Published in final edited form as:

Exp Brain Res. 2012 October ; 222(3): 303–320. doi:10.1007/s00221-012-3217-2.

Signal Detection Theory and Vestibular Perception: II. Fitting Perceptual Thresholds as a Function of Frequency

Koeun Lim¹ and Daniel M. Merfeld^{1,2}

¹Massachusetts Eye and Ear Infirmary, Jenks Vestibular Physiology Lab, 243 Charles St. Suite 421, Boston, MA, USA

²Harvard Medical School, Otology and Laryngology, MA, USA

Abstract

Vestibular perceptual thresholds are defined by a dynamic sensory system. To capture these dynamics, thresholds were previously fit as a function of frequency. In this paper, we compare fits using two published models with two new models. Furthermore, a new fitting method that utilizes vestibular perceptual dynamics is developed to improve fit quality and overcome problems associated with the conventional approach. Combinations of the four models and two fitting methods are tested using both simulated data and previously published experimental data. Simulations reveal that the conventional approach underestimates thresholds when the number of trials at each frequency is limited (circa 50); this underestimation is reduced fivefold by the new fitting method that simultaneously utilizes data across frequencies. The new fitting method also scored best for goodness of fit for both the simulations and experimental data. In fact, the new approach of fitting simultaneously across frequencies proved more accurate, more precise, more robust, and more efficient than the conventional approach of fitting the responses at each frequency individually and then fitting these threshold data across frequency. The revised fit of published yaw rotation threshold data show that these are best fit by a first-order high-pass filter having a plateau of 0.5°/s (roughly a factor of 4 higher than the motion platform vibration) at frequencies above the cut-off frequency of 0.26 Hz, which is well above the cut-off frequency of the semicircular canals (circa 0.03 Hz). This dynamic analysis suggests the contributions of a velocity leakage mechanism to human yaw rotation thresholds.

Keywords

Detection theory; Vestibular perception; Threshold; Psychophysics; Generalized linear model

Introduction

As mentioned in the first paper in this series (Merfeld 2011), “signal detection theory has long been used to guide the design and analysis of vestibular studies”, and “there has been a recent resurgence of interest in the application of signal detection theory to vestibular

tel: +617-573-5595; fax: +617-573-5596; dan_merfeld@meei.harvard.edu.

Conflict of Interest

The authors declare that they have no conflict of interest.

responses.” Given this resurgence in interest, we (Merfeld 2011) theoretically analyzed the application of signal detection theory to vestibular discrimination – including an analysis of both vestibular detection and vestibular direction recognition. We also analyzed both one-interval and two-interval designs. Among several other contributions provided by this analysis, we concluded that theoretical and practical considerations (detailed in the paper) showed that one-interval direction-recognition has certain advantages over one-interval detection, two-interval recognition, and two-interval detection, so we focus exclusively on one-interval direction-recognition herein.

This second paper in the series presents a computational study that investigates how to fit binary discrimination responses from a one-interval direction-recognition task as a function of frequency. Since such fits provide a fundamental way to study the dynamics of vestibular thresholds, it is crucial that such fits be performed so as to yield efficient, accurate, and precise results. We explicitly assess fit accuracy (closeness to true value), precision (repeatability), and efficiency, where we define efficiency as extracting maximal information from the minimal number of trials..

Early studies of vestibular thresholds focused on non-dynamical aspects of the responses (Mann et al. 1949; Graybiel et al. 1946; Clark and Graybiel 1963; Walsh 1961). About 25 years ago, Benson recognized the importance of threshold dynamics, which he studied by measuring motion perception evoked by single-cycle sinusoid accelerations using a one-interval direction recognition task (Benson et al. 1989; Benson et al. 1986). He used standard protocols in which the subjects were required to indicate their motion direction after each trial. Equipment limited his threshold studies to frequencies between 0.05 and 1.11 Hz for yaw rotation and 0.14 and 1.02 Hz for inter-aural translation. Over this frequency range, thresholds appeared to more-or-less linearly decrease as frequency increased, which prohibited fits more complicated than lines. With modern equipment, Grabherr and colleagues (Grabherr et al. 2008) used similar single cycle sinusoidal stimuli and measured thresholds across a full two-decade range between 0.05 and 5.0 Hz and found that the threshold dynamics might be modeled empirically using a high-pass filter. It is important to note that this fit should be considered phenomenological as it fit a steady-state transfer function to the data, but the underlying signals evoked by a single cycle of sinusoidal acceleration are not at steady-state.

