the term “threshold” replaces bidirectional “bias”, and “slope” replaces bidirectional “threshold.”

Note that we are not claiming that vestibular responses are the only sensations that are bidirectional as there are bidirectional aspects of other modalities. We are simply noting that vestibular responses are bidirectional and that this fundamental characteristic impacts the application of detection theory. Specifically, vestibular bias—often simply defined as an offset from zero—arises, at least in part, due to the bidirectional nature of vestibular responses. For example, unequal contributions from the left and right labyrinths can lead to vestibular bias. A similar vestibular bias can arise centrally from asymmetric processing of peripheral information. The cause/source of a vestibular bias—whether peripheral or central—is not crucial to the

Table 1 Threshold comparison for different paradigms

	$k_{79.4\%} = T_{79.4\%}/\sigma$	$k_{70.7\%} = T_{70.7\%}/\sigma$	Theoretical threshold relationships	% Correct at $T_{1\sigma}$
One-interval recognition	0.82 icdf('norm',0.794,0,1)	0.54 icdf('norm',0.707,0,1)	T_{1r}	84.1 1- cdf('norm',0,1,1) or cdf('norm',0,-1,1)
Two-interval recognition	0.58 icdf('norm',0.794,0,sqrt(2))/2	0.39 icdf('norm',0.707,0,sqrt(2))/2	$T_{2r} = T_{1r}/\sqrt{2}$	92.1 1-cdf('norm',0,2,sqrt(2))
One-interval detection	1.64 icdf('norm',0.794,0,1)*2	1.09 icdf('norm',0.707,0,1)*2	$T_{1d} = 2T_{1r}$	69.1 1-cdf('norm',.5,1,1)
Two-interval detection	1.16 icdf('norm',0.794,0,sqrt(2))	0.77 icdf('norm',0.707,0,sqrt(2))	$T_{2d} = T_{1d}/\sqrt{2} = \sqrt{2}T_{1r}$	76.0 1-cdf('norm',0,1,sqrt(2))

Table summarizes the relationship between thresholds (T) and the standard deviation of the physiologic noise (σ) and also shows theoretic calculations for each of four testing paradigms. $T_{79.4\%}$ is the 79.4% threshold for a 3-down/1-up staircase paradigm, $T_{70.7\%}$ is the 70.7% threshold for a 2-down/1-up staircase paradigm, $k_{79.4\%}$ is the different constant multiplier for each of the four 3-down/1-up testing paradigms, $k_{70.7\%}$ is the different constant multiplier for each of the four 2-down/1-up testing paradigms. $T_{1\sigma}$ is the “one-sigma threshold”, which is the stimulus magnitude that equals the standard deviation of the noise ($T_{1\sigma} = \sigma$). Example calculations assumed that vestibular bias and criterion bias were zero; both biases need to be considered when non-zero. While explicit values are shown only for 70.7% (2-down/1-up) and 79.4% (3-down/1-up) correct, the theoretical relationships (3rd column) generalize to other threshold criteria (e.g., 75% correct)

following analysis as either can yield vestibular bias. Unfortunately, as described in more detail below, “bias” is used in the detection theory literature to refer to the fact that the detection criteria might not be the same for all subjects, which forms a basis for “criterion bias”. Criterion bias will be introduced mathematically later, but, briefly stated criterion bias simply represents the tendency for a subject to prefer one choice over another. To help distinguish these two independent effects, we will generally avoid the use of “bias” by itself and will instead refer specifically to “vestibular bias” or “criterion bias”.

We assume that any vestibular bias is constant (e.g., independent of stimulus amplitude and duration). The presence of vestibular bias means that we must distinguish the actual (“objective”) stimuli—known to the experimenter—from the sensed (“subjective”) stimuli—experienced by the subject. To do so, we use an example comparing two stimuli with specific values. Like some other examples, the values are arbitrary. (When values are not arbitrary, we will specifically state this.) Assume objective stimulus amplitudes of 0 and 6°/s and a subjective vestibular bias (μ) of -2°/s, where the vestibular bias simply means that, in the presence of null stimuli (zero amplitude), this subject will, on average, subjectively experience stimuli equivalent to -2°/s. (For example, the presence of a vestibular bias might manifest as a positive VOR bias, though we do not assume that the VOR bias and the subjective bias are necessarily one and the same.)

The probability density functions for the objective stimuli are shown as impulse functions (Fig. 1a), since the objective stimuli are presumed to have much less variability (“noise”) than the subjective experience (Fig. 1c). Specifically, it is presumed that the motion devices are well controlled and provide nearly the same stimuli each time.

The vestibular bias is represented by a rightward shift of the subjective axes relative to the objective axes (Fig. 1b). Therefore, the mean subjective motions sensed are -2 and +4°/s, respectively. But the sensed stimuli will have physiologic noise that is assumed Gaussian, with a standard deviation (σ) of 2°/s chosen for this example. This noise includes all physiologic sources of variability (afferent noise, processing noise, etc.).

For all analyses included herein, the noise for small near-threshold stimuli is assumed constant and is assumed to sum with the signal. The distributions in Fig. 1c can be interpreted as indicating that for a given stimulus amplitude the subjective sensation (sensed signal) for a given trial will be randomly selected from this probability distribution. A sensed signal near the mean is most likely, but individual trials can yield sensed signals above or below the mean, with the prevalence proportional to the magnitude of the probability density function (PDF). The equation for a Gaussian PDF can be written as:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

The cumulative distribution functions (CDFs) for these PDFs are shown Fig. 1d. The CDFs represent the percentage of times that the subject’s perception would be less than the value on the abscissa (x -axis) for the given mean stimulus. (An example follows three paragraphs below.) The CDF is the integral of the PDF:

$$\phi(x) = \int_{-\infty}^x f(x')dx' = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x'-\mu)^2}{2\sigma^2}} dx' \tag{2}$$

This integral does not have a closed form solution, so it is solved using standard numerical methods that often

involve a special function called the error function (Hildebrand 1976; Wikipedia 2010b). Figure 1e and f show the PDFs and CDFs in objective coordinates, which is accomplished by simply reversing the shift due to vestibular bias from objective (Fig. 1a) to subjective coordinates (Fig. 1b).

The second column of Fig. 1 shows the same variables but the units have been changed. Specifically, we have normalized all values by the standard deviation of the original distribution. This process of normalizing by the standard deviation is sometimes called “standardizing” the variable as this normalization yields one as the standard deviation. For the rest of the paper, we will only use distributions with a standard deviation of one (implicitly assuming standardization). For simplicity, we further assume that this standard deviation of the noise is always constant and does not depend upon the stimulus but this assumption (like others) can be relaxed if required by experimental findings. For this distribution, this normalization yields objective stimuli of 0 and +3 and subjective stimuli means of -1 and $+2$. This normalization is equivalent to changing units and does not limit the generality of any findings.

The CDFs of Fig. 1j represent the percentage of times that the subject’s perception would be less than the value on the abscissa (x -axis) for the given mean stimulus. For example, for the dashed distribution, the subjective mean is $+2$, so 50% of the trials would be perceived as less than $+2$ (and 50% greater than $+2$), and 2.28% of the trials involving a mean subjective stimulus of $+2$ (objective stimulus of $+3$) would be perceived as being negative.

This value of 2.28% can be calculated using the cumulative distribution function in MATLAB Statistics Toolbox as `cdf('norm',0,2,1)`, where ‘norm’ indicates that the distribution is normal (Gaussian), zero represents the “decision boundary”, one is the standard deviation in standard deviation units, and two is the mean of the subjective distribution. Placing the decision boundary at zero represents that we asked subjects to indicate whether their perception was positive or negative. (Decision boundaries will be discussed in more detail later.) We will show similar MATLAB functions in the text to help directly illustrate the calculations. These can easily be mapped to any other program (e.g., Excel, etc.).

The Gaussian assumption is justified by the central limit theorem of statistics (Larsen and Marx 1986; Wikipedia 2010a). More fundamentally, it is not essential that the distributions be Gaussian, though for bidirectional vestibular responses, the noise distribution will typically be symmetric (or at least nearly symmetric). If the distributions are not Gaussian, the approach outlined here is still valid, though calculations would need to be redone using an appropriate distribution. (For example, all of the standard z -scores and d' calculations—to be discussed in the

following paragraphs—assume Gaussian noise.) We will take a few paragraphs below to introduce some standard signal detection metrics but readers seeking details should refer to another source, like Macmillan and Creelman (2005).