Recently, Soyka and colleagues (Soyka et al. 2011) measured translation thresholds using three different motion trajectories – trapezoids, triangular waves, and sinusoids - at three durations (1.5, 2.36, and 5.86 s) each. For comparison to the Benson and Grabherr studies, the sinusoidal stimuli used by Soyka and colleagues were equivalent to single-cycle stimuli at frequencies between 0.17 and 0.67 Hz. In their study, Soyka and colleagues passed the relevant signal (trapezoidal, triangular, or sinusoidal) through a second order transfer function having three free parameters – two transfer function time constants and one gain. As noted in their paper, this model is physiologically motivated and provides a way to compare stimuli having different shapes as well as different periods.

Building upon these two pre-existing models, we developed and tested two new models. We use computer simulations to create data from an “ideal” subject having known parameters to

test each of these four models. As will be described in the methods, one of the new models is a simple phenomenological fit using the curve $\sigma(1 + f/f_0)$, where f is frequency, and σ and f_0 are fitted parameters to be described in the methods. Like the model used by Grabherr et al. (2008), this model can not describe an implementation representative of the brain's implementation. However, because a polynomial fit is more general than a transfer function, it can be applied more generally to other threshold measures that show any kind of parametric dependence. In fact, it is important to note that in the general case, the polynomial fit could be replaced with any equation that captures the form of threshold variations as a function of another parameter (e.g., spatial frequency, time, etc.). In this limited sense, parameters obtained from phenomenological models are more robust than parameters estimated using transfer function models that consider dynamic characteristics. However, when the dynamic model function is representative of the implementation, the dynamic model is likely to yield a better fit than the phenomenological model, and the estimated parameters are truly indicative of neural implementation.

The second new model is similar to the transfer function fit described above (Soyka et al. 2011) but assumes unity-gain in a transfer function fit. This reduces by one the number of free parameters fit to the data. This model includes the implicit assumption that the underlying noise does not vary with stimulus frequency. This was also an implicit assumption for the Soyka (2011) model but not for either of the phenomenological models. Noise here refers to all physiological noise accumulated along all neural pathways associated with motion direction discrimination process. A recent report of VOR (vestibulo-ocular reflex) thresholds (Haburcakova et al. 2012) suggests that this assumption may be valid. However, it is certainly possible that this assumption may not hold true for perception, since perception utilizes different neural pathways and mechanisms than the VOR (Merfeld et al. 2005a; Merfeld et al. 2005b).

The primary contribution demonstrated in this paper is a new fitting *method* that we developed, implemented, and tested. Before proceeding, it is illustrative to point out that all four models described above first fit a psychometric function to the data at each frequency, yielding a bias and threshold estimate at each frequency, and then fit the resultant thresholds across frequencies. We show a new method that fits a model across frequencies simultaneously. No additional assumption was made for the new method beyond those required for fitting the transfer function models. (See previous paragraph.) We test this new approach using the two new models outlined in the previous paragraph. We find that this new method of fitting a single psychometric function across frequencies yields more accurate, more precise, and more efficient parameter estimation than the approach of fitting a psychometric function at each frequency followed by a fit of the resultant thresholds across frequency.

Finally, to test the method and illustrate some practical advantages using “real” data, we test and compare these modeling approaches using real, published yaw rotation threshold data (Grabherr et al. 2008).

Methods

An overview – fitting model functions to data

The goals of this paper are 1) to test existing models for fitting vestibular threshold data numerically, 2) to introduce and test two new models, 3) to introduce and test a new fitting method, and 4) to refit previously published data. While much effort is dedicated to testing new and existing models, the primary contribution of this paper are the new fitting method and the reanalysis of published data. The psychometric fit, model functions, fitting algorithms, and simulation methods are detailed below.