One-interval versus two-interval designs

Stimuli can be presented in different temporal order (“intervals”) or different spatial locations, which yield an n -alternative forced choice paradigm. Vestibular stimuli cannot be applied in different spatial locations. Furthermore, an early study (Blackwell 1952) concluded that forced choice procedures involving temporal intervals were preferred over providing stimuli in alternate spatial locations. This, alongside the fact that most hearing studies lend themselves more to sequential application of the stimuli, helps explain that the two most common experimental designs used in psychophysical detection theory paradigms are one-interval and two-interval designs (e.g., Macmillan and Creelman 2005). In a one-interval design, a single stimulus is provided, which the subject must classify. For example, for one-interval detection, motions having different amplitudes will be provided. For each trial, the subject must report whether they perceived motion or not. In a one-interval recognition task, positive or negative motion will be provided, and, for each trial, the subject must report whether they perceived motion in the positive or negative direction.

In a two-interval design, both alternatives are provided on every trial in random order and the subject must report the order of the stimuli (i.e., which came first). For example, in a two-interval detection task, motion is provided in one interval and no motion in the other interval, with the order randomized. The subject will be asked to report which interval included the motion (or which interval included the null stimulus). In a two-interval positive/negative recognition task, motion in the positive direction would be provided in one interval and motion in the negative direction in the other interval; again the order would be randomized and the subject would report the direction of the motion in the first (or second) interval.

Adaptive versus non-adaptive methods

Various techniques are used to select the stimulation amplitudes so that the threshold can be accurately estimated with a limited number of trials. There are two classes of experimental methods—adaptive and non-adaptive—that are used to define thresholds and/or apply detection theory to psychophysical data. For the non-adaptive approach, the subject’s responses do not affect the stimuli presentation. In one common non-adaptive method, sometimes called the method of constant stimuli, the

investigator decides both the amplitude and presentation order in advance. The data are then fit to determine a psychometric function (e.g., Wichmann and Hill 2001a, b). The non-adaptive approach is often less efficient than adaptive procedures (i.e., more trials are required); this can lead to more lapses, which are simply defined as stimulus-independent errors (e.g., inattention, fatigue, sleep, etc.) that can introduce significant biases into parameter estimates (Wichmann and Hill 2001a)—especially when such a lapse occurs at a large stimulus level that a subject should almost always identify correctly.

For the adaptive approach, the stimuli provided are determined by the subject's responses on previous trials. Two common adaptive methods include staircase and maximum likelihood paradigms (Leek 2001). The basic characteristic of staircase methods when applied to threshold measurements is that they increase the stimulus magnitude when the subject makes a mistake and lower the magnitude when the subject is correct. One common staircase procedure is called an *n*-down/1-up procedure. For an *n*-down/1-up staircase, the stimulus level decreases when the subject correctly discriminates the stimuli *n*-times in a row and increases for each incorrect response. For such a paradigm, the stimulus level varies above and below an asymptote that falls at the stimulus level at which increases and decreases in the stimulus are equally likely. At the staircase asymptote, the probability of providing a wrong answer before *n* correct answers are provided equals the probability (*P*) of providing *n* correct answers in a row ($P^n = 0.5$). As a specific example, for a 3-down/1-up (3D/1U) paradigm, the level of the stimulus decreases when a subject is correct three trials in a row and increases for each mistake. Therefore, the chance of a single correct answer is calculated as the cube root of 0.5, which is $P = 0.794$. Therefore, a 3D/1U paradigm targets a subject performance level of 79.4% correct (Leek 2001). Advantages of this paradigm are that it is simple and doesn't require a priori knowledge of the response distribution. Disadvantages are that the stimulus level depends only upon recent trials. In other words, not all data are utilized to determine the next stimulus level.

Maximal likelihood models provide the basis for a second adaptive method that combines some of the advantages of fixed-interval and staircase procedures. When enough data have been obtained to yield an acceptable fit, a cumulative distribution (e.g., cumulative Gaussian) is fit to the data, and the target stimulus level is determined from this fitted model. Additional data are collected to improve the quality of the fit until the fit quality is acceptable. Specifically, this process of fitting a psychometric function and selecting a target stimulus from the fitted model is iteratively repeated until the desired endpoint set by the investigator (e.g., fixed # of trials,

variance of parameters below that specified, etc.) is reached (Leek 2001). Maximal likelihood methods are a little more difficult to implement than staircase or fixed-interval methods. But using maximum likelihood estimation, methods to guide stimuli selection has been shown to be more efficient—requiring fewer trials to converge to a predetermined confidence interval—than staircase methods (e.g., Pentland 1980; Watson and Pelli 1983; Leek 2001).

Standard detection theory metrics

Having introduced the basic distributions underlying detection theory, we will now summarize some standard detection theory techniques that utilize these distributions. We do so because this will introduce some tools that we will use later and because an understanding of these metrics will guide an understanding of the main analytical results.

One standard analysis approach—sometimes even considered a “gold standard”—is to fit a cumulative distribution function like those shown in Fig. 1f to the data. The data set consists of many trials each of which yields one subjective decision (e.g., left or right?). This is often done using algorithms that yield maximum likelihood fits to the data, like those described in detail in Wichmann and Hill (2001a, b). Such curve fits will typically yield estimates $\hat{\sigma}$ and $\hat{\mu}$ that represent maximum likelihood estimates of the noise standard deviation and bias, respectively.

Because such curve fits utilize all of the available trials and yield a maximum likelihood fit to the data, such fits often provide the best estimate of the underlying distributions. Such curve fits will generally asymptotically converge to the actual underlying distribution function if enough data are available. Typically, at least 100–200 trials are required for a high-quality fit.

Given this, what is a threshold as defined by detection theory? A threshold is the stimulus level at which a subject is able to detect or recognize the stimulus in some appropriate fraction of the trials set by the investigator. For a recognition task, this will fall some fraction above/below the 50% level, where 50% represents pure guessing. (This 50% level is sometimes referred to as the “point of subjective equality” (PSE) in the literature. PSE is discussed in more detail below.) As one specific example, for a 3-down/1-up paradigm, the threshold occurs when the fit CDF equals 79.4/20.6% (e.g., left/right or positive/negative).

A differential threshold, which is also called a just noticeable difference (or JND), is the smallest stimulus that can be discriminated from a reference stimulus. The JND, when the reference is no motion, is an absolute threshold that corresponds to the two-interval detection analyses presented herein. By always assuming that the reference is

null motion, our analyses (Figs. 4 through 5) focus exclusively on absolute thresholds but can easily be extended to differential thresholds by choosing an appropriate decision boundary and by replacing the null motion distribution with the appropriate reference motion distribution.

The above methods are viable when enough trials are available to fit the data. In the absence of a complete data set, thresholds or analogous information can be extracted using other standard approaches. A few such approaches are outlined below; such approaches can provide meaningful information but seldom equal the quality of a maximum likelihood fit.

Slope approximation First, imagine that data are only available or have only been acquired near the midpoint (50% correct) with little or no data available for very large magnitude (positive and/or negative) stimuli. One approximation is to realize that the cumulative distribution function is roughly linear near the midpoint (e.g., see Fig. 1d). In fact, because CDFs (Eq. 2) are simply the integral of PDFs (Eq. 1), the slope of the Gaussian CDF at its midpoint (s) equals the peak value of the Gaussian PDF, or $s = \frac{1}{\sqrt{2\pi}\sigma^2}$. Therefore, if one measures the slope (s) near the center, this provides an approximate estimate of σ , $\hat{\sigma} = \frac{1}{\sqrt{2\pi}s}$.

z-transformations and d' Other approaches involve what is called a z -transformation. In the context of threshold estimation, a z -transformation (not to be confused with the z -transform of discrete time dynamic systems) converts data, for example, correct or incorrect detection rates, into a z -score. A z -score $(x - \mu)/\sigma$, is simply a metric that indicates where one falls on the cumulative distribution in standard deviation units. More specifically, the z -score represents distance from the distribution mean in standard deviation units.

We'll use the arbitrary example of Fig. 1 to elaborate, including the introduction of a few ideas that will be more fully defined later. The CDF for the null motion stimulus (Fig. 1j) equals 0.841 at subjective zero (at x equals zero). This means that 84.1% of the trials with no motion will be perceived as negative motion due to the presence of a small vestibular bias. In Matlab, the z -transformation can be calculated as `icdf('norm',0.841,0,1)`, where `icdf` is the inverse cumulative distribution function, 0.841 is the fraction correct for a Gaussian distribution with mean 0 and a standard deviation of 1, which yields a z -score of +1.00. For the +3.0 stimulus, which corresponds to +2 in subjective coordinates, subjects will on average experience negative motion 2.28% of the time [`cdf('norm',0,+2,1)`]. This means that, when exposed to motion stimuli having subjective amplitude of +2, the subject should correctly perceive positive motion 97.72% (100% minus 2.28%)

of trials. The z -transformation of 0.9772 [`icdf('norm',0.9772,0,1)`] is +2; this informs us that, the mean of this distribution is 2 standard deviations to the right of zero. Note that the sum of the z -scores equals 3, which equals the distance between the distribution means. (Shortly, we will define this sum as d' , a standard detection theory metric.)

We also calculate z -scores in the objective coordinates (Fig. 1k and l, expanded in Fig. 2). The CDF for the null motion equals 0.5 at objective zero (at x equals zero). This means that 50% of the trials with no motion will be perceived as negative motion and 50% as positive motion. This yields a z -score of 0 (`icdf('norm',0.5,0,1)`), since the mean of the null motion occurs at the objective origin. The CDF for the motion (Fig. 2b) equals 0.00135 at objective zero [`cdf('norm',0,2,1)`], which means that 0.135% of the motion trials will be perceived as negative motion and 99.865% as positive motion. This yields a z -score of 3, `icdf('norm',0.99865,0,1)`, since the mean of the motion distribution falls at the distance of three from the objective origin. Note that the sum of the z -scores again equals 3.

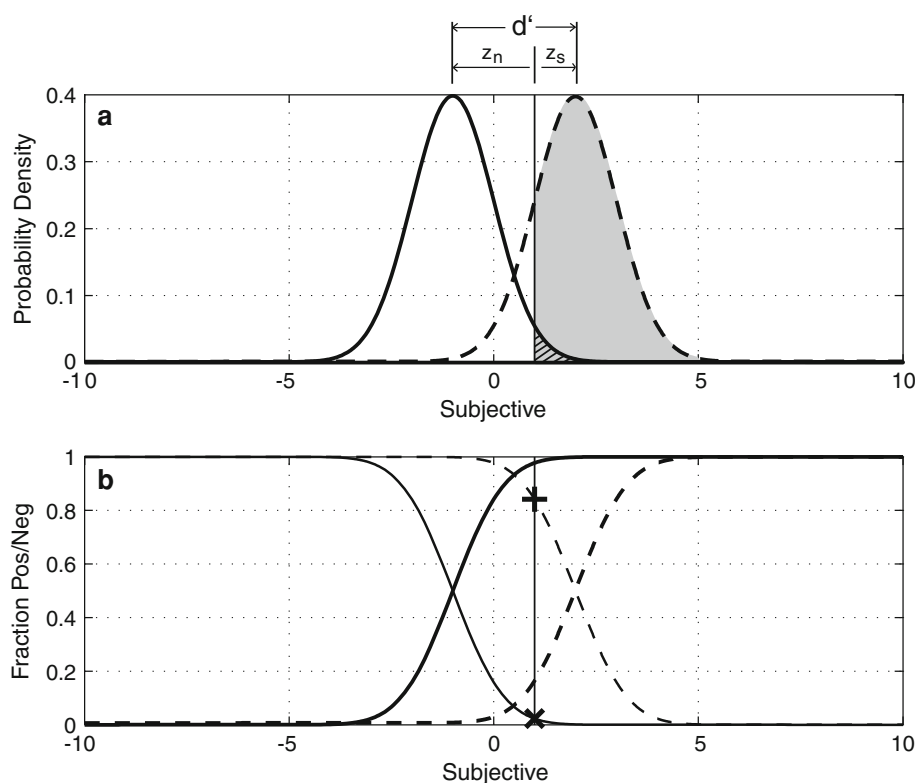
Ignoring vestibular bias when using a z -score can be misleading. For the example of the previous paragraph, the operator knows that the stimulus was 3 and they know that this corresponded to positive responses 97.72% of the time, which yields a z -score of 2. In the absence of bias, we could use the stimulus amplitude (A) and z -score to calculate the standard deviation. Specifically, $z = A/\sigma$ so $\sigma = A/z = 3/2$. This does not match the actual standard deviation, which was 1, because the vestibular bias of -1 was neglected.

In theory, a signal detection parameter called d' (pronounced d prime) will correct for any vestibular or criterion bias present. For a detection task, d' is the distance between the means of the signal distribution and the null motion distribution after normalizing by the standard deviation of the noise distribution. For example, note that the distance calculated in subjective and objective coordinates above equaled 3, independent of the coordinates used.

Figure 2, which simply expands Fig. 1i and j, shows an example calculation of d' in subjective coordinates. We arbitrarily place a decision boundary at +1, which means that this theoretical subject will decide that all trials yielding stimuli sensed to be less than +1 are due to the null stimulus and all trials yielding sensed stimuli greater than +1 are due to motion. (It is straightforward to show other decision boundaries do not change d').

We first define the hit rate, which is the percentage of trials correctly identified as motion. This is represented graphically using solid shading (Fig. 2a) and a plus sign (Fig. 2b). For this example, with a decision boundary at +1 and a distribution with a mean of +2 and standard deviation of 1, the hit rate is 84.13% [`1-cdf('norm',1,2,1)`].

Fig. 2 a shows subjective PDFs for sensed stimuli having mean amplitudes of -2 and $+4$. The *shaded area* represents the hit rate. The *hatched area* represents the false alarm rate. The *thick curves* in **b** show the CDFs for these distributions. The *thin curves* show one minus these CDFs. The $+$ represents the hit rate. The x represents the false alarm rate



For this example, the false alarm rate, which is the percentage of null motion trials incorrectly identified as motion, is 2.275% [$1 - \text{cdf}(\text{'norm'}, 1, -1, 1)$]. The false alarm rate is represented graphically using hatched lines (Fig. 2a) and an x (Fig. 2b).

Recalling that a z -score defines a distance from the mean of a distribution, we can use the hit rate and false alarm rate to define the distance between the distribution means. The z -score for the hit rate of 0.8413 is $+1$, which indicates that the mean of the motion stimulus is one standard deviation above the decision boundary, which is indicated on the figure as z_s . The z -score for the false alarm rate of 0.02275 is -2 , which indicates that the mean for the null stimulus is two standard deviations below the decision boundary, which is represented by z_n . d' is simply the distance between the two means, which is simply the difference of z_s minus z_n . so d' can be calculated as the z -score for the hit rate (calculated as $+1$ above) minus the z -score for the false alarm rate (calculated as -2 above), which can be written as: $d' = z_s - z_n = z(\text{hit rate}) - z(\text{false alarm rate})$. So for this example, d' is again correctly estimated as 3.0.

Point of subjective equality The point of subjective equality (PSE) is another psychophysical detection theory metric that is commonly evaluated for studies that use the method of constant stimuli, which, according to (Jones 1974), dates back at least to Fechner. The method of constant stimuli is used to measure subjects' ability to

discriminate between a standard stimulus and comparison stimuli (analogous to the two-interval vestibular task described earlier). In this context, the PSE is defined as the test stimulus that is perceived to be the same as the reference, but this definition has been generalized to other applications.

A PSE example will be presented later (Fig. 6), where we will show that the PSE for a two-interval direction-recognition task occurs, when the fraction positive is 50%. This occurs when the stimulus is the same as the reference. This, of course, is the expected theoretical result since a vestibular test stimulus should theoretically be perceived the same as the reference when it does not differ from the reference.

PSE is often used to compare two or more sensory modalities. The two-interval direction-recognition analyses presented later (Fig. 6) can easily be extended for such cross-modality investigations. For example, the reference distribution might be vestibular and the test stimulus visual, or the reference might be vestibular and the test stimuli might combine visual and vestibular motion. See Carpenter-Smith et al. (1995) for an example.

Results and analyses

One-interval direction-recognition

For one-interval direction-recognition, the subject reports whether they moved in one direction or the other (left vs.

right, up vs. down, forward vs. back, etc.). For this analysis, we initially assume that the task is symmetric. For example, the subject might be discriminating yaw rotation, head roll tilt, or roll tilt of the subjective visual vertical—each of which is left/right symmetric. Given a symmetric task, each subject would typically set the decision boundary at subjective zero, which is represented by the vertical line passing through the subjective origin (Fig. 3b, c). A decision boundary simply represents the rule that the subject will use on a given trial to make their decision. This specific decision boundary indicates that each subject will report positive motion on a specific individual trial if they sense positive motion (signal plus noise) on that trial, and they will report negative motion if they sense negative motion on that trial.