Threshold in relation to bias and spread of a psychometric function

For the i^{th} trial of a one-interval direction-recognition threshold task, a motion stimulus with amplitude of Ω_i in either the positive or negative direction is presented. In response, the subject reports the perceived direction ($y_i = 1$ if positive, 0 if negative). These binary decision data can be fit with a continuous distribution $\psi(\Omega)$ called a psychometric function, which is a plot of the probability of a positive response versus the motion stimulus amplitude Ω . The first and second central moments – sometimes referred to as the mean (μ) and spread (σ), respectively, of the psychometric function – can be estimated by assuming it can be modeled by a cumulative Gaussian probability distribution $\psi(\Omega) \approx \Phi(\Omega; \mu, \sigma)$, where Φ represents the cumulative Gaussian distribution. The Gaussian distribution is justified by the central limit theorem of statistics (Jaynes and Bretthorst 2003) and the ability to fit experimental data. For direction-recognition, μ indicates the point of subjective equality (PSE) where perceived motion is zero.

In signal detection theory, a continuous psychometric function infers background noise of the system where the spread σ corresponds to the standard deviation of the physiologic noise (Merfeld 2011). Threshold T is arbitrarily defined as some multiple γ of σ , specifically, $T = \gamma\sigma$. For a Gaussian distribution, γ is equivalent to a z-score at a target decision confidence

level of $p = p_T$ such that $\gamma = \frac{\Phi^{-1}(p_T; \mu, \sigma) - \mu}{\sigma}$ (in MATLAB, `icdf('norm',  $p_T$ ,  $\mu$ ,  $\sigma$ )`), where Φ^{-1} is the inverse of the cumulative Gaussian distribution. It is worth noting that γ is not a physical parameter, and hence it can be set at any arbitrary level, typically on the order of 1.

γ is sometimes predetermined by experimental procedures; for instance, in the 3-down 1-up staircase paradigm that was used by Grabherr et al. (2008), the target decision confidence level is set to $p_T = 0.5^{1/3}$ ($= 0.794$) (Leek 2001), which yields $\gamma = 0.82$. Once γ is known, thresholds can be converted to any desired level of γ . In this paper, thresholds estimated using standard staircase methods $\hat{T}_s$ were recalculated at $\gamma = 1$ ($p_T = 0.84$). Here a hat $\hat{}$ is placed on staircase threshold because $\hat{T}_s$ is an approximation that the staircase procedure yields, not a true value. Hereafter, a value with a hat $\hat{}$ indicates an estimate.

In order for the staircase procedure to work well for a direction recognition task, two conditions must be satisfied: 1) the psychometric function must be symmetric about the PSE; and 2) the PSE must be at zero, which means no vestibular bias ($\mu = 0$). If the PSE deviates from zero, $\hat{T}_s$ is no longer accurate. (See appendix.) Furthermore, because a staircase threshold is calculated from only the last portion of a session, not the entire test

session, it typically results in reduced precision (i.e., increased variance of the estimated threshold). To overcome such inaccuracy and imprecision, various hybrid procedures were developed (Hall 1981) that combined adaptive and psychometric function fitting procedures. We mimic this approach and apply maximum-likelihood (ML) estimation method to staircase data. In order to calculate ML estimation threshold T_{ML} , we estimated the mean $\hat{\mu}$ and spread $\hat{\sigma}$ that maximized the likelihood of the psychometric fit to the data (Ω and y) acquired using a staircase procedure. For convenience, T_{ML} was also defined as $1\hat{\sigma}$ ($\gamma = 1$), which we call the “one-sigma threshold.”