We specifically note that setting a decision boundary at subjective zero does not prohibit vestibular biases. For

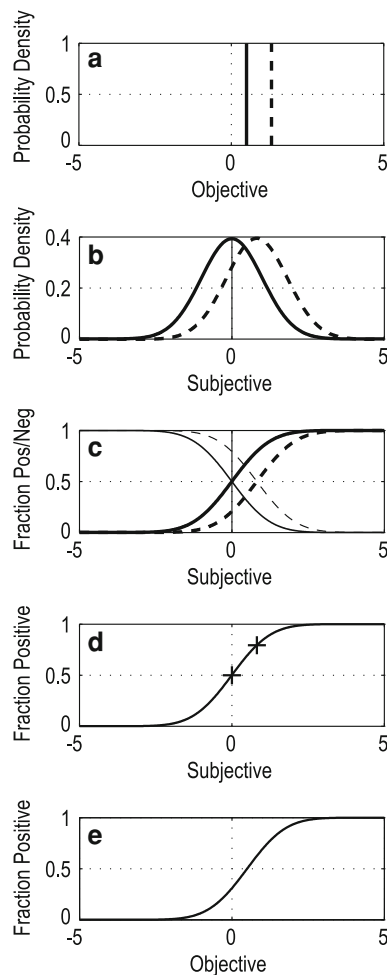


Fig. 3 One-interval recognition **a** Two separate trials having objective amplitudes of +0.5 and +1.32 are shown. **b** and **c** *Thick* curves show the subjective PDF **b** and CDF **c** given a bias of -0.5 and standard deviation of one. *Thin* curves show one minus the CDF. The *vertical line* at the origin indicates the assumed decision boundary. **d** and **e** show the expected subject performance for many different stimulus levels in subjective and objective coordinates, respectively

example, this decision boundary does not indicate that subjects will always report left tilt when they *objectively* tilt to the left—only that the subjects will report left tilt when they *subjectively* perceive left tilt.

We also specifically note that a criterion bias can be present for recognition that would yield a decision boundary at a level other than subjective zero for some vestibular recognition tasks. For example, one can imagine that it is easier and/or more costly to fall backwards than forwards, which *could* lead to a criterion bias even for recognition tasks, especially if falling is possible. One could even provide rewards and punishment meant to encourage more positive responses than negative response for a symmetric task like left/right tilt recognition, but, for simplicity, we assume that no such criterion biases are present for symmetric recognition tasks. Criterion biases are presented in more detail below when we consider one-interval detection tasks, and we also return to this issue in the discussion. In fact, in the context of many vestibular discrimination tasks, criterion bias cannot be directly distinguished from vestibular bias behaviorally.

Two separate trials having objective amplitudes of $+0.5$ and $+1.32$ are shown (Fig. 3a). Figure 3b and c, respectively, show the subjective PDF and CDF, given a bias of -0.5 . The decision boundary is located at zero in the subjective coordinates—corresponding to $+0.5$ in objective coordinates. The value of the bias is arbitrary, but the difference (0.82) between the two trials was specifically chosen for reasons that will become apparent.

The value for the cumulative distribution that represents the positive stimuli having a mean 0.82 in the subjective frame (thick dashed curve) is 0.206 . This means that for subjective stimuli of $+0.82$, the subject will sense that the stimulus is negative on 20.6% of the trials. This, of course, implies that the subject will correctly sense that the stimulus is positive on 79.4% of the trials (thin dashed curve). Also, the value for the cumulative distribution that represents the null stimulus in subjective coordinates is 0.5 at the decision boundary. This indicates that the subject will determine the motion direction was negative on 50% of the trials when the average subjective sensation is zero.

These two data points are plotted in Fig. 3d, where we are now plotting average expected subject performance for many different stimulus levels. We can go through the above process for many different levels of sensed stimuli, each time finding the percent of trials reported as positive for a given stimulus. This would yield the curve shown in Fig. 3d. We have to change our perspective to interpret this figure. In Fig. 3c, the stimuli were fixed, and we were determining the cumulative probability density of subjective experience. In Fig. 3d, we vary the stimulus level and plot average subject performance for different subjective stimuli [$\text{cdf}(\text{'norm'}, \text{stimulus}, 0, 1)$].

Furthermore, remember that, because of vestibular bias, the objective stimuli are shifted by 0.5 relative to the subjective stimuli. The plot of subject performance versus objective stimuli is shown in Fig. 3e [cdf('norm',stimulus,-bias,1)]. Note that the experimenter would determine that the subject got 50% correct at 0.5, which would inform the operator that the subject's vestibular bias was -0.5 .

Why did we choose the stimulus levels shown in Fig. 3? The motion of 0.5 was chosen because, given a subjective vestibular bias of -0.5 , it leads to a subjective distribution having zero mean. In other words, the negative motion counteracted the subjective vestibular bias leading to a subjective perceptual state of no motion. The objective stimulus of 1.32 yielded a subjective distribution with a mean of 0.82. This subjective distribution is the same as would be experienced with zero vestibular bias and 0.82 stimulus amplitude. As shown, the subject should, on average, correctly determine that this stimulus is positive 79.4% of the time. This number might seem familiar because we earlier showed that the percent correct targeted by a 3-down/1-up staircase is 79.4%. Therefore, we chose the subjective distribution to have a mean of 0.82 to show that, in the absence of a significant vestibular bias, the 79.4% threshold for a 3-down/1-up staircase occurs for a subjective stimulus level of 0.82σ . As we will show, this

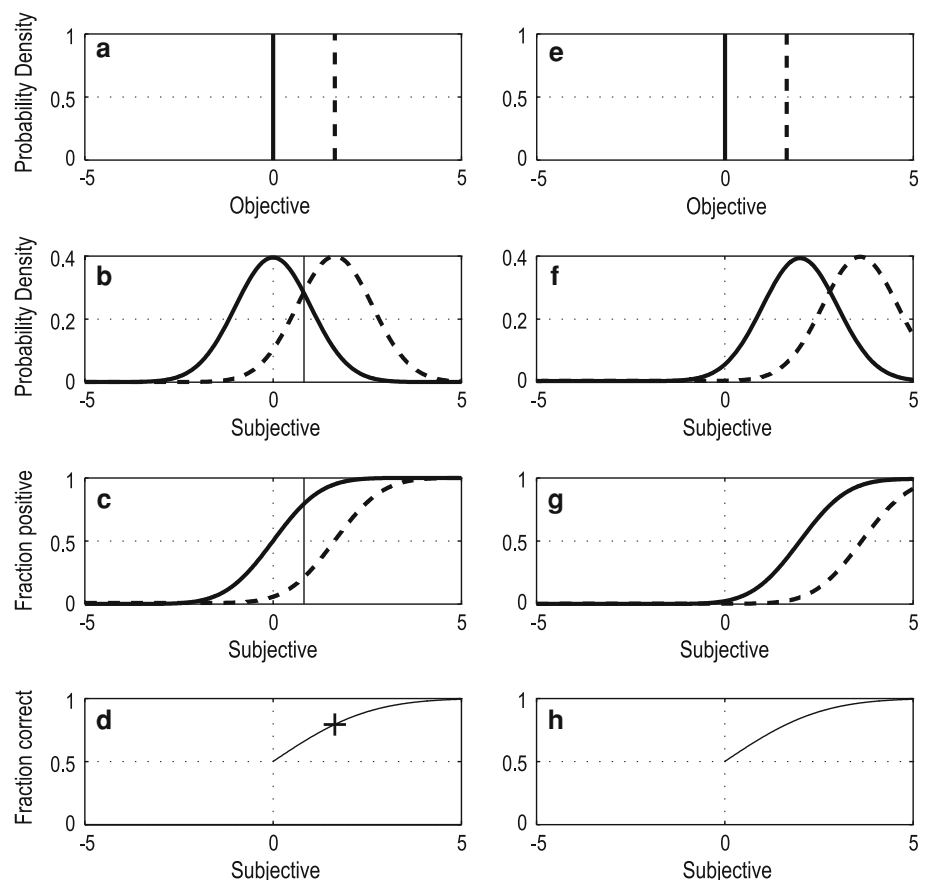
relationship holds only for one-interval recognition. Different values will be found for the other paradigms.

One-interval detection

For one-interval detection, the subject's task is to determine whether motion is present or not (e.g., yes or no). For simplicity, we will first proceed assuming that the vestibular bias equals zero. Therefore, the comparison is always made to no motion, which is represented by the solid stimulus (Fig. 4a) and solid subjective distribution (Fig. 4b). The actual motion (Fig. 4a, dashed line) will always be in the positive direction, and the subject knows this. (Analysis yields identical results for all negative motion.) Since the mean value at threshold for a 3-down/1-up paradigm targets 79.4% correct, we will show below that, in the absence of any bias (vestibular or criteria bias), this is achieved for a stimulus level of 1.64σ . Therefore, with σ equal 1 for this example, the mean sensed stimulus level is set to 1.64 (Fig. 4b).

Before proceeding, we must briefly discuss criterion bias, which influences all detection processes and is not unique to vestibular detection. Unlike the symmetric recognition example immediately above, reporting no motion and reporting motion are inherently asymmetric.

Fig. 4 One-interval detection **a** Two separate trials having objective amplitudes of 0 and +1.64 are shown. **b** and **c** show the subjective PDF and CDF, respectively, given a bias of zero. The *line* at the origin indicates the assumed decision boundary. **d** shows the theoretical percentage correct expected at each stimulus level. With the mean stimulus equal to 1.64, a subject will correctly detect motion 79.4% of the time, which we define as threshold. This is shown on Fig. 3d with a +. **e**, **f**, **g**, **h** The right column shows the exact same distributions as the left column but with a vestibular bias of +2.0



Specifically, a given subject may be more concerned about making an error by missing a motion that was present than about making an error by detecting motion when no motion was present. This is referred to as bias in the detection literature and is an issue that must be taken seriously whenever one performs one-interval detection.

In order to compare the different test paradigms, we must set a decision boundary. In the absence of vestibular and criterion bias, the decision boundary should be located where the two PDFs equal one another. For this example, where the standard deviations are assumed equal, the point of equality is 0.82. Below this point, the null motion is more probable (more likely); above this point, motion is more likely than no motion. At 0.82, the values for the two cumulative distributions (Fig. 4c) are 79.4 and 20.6%. This means that for stimuli with an amplitude of 1.64, null motion (solid) will be correctly identified 79.4% of the time, and positive motion will be incorrectly identified 20.6% of the time, which, in turn, means that the subject will correctly determine that there was motion 79.4% of the time. In summary, with the mean stimulus equal to 1.64, a subject will correctly detect motion 79.4% of the time, which we define as threshold. This is shown as a+ on Fig. 4d. Therefore, if one defines a threshold using a 3-down/1-up one-interval detection paradigm in the absence of vestibular bias, the standard deviation would be calculated by dividing the threshold (T) by 1.64 ($\sigma = T/1.64$). We can repeat the above process for many different levels of sensed stimuli—each time finding the percent correct for a given stimulus—yielding the curve shown in Fig. 4d.

Note that the optimal decision boundary is determined by the magnitude of the stimulus. If the stimulus were twice as big, the optimal decision boundary would move to the right. So the optimal decision boundary varies with stimulus magnitude, presumably based on the subject's recent experience. This characteristic makes one-interval detection somewhat less reliable than one-interval recognition, since recognition tasks typically deliver a non-varying decision boundary.

What if there is a vestibular bias? Figure 4e–h show the same distributions as Fig. 4a–d but with a large positive vestibular bias. Note that even for null stimulus (solid), the subject almost always senses positive motion even when no motion is present. Where should the subject set the decision boundary? For a system designed by engineers without any criterion bias, the decision boundary should be located where the two probability density functions equal one another, but this would mean that—during the null motion condition—no motion would be reported almost 50% of the time that positive motion was sensed. While this solution will work for an engineered system, it seems possible that substantial training will be required to train a subject to report no motion when they clearly sense

positive motion. While this was, admittedly, an extreme example—with the bias chosen to be twice the standard deviation of the noise to emphasize the point, smaller biases have the same qualitative effect. Additional theoretical analysis of the vestibular detection paradigm is warranted but this negative characteristic makes one-interval detection less appealing than both two-interval detection and one-interval direction-recognition unless a specific justification for pursuing one-interval detection is identified.

Furthermore, the vestibular bias and criterion bias discussed earlier will each contribute to subjective decisions during a one-interval detection task. There are no simple analyses that will separate vestibular bias from criterion bias, though one could assay vestibular bias using a different test paradigm. Finally, it is also worth briefly noting here that the discussion will also point out that vibration cues can have a substantial influence on one-interval motion detection paradigms, which is just one more factor weighing against the use of one-interval detection.

Two-interval detection

Two-interval detection addresses some concerns associated with one-interval detection, so we discuss this paradigm next. As mentioned earlier, in a two-interval detection paradigm, motion is provided in one interval and no motion in a second interval with the order randomized. The subject's task is to identify whether the motion occurs in the first or second interval. The motion trials for a given test session will always be positive (or always negative), and the subject knows this.

There are different approaches to analyzing two-interval detection tasks. We begin with one positive objective stimulus and one null stimulus (Fig. 5a). The order that these stimuli are provided is important but is not represented in the figure. This paradigm was specifically designed to eliminate criterion bias, since there is no a priori reason that subjects should prefer the first interval over the second (or vice versa), and experimental results confirm that this approach often successfully eliminates criterion bias. Furthermore, as shown below, the effect of vestibular bias is eliminated by the two-interval detection task.

Since the mean value at threshold for a 3-down/1-up paradigm targets 79.4% correct, we will show below that this is achieved for a stimulus level of 1.16σ . Therefore, with σ equal 1 for this example, the mean positive stimulus level is chosen to equal 1.16 (Fig. 5a). As before, choice of this amplitude is illustrative, but any arbitrary value could have been chosen. The subjective PDFs for these two stimuli—given a bias of -0.5 —are shown in Fig. 5b. One way that the subject can compare these data is to subtract

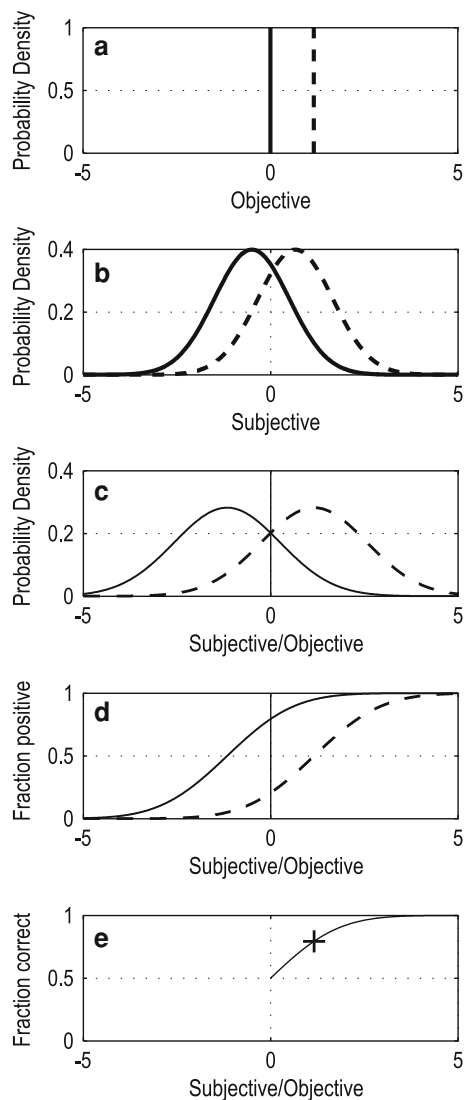


Fig. 5 Two-interval detection **a** The null stimulus and an objective stimulus having an amplitude of 1.16 are shown. **b** The subjective PDFs are shown with a vestibular bias of -0.5 . **c** shows the probability distributions after subtracting the PDF experienced second from that experienced first. The *dashed* curve represents the distribution when the positive motion occurs first, while the *solid* curve represents the distribution when the null motion occurs first. Since vestibular bias is eliminated via subtraction, the magnitude of the mean value of the distributions (peak of the distributions) is 1.16. The variances of the two distributions add, so the standard deviations now equal $\sqrt{2}$. **d** shows the CDFs associated with the above PDFs. **e** shows the theoretical percentage correct at each stimulus level. The + represents 79.4% correct for a stimulus of 1.16

their sensation of the second trial from their sensation of the first trial. (Note that it is not necessary to assume that the subjects are actually performing such a subtraction. In fact, the same result that we derive below is derived by Macmillan and Creelman (2005) using 2-dimensional probability distributions without explicitly performing subtraction.)