Psychometric fits were always performed using a generalized linear model (GLM). Specifically, a Gaussian distribution function $\phi(\Omega; \hat{\mu}, \hat{\sigma})$ was fitted to the data using a GLM fit with a probit link function. The data fit was performed with two vectors, Ω and y . The function call in MATLAB was $b = \text{glmfit}(\Omega, y, 'binomial', 'link', 'probit')$ using the statistics toolbox (v 7.0), where b is a 2-element vector. The elements of b are related to underlying Gaussian distribution parameters, such that the estimated bias is $\hat{\mu} = -b_1/b_2$, and the estimated spread is $\hat{\sigma} = 1/b_2$ (Dobson and Barnett 2008). (An outline of a derivation of these two relationships can be found in the appendix.) In fact, the invariance property of maximum likelihood estimators (Dobson and Barnett 2008) guarantees that $\hat{\mu}$ and $\hat{\sigma}$ are maximum likelihood estimates even though the maximum likelihood model fit was performed on b_1 and b_2 . Here the likelihood L of $\hat{\mu}$ and $\hat{\sigma}$ for given stimulus and response vectors Ω and y can be defined as:

$$L = \prod_i \phi_i(\Omega_i^{y=1}; \hat{\mu}, \hat{\sigma}) \cdot \prod_i \phi_i(\Omega_i^{y=0}; \hat{\mu}, \hat{\sigma}), \quad \text{equation (1)}$$

where $\phi_i(\Omega_i^{y=0(1)}; \hat{\mu}, \hat{\sigma})$ indicates probability of the response 0 (or 1) happening at i^{th} motion (Ω_i with direction) for the estimated set of $\hat{\mu}$ and $\hat{\sigma}$. Deviance D can then be defined in terms of negative log of L such that $D = -2 \log(L)$. Maximizing likelihood is equivalent to minimizing deviance.

Frequency content of motion stimulus

Motion stimuli used in this study are unidirectional angular velocity trajectories that result from the integration of a single cycle of sinusoidal angular acceleration (Grabherr et al.

2008; Benson et al. 1989). The velocity is $\Omega \left(\frac{1 - \cos(2\pi ft)}{2} \right)$, where Ω is the peak velocity, and f is the frequency. The velocity stimulus is a raised cosine bell that starts and ends with 0 velocity and 0 acceleration. In this study, we examine the dependence of vestibular perception on the frequency components f by modeling direction recognition thresholds as a function of this frequency.

Four dynamic vestibular threshold models

In this paper we propose two new dynamic models and compare these with two pre-existing models previously introduced by Grabherr et al. (2008) and Soyka et al. (2011). All four models can be generalized to estimate thresholds (T) as a function of frequency f such that

$$\hat{T}(f) = \frac{\alpha}{\beta(f; \tilde{f})}, \quad \text{equation (2)}$$

where α is a coefficient to which the curve fit asymptotes (“plateaus”), and $\beta(f; \tilde{f})$ is a function of frequency f that gives the shape of the fit with $\tilde{f}$ as a fit parameter.

The four models can be categorized into two types: 1) phenomenological models that describe threshold patterns as a function of frequency (models #1 and #2); and 2) dynamic filtering models that provide possible implementations for how the brain dynamically processes information when generating signals that yield perception discrimination (models #3 and #4).

The first model is a first-order function of the form:

$$\hat{T}_1(f) = \tilde{\sigma}(1 + \tilde{f}/f). \quad \text{equation (3)}$$

This fit is simple and provides a good phenomenological fit. A primary shortcoming is the absence of a direct physiological interpretation. The primary advantages are: 1) that the approach can be applied generally using any function that captures a relationship between thresholds and a characteristic of the physical stimuli (not limited to frequency relations embedded in transfer functions), and 2) that the fit is independent of the underlying mechanism, so the fit values do not need to change as we learn more about the dynamics of the underlying neural mechanisms. This may be particularly relevant to the study of threshold dynamics at this time, since the available dynamic data are not extensive. It is straightforward to extend this fit to a 2nd (or higher) order polynomial, but this did not yield an improvement in accuracy for data fit herein, so we only present the first order model.