The probability distributions following such subtraction are represented by the shorter, broader distributions (Fig. 5c), where the dashed curve represents the distribution when the positive motion occurs first (dashed minus solid from Fig. 5b), while the solid curve represents the distribution when the null motion occurs first (solid minus dashed from Fig. 5b). Note that vestibular bias—which is assumed the same for both intervals—is eliminated via the subtraction process. Therefore, the magnitude of the mean value of the distributions (peak of the distributions) is the stimulus level 1.16, for this example, and there is no longer a difference between subjective and objective coordinates following subtraction, which is represented by labeling the x -axis as subjective/objective. The variances of the two distributions add, so the standard deviations now equal $\sqrt{2}$, which is seen as the broader distributions.

As discussed earlier, there is no a priori reason to prefer the first or second interval, so the decision boundary belongs at zero as shown by the vertical lines in Fig. 5c and d. Using simple logic, if the difference between the two trials is positive, it is reasonable to decide that the motion in the first interval was greater than the second, so the first must have been the motion trial. If the difference is negative, it is reasonable to decide that the second interval was greater than the first, so the second must have been the motion trial.

The cumulative distributions for the probability density function difference are also shown (Fig. 5d). The dashed curve represents when the motion interval was first. At zero, the value of this CDF is 20.6%. This means that the subject should on average incorrectly determine that the 2nd interval included the motion 20.6% of the time, which means that they would be correct 79.4% of the time. The solid curve represents when the null motion interval was first. At zero, the value of this CDF is 79.4%. This means that the subject should on average correctly determine that the 2nd interval included the motion 79.4% of the time. In summary, with the mean stimulus equal to 1.16, a subject will correctly detect motion 79.4% of the time, which is defined as a 3-down/1-up threshold. Therefore, if one runs a 3-down/1-up paradigm in the absence of vestibular bias and finds a threshold (T), the standard deviation would be calculated by dividing the threshold (T) by 1.16 ($\sigma = T/1.16$). This point is indicated using a + on Fig. 5e. As above, we can go through the above process for many different levels of sensed stimuli, each time finding the percent correct for a given stimulus. This would yield the curve shown in Fig. 5e.

Two-interval direction-recognition

The analysis of the two-interval recognition task is similar to the two-interval detection analysis. For this task, the subject experiences both positive and negative stimuli—

one immediately after the other. Therefore, we begin with two stimuli that have the same magnitude but opposite signs (Fig. 6a). With a bias of -0.5 , this yields the two distributions shown in Fig. 6b. Since the mean value at threshold for a 3-down/1-up paradigm targets 79.4% correct, we will show below that this is achieved for a stimulus level of 0.58σ . Therefore, with σ equal 1 for this example,

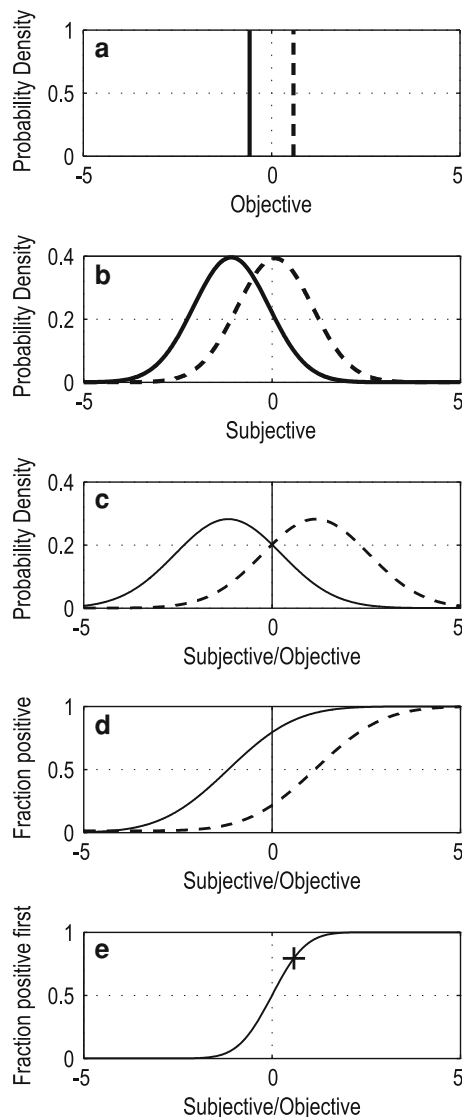


Fig. 6 Two-interval recognition **a** Objective stimuli having amplitudes of -0.58 and $+0.58$ are shown. **b** The subjective PDFs are shown with a vestibular bias of -0.5 . **c** shows the probability distributions after subtracting the PDF experienced second from that experienced first. The *dashed* curve represents the distribution when the positive motion occurs, while the *solid* curve represents the distribution when the null motion occurs first. The magnitude of the mean value of the two distributions is 1.16. The variances of the two distributions add, so the standard deviations now equal $\sqrt{2}$. **d** shows the CDFs associated with the above PDFs. **e** shows the theoretical percentage correct at each stimulus level. The + represents 79.4% correct for the $-0.58/+0.58$ stimulus pair

the mean sensed stimulus magnitude is set to equal 0.58 for illustrative purposes. As above, by subtracting the second sensation from the first sensation and vice versa, we find two broader distributions (Fig. 6c). Just as for the two-interval detection, the effect of vestibular bias is subtracted out. The mean difference is twice the magnitude of the mean stimuli, but the variances of the two distributions add, so the standard deviation increases to equal $\sqrt{2}$. For the same reasons as for the two-interval detection task, we place the decision boundary at zero.

The CDFs are also shown (Fig. 6d). The dashed curves represent when the motion interval was first. At zero, the value of this CDF is 20.6%. This means that the subject should on average incorrectly determine that the 2nd interval included the motion 20.6% of the time, which means that they would be correct 79.4% of the time. The solid curves represent when the null motion interval was first. At zero, the value of this CDF is 79.4%. This means that the subject should on average correctly determine that the 2nd interval included the motion 79.4% of the time. This is shown as the + on Fig. 6e.

As above, we can go through the above process for many different levels of sensed stimuli, each time finding the percent correct for a given stimulus level. This would yield the curve shown in Fig. 6e. In summary, with the mean stimulus equal to 0.58 , a subject will correctly discriminate motion 79.4% of the time, which we define as threshold. Therefore, if one runs a 3-down/1-up paradigm and finds a threshold (T), the standard deviation would be calculated by dividing the threshold (T) by 0.58 $\sigma = T/0.58$.

Comparison across different tasks

Table 1 summarizes theoretical threshold predictions for the different conditions as derived in the previous sections (Figs. 3 through 6). Note that the values for the one-interval detection task assume no vestibular bias and that the decision boundary is set to the stimulus level where the probability density functions equal one another. This is done to allow a comparison to the other tasks, but these assumptions would need to be validated for each given study. The table shows that detection thresholds are expected theoretically to be two times recognition thresholds and that two-interval thresholds are proportional to one-interval thresholds divided by the $\sqrt{2}$. It is worth noting that deviations from such theoretical predictions have sometimes been observed. Nonetheless, given the underlying assumptions, such deviations do not provide reason to reject detection theory without thoughtful deliberations. See Wickelgren (1968) and MacMillan and Creelman (2005) for detailed discussions of this point.

We also utilize the exact same approach outlined above to yield equivalent threshold results for a 2-down/1-up paradigm, which targets a 70.7% threshold. Findings are shown in Table 1.

While not derived above, another way to compare these methods is to set the stimulus magnitude (s) equal to the standard deviation of the noise (σ). If one follows the exact approach outlined above but with the normalized stimulus equal to one, this yields the following theoretically predicted percent correct: 84.1% for a one-interval recognition task, 69.1% for one-interval detection, 76.0% for two-interval detection, and 92.1% for two-interval recognition. See Table 1 for details regarding these calculations.

Discussion

Both detection and recognition are standard discrimination procedures (Macmillan and Creelman 2005) based on the statistical signal processing approach provided by detection theory (Kay 1998). The same is true for one-interval and two-interval tasks. There is no fundamental difference in the theory underlying these approaches. To further compare/contrast these procedures, we begin by discussing several observations/facts.

Impacts of criterion bias

Criterion bias potentially impacts all detection tasks and could affect some recognition tasks. One reason for the difference in impact is that detection is inherently asymmetric—“yes I am moving” is not symmetric to “no, I am not moving.” In comparison, many recognition tasks are symmetric. For example, leftward tilts are symmetric with rightward tilts and rightward yaw rotations are symmetric with leftward yaw rotations. When such response symmetries are present, as they often are for recognition tasks, undesirable criterion bias effects are much less likely as it is improbable that trained subjects will report leftward tilt when they, in fact, perceive rightward tilt, and vice versa. The same cannot be said for yes/no detection tasks, where subjects can be more afraid to miss a motion than to report motion when none was present or vice versa.