The second model (Grabherr et al. 2008) fits the steady-state magnitude of the inverse of a high-pass filter, $T_2(f) = |(1 + \tau\tilde{S})/K\tau\tilde{S}|$, where s is the Laplace variable. Grabherr et al. (2008) chose this transfer function because the data appeared loosely consistent with the inverse of a high-pass filter. The transfer function is inverted because the output signal from a unity-gain high-pass filter is smaller than the input, especially at low frequencies, which means that a larger input signal (stimulus amplitude) is required for the output to reach threshold. Substituting $2\pi f\tilde{f}$ for s and $1/2\pi f$ for τ , we obtain

$$\hat{T}_2(f) = K^{-1} \sqrt{1 + \left(\frac{\tilde{f}}{f}\right)^2}. \quad \text{equation (4)}$$

With $K^{-1} = \sigma'$, this fit has a form similar to the first fit.

This second fit assumes steady-state, which is an invalid assumption for a raised cosine bell velocity trajectory (as well as other motion trajectories). Therefore, this model provides a phenomenological fit. An advantage is that the fit parameters can be loosely interpreted – to the extent that transient effects (i.e., absence of steady-state) can be ignored – as transfer function parameters that represent neural dynamics.

The third model is similar to the second model in the sense that it also assumes a transfer function with first order high-pass dynamics. However the fundamental difference is that this model includes the full effect of transfer function dynamics including transient effects by filtering the motion input – in our case the bell-shaped velocity trace – through a high-pass filter. The parameters of the transfer function are varied to minimize the cost function error. Soyka et al. (2011) have used a similar method that included peripheral dynamics.

Using the `lsim` function in MATLAB, we passed a set of angular velocity input signals having unity amplitude at each frequency through a linear system having a first order high-pass transfer function $\tau\tilde{S}/(1 + \tau\tilde{S})$. The basics for this fitting algorithm are similar to those previously described for the Soyka et al. (2011) approach, with a few differences are highlighted in the following sentences. Unlike in Soyka's work presenting linear acceleration thresholds where the absolute peak magnitude of filtered acceleration was used, we instead used the magnitude of the peak velocity. (Recall that acceleration was always symmetric for our single cycle sinusoidal stimuli with the magnitude of the peak negative acceleration equal to the peak of the positive acceleration.) One divided by the peak of the output signal yielded a curve that had a constant proportion to $\hat{T}_{ML}(f)$, and a scalar parameter K^{-1} was also fitted. Mathematically this can be written as:

$$\hat{T}_3(f) = \frac{K^{-1}}{(\tilde{\Omega}(t;f) * \text{hpf}\{\delta; \tilde{f}\})_{peak}}, \quad \text{equation (5)}$$

where $\tilde{\Omega}(t;f)$ is an angular velocity time trace having a peak 1 at each frequency, $\text{hpf}\{\delta; \tilde{f}\}$ is the impulse response of the first order high-pass filter with cutoff frequency f , and $*$ represents the convolution operator. As in the second model, with an adequate set of data, K^{-1} converges asymptotically to σ , representing the underlying noise level that determines the threshold.

The fourth model implements a high-pass filter like the third model, and includes the same implicit assumption described immediately above, but it utilizes additional information – specifically that detection theory (Green and Swets 1966; Macmillan and Creelman 2005) defines vestibular thresholds as some constant fraction of the noise (Merfeld 2011). In other words, the fourth model is not derived from peripheral dynamics, but instead, it aims to describe the predominant perceptual dynamics of peripheral and central vestibular signal processing.

Assuming that the neural signal is proportional to the peak of the perceived angular velocity, the peak of the high-passed threshold motion at threshold should be constant regardless of the stimulation frequency. Following this logic, we first pass a threshold motion time trace (in our case a raised cosine-bell velocity) through the first-order high-pass filter with cutoff frequency f to yield a filtered output time trace whose peak should be constant at threshold across the observed frequency range. This can be represented mathematically as:

$$\hat{T}_4(f) = \frac{(\Omega_T(t;f) * \text{hpf}\{\delta; \tilde{f}\})_{peak}}{(\tilde{\Omega}(t;f) * \text{hpf}\{\delta; \tilde{f}\})_{peak}}, \quad \text{equation (6)}$$

where $\Omega_T(t; f)$ represents the angular velocity time trace at threshold for frequency f . This equation is the same as the third model fit