Two-interval reduces criterion bias

Because there is often no inherent reason that a subject should prefer interval 1 to interval 2 in a two-interval task, two-interval tasks often help reduce the influence of criterion bias (Macmillan and Creelman 2005). However, such two-interval detection paradigms deserve to be experimentally validated for vestibular tasks. In fact, vestibular dynamics could complicate use of a two-interval

paradigm, especially when there is little delay between the two intervals, since the stimuli in interval 1 could have lingering dynamic influences on what is sensed in interval 2. Furthermore, it is worth noting that timing/order effects have been reported in non-vestibular domains (Macmillan and Creelman 2005) and may be present for vestibular tasks as well.

Impacts of vestibular bias

The influence of vestibular bias on a detection task was demonstrated graphically (Fig. 4e–h). It was shown that for a large vestibular bias, an optimal detector should report no motion even when a large motion was sensed. While this may be simple to implement for a system designed by engineers, it seems possible that it may not be straightforward to train a human to report no motion when motion is clearly and unambiguously perceived.

While not a direct result of this analysis, it is also important to point out that vestibular biases can impact thresholds estimated using staircase procedures. For example, imagine that a subject has a large bias—like that shown in Fig. 4e–h—and is performing a direction-recognition task. This subject will almost always perceive positive motion—except for very large negative stimuli. Thus, this subject will almost exclusively make mistakes when stimuli are negative, as even tiny positive stimuli will almost always be sensed as positive due to the positive bias. Thus, staircase procedures only provide accurate threshold estimates when biases are small relative to inherent variability. Therefore, staircase procedures should only be used when testing demonstrates that the magnitude of each individual subject’s bias is well below that subject’s physiologic noise level (σ) or well below the measurement variability. Or, alternatively, data from subjects with evidence of a substantive vestibular bias must be discarded. In fact, note that two subjects excluded from an earlier study (Grabherr et al. 2008) had demonstrated and/or had a substantial risk for such a vestibular bias.

Vibration

Vibration cues have been shown to provide an indication that one is moving (e.g., Seidman 2008). Since some amount of vibration is unavoidable for controlled motion stimuli, such vibrations can influence experiments measuring vestibular thresholds. For example, air bearings reduce bearing vibrations immensely, but motor drive systems still introduce other vibrations.

Vibration can be sensed by the specific vestibular modality being tested and/or by a number of other parallel sensory modalities (e.g., touch) that potentially also include the vestibular modalities not being tested (e.g., inter-aural

translation modality when testing yaw rotation modality). Vibrations affect the vestibular modality being tested by adding noise to the inherent physiologic noise. If this vibrational noise is much less than the inherent physiologic noise of the modality being tested, then such direct effects of vibration are negligible. The validity of such an assumption can be evaluated experimentally.

For the remainder of this section, we consider the influence of vibrational cues to parallel sensory modalities. For detection paradigms, the difference between vibration during motion and vibration during the null condition provides a motion cue. For recognition paradigms, the difference between vibrations for motion in the two directions provides a motion cue. More specifically, to the extent that a motion device has the same vibration characteristics for motion in each direction, this cannot provide a directional cue for direction-recognition—even if the vibration is large. This means that for detection tasks subjects can use the entire vibration cue available, while for direction-recognition tasks, subjects can only use the difference between vibrations in one direction versus the other direction. Practically speaking, well-designed motion devices provide symmetric stimuli and have relatively symmetric vibration characteristics. Therefore, while one cannot rule out the influence of vibration on recognition tasks, vibration is a substantially bigger issue for detection tasks than for recognition tasks. This is no less true for two-interval tasks than for one-interval tasks.

The contributions of vibration cues to vestibular detection paradigms are impossible to avoid but vibrations can be measured and analyzed by the experimenter. But even when this is done well, it still leaves the problem that it is difficult to prove a negative—i.e., it is difficult to prove that some subtle difference in the noise is not being sensed and used to help with signal detection. As just one realistic example, a vibration component at a single frequency could provide a readily detectable cue to help subjects detect motion even if all other frequency components of the vibration were the same.

Time required for one- and two-interval trials

An advantage of one-interval tasks over two-interval tasks is that they always take less time. The difference may not be large when trials are short since overhead tasks (e.g., subject decision time, subject positioning, delays between trials, etc.) take up nearly as much time as the actual motion stimuli for short trials. For example, for single-cycle sinusoidal stimuli having a frequency of 1 Hz, the time required for one- and two-interval tasks may be about the same. However, for single-cycle sinusoids at 0.1 Hz and lower, testing time will roughly double (and may

increase even more if a substantial delay between intervals is required due to neural filtering dynamics). Beyond simply increasing the time to acquire the same number of trials, doubling testing time also increases the likelihood of fatigue, which is a confounding factor that can increase the subject lapse rate.

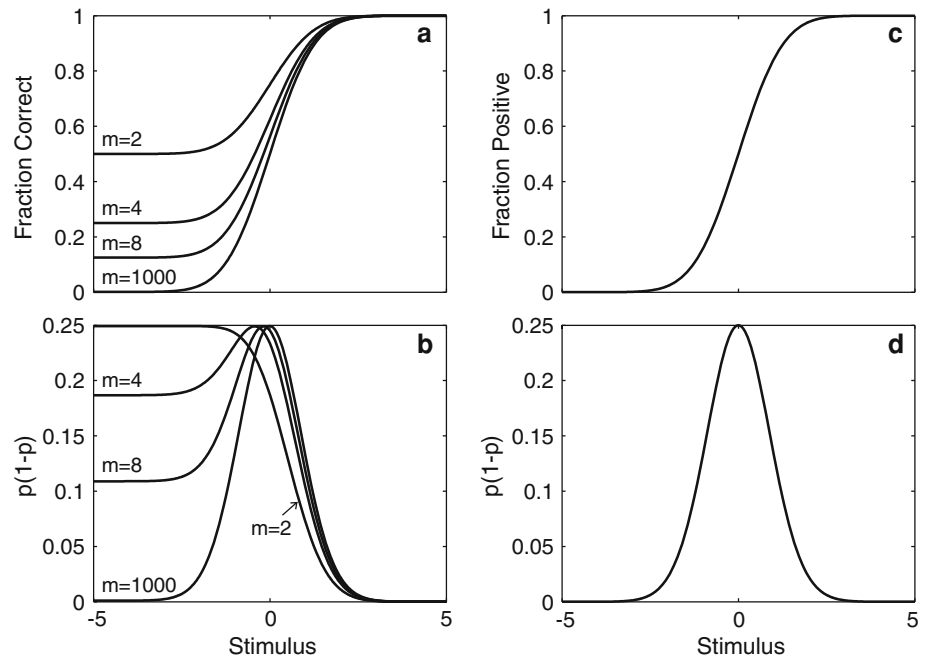
Which paradigm to choose?

There is no simple general answer to this fundamental question. Each specific question(s) and/or hypothesis being investigated must be evaluated while considering the aforementioned results and facts. Nonetheless, some general guidelines can be distilled. First, in general, vibration noise has less of an influence on recognition paradigms than detection paradigms. In fact, in preliminary studies (unpublished) on a Moog 6DOF motion platform, we measured two-interval detection thresholds that were more than two orders of magnitude lower than two-interval recognition thresholds. While others have found smaller differences (Mallery et al. 2010), we are relatively confident that our detection paradigm was assessing a vibration threshold and not a vestibular threshold per se. Second, all else equal, detection paradigms require more trials because null motion trials, which are not required for recognition tasks, must be provided for detection tasks. Third, detection paradigms do not lend themselves to simple decision boundaries because the optimal decision boundary changes with stimulus amplitude (which of course is unknown to the subject).

Two-interval tasks cannot be used to characterize vestibular bias directly. Furthermore, practical considerations (e.g., less time, simpler subjective decision process) lead us to prefer one-interval recognition tasks over two-interval recognition tasks, especially as there are no fundamental advantages of two-interval recognition over one-interval recognition. The above considerations lead us to conclude that a one-interval direction-recognition task will often be a very good choice. Of course, two-interval tasks are essential for tasks having reference stimuli.

Two other procedures can be considered to help improve detection procedures. First, a two-interval task can be used to help minimize the contributions of vibration cues. For example, one could ask subjects to detect whether the first or second interval contained the larger stimuli (Mallery et al. 2010). While such a design could yield a substantial improvement, larger vibrations might accompany a larger motion, which could provide a subjective cue. Second, one can also add artificial high-frequency vibration noise, but then the characteristics of the artificial noise become another design choice that could affect the results. Such artificial vibrations are not bad—but also are not a panacea and deserve cautious consideration, especially as

Fig. 7 Expected response variance. **a** Cumulative distribution function for several m -alternative detection tasks showing the probability of being correct (p). **b** Expected variance for a binomial distribution for N trials, with p the probability of being correct, is $Np(1 - p)$, which is represented by $p(1 - p)$, for the same m -alternative detection tasks. The above plots augment Fig. 1 from a paper by Jakel and Wichman (2006). **c** A cumulative distribution function for a 2-alternative direction-recognition task. **d** Expected variance, again represented by $p(1 - p)$, for the direction-recognition task. Note that the expected variance for the direction-recognition task is always smaller than that for a detection task



thresholds are often used to provide a window to assay physiologic noise.

Finally, an earlier study (Jakel and Wichmann 2006) demonstrated that using more alternatives reduces the expected response variance for a detection task. Using the same approach, Fig. 7a shows Gaussian psychometric cumulative distribution functions (p) versus stimulus, and Fig. 7b shows expected variance, $Np(1 - p)$, for 2-, 4- and 8-alternative detection tasks, where N is the number of trials. For details, see (Jakel and Wichmann 2006). Note that the 4- and 8-alternative detections have substantially less expected variability than the 2-alternative task across a broad range of negative stimulus values. While a 1000-alternative psychophysical task is not feasible, we also show the expected variance for such a task. For comparison, we also show the cumulative distribution function (Fig. 7c) and expected variance (Fig. 7d) for a direction-recognition task.

For a direction-recognition task, neither the CDF (Fig. 7c) nor expected variance (Fig. 7d) change for an m -alternative task. Therefore, m -alternative direction-recognition tasks do not provide greater efficiency than 2-alternative direction-recognition tasks and hence our focus on 2-alternative tasks.

Finally note that as m gets large, the expected variance for m -alternative detection tasks (Fig. 7b) gets smaller and approaches that of a direction-recognition task (Fig. 7d). Therefore, the direction-recognition task has less variability than an m -alternative detection task unless m approaches ∞ (or at least becomes very large). This aligns with the analyses presented earlier herein that independently concluded that vestibular direction-recognition tasks are, in general,

preferable to vestibular detection tasks. Furthermore, while there are disadvantages associated with the bidirectional nature of vestibular response (e.g., vestibular bias), this analysis shows that such disadvantages are accompanied by advantages like smaller estimated variance. Such differences—both advantages and disadvantages—emphasize that analyses applied to unidirectional tasks do not apply to bidirectional tasks. Further analyses and simulations of recognition tasks are warranted.

Acknowledgments Funding for this experiment was provided in part by the NIH (NIDCD R01 04158). We would also like to thank Margaret Lankow for administrative assistance, Csilla Haburcakova and Saori Fukuda for help with the graphics, and Shomesh Chaudhuri, Michael Barnett-Cowan, Luzia Grabherr, Tim Hullar, Faisal Karmali, Koeun Lim, Fred Mast, Chuck Oman, and Florian Soyka, who provided comments on preliminary drafts. Thanks also to Jay Goldberg, who has inspired the use of such theoretical analyses in the vestibular sciences alongside thorough experimental evaluations of such analyses.

References

- Barnett-Cowan M, Harris LR (2009) Perceived timing of vestibular stimulation relative to touch, light and sound. *Exp Brain Res* 198:221–231
- Benson AJ, Spencer MB, Stott JR (1986) Thresholds for the detection of the direction of whole-body, linear movement in the horizontal plane. *Aviat Space Environ Med* 57:1088–1096
- Benson A, Hutt E, Brown S (1989) Thresholds for the perception of whole body angular movement about a vertical axis. *Aviat Space Environ Med* 60:205–213
- Blackwell HR (1952) Studies of psychophysical methods for measuring visual thresholds. *J Opt Soc Am* 42:606–616

- Brown RG, Hwang P (1992) Introduction to random signals and applied Kalman filtering. Wiley, New York
- Carpenter-Smith TR, Futamura RG, Parker DE (1995) Inertial acceleration as a measure of linear vection: an alternative to magnitude estimation. *Percept Psychophys* 57:35–42
- Clark B, Stewart JD (1968) Comparison of three methods to determine thresholds for perception of angular acceleration. *Am J Psychol* 81:207–216
- De Vrijer M, Medendorp WP, Van Gisbergen JA (2008) Shared computational mechanism for tilt compensation accounts for biased verticality percepts in motion and pattern vision. *J Neurophysiol* 99:915–930
- Doty RL (1969) Effect of duration of stimulus presentation on the angular acceleration threshold. *J Exp Psychol* 80:317–321
- Gelb A (1974) Applied optimal estimation. MIT Press, Cambridge
- Grabherr L, Nicoucar K, Mast FW, Merfeld DM (2008) Vestibular thresholds for yaw rotation about an earth-vertical axis as a function of frequency. *Exp Brain Res* 186:677–681
- Green D, Swets J (1966) Signal detection theory and psychophysics. Wiley, New York
- Gu Y, DeAngelis GC, Angelaki DE (2007) A functional link between area MSTd and heading perception based on vestibular signals. *Nat Neurosci* 10:1038–1047
- Guedry F (1974) Psychophysics of vestibular sensation. In: Kornhuber HH (ed) Handbook of sensory physiology, vol VI. Springer-Verlag, New York, pp 1–154
- Hildebrand FB (1976) Advanced calculus for applications. Prentice-Hall, Inc., Englewood Cliffs
- Jakel F, Wichmann FA (2006) Spatial four-alternative forced-choice method is the preferred psychophysical method for naive observers. *J Vis* 6:1307–1322
- Jones FN (1974) History of psychophysics and judgment. In: Carterette EC, Friedman MP (eds) Handbook of perception: vol. 2. Psychophysical judgment and measurement. Academic Press, New York, pp 1–22
- Kay SM (1993) Fundamentals of statistical signal processing: estimation theory. Prentice Hall PTR, Upper Saddle River
- Kay SM (1998) Fundamentals of statistical signal processing: detection theory. Prentice Hall PTR, Upper Saddle River
- Larsen RJ, Marx ML (1986) An introduction to mathematical statistics and its applications. Prentice-Hall, Englewood Cliffs
- Leek MR (2001) Adaptive procedures in psychophysical research. *Percept Psychophys* 63:1279–1292
- Macmillan NA, Creelman CD (2005) Detection theory: a user's guide. Lawrence Erlbaum Associates, Mahwah
- MacNeillage PR, Banks MS, DeAngelis GC, Angelaki DE (2010) Vestibular heading discrimination and sensitivity to linear acceleration in head and world coordinates. *J Neurosci* 30:9084–9094
- Mah RW, Young LR, Steele CR, Schubert ED (1989) Threshold perception of whole-body motion to linear sinusoidal stimulation. In: AIAA flight simulation technologies conference and exhibit, vol AIAA-89-3273, Boston, MA
- Mallery RM, Olomu OU, Uchanski RM, Militchin VA, Hullar TE (2010) Human discrimination of rotational velocities. *Exp Brain Res* 204:11–20
- Ormsby C (1974) Model of human dynamic orientation. In: Aeronautics and Astronautics, vol Ph.D. MIT, Cambridge
- Pentland A (1980) Maximum likelihood estimation: the best PEST. *Percept Psychophys* 28:377–379
- Sadeghi SG, Chacron MJ, Taylor MC, Cullen KE (2007) Neural variability, detection thresholds, and information transmission in the vestibular system. *J Neurosci* 27:771–781
- Seidman SH (2008) Translational motion perception and vestibuloocular responses in the absence of non-inertial cues. *Exp Brain Res* 184:13–29
- Treutwein B (1995) Adaptive psychophysical procedures. *Vis Res* 35:2503–2522
- Watson AB, Pelli DG (1983) QUEST: a Bayesian adaptive psychometric method. *Percept Psychophys* 33:113–120
- Wichmann FA, Hill NJ (2001a) The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept Psychophys* 63:1293–1313
- Wichmann FA, Hill NJ (2001b) The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Percept Psychophys* 63:1314–1329
- Wickelgren WA (1968) Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *J Math Psychol* 5:102–122
- Wikipedia (2010a) Central limit theorem. http://www.en.wikipedia.org/wiki/Central_limit_theorem
- Wikipedia (2010b) Error function. http://www.en.wikipedia.org/wiki/Error_function
- Zupan LH, Merfeld DM (2008) Interaural self-motion linear velocity thresholds are shifted by roll vection. *Exp Brain Res* 191:505–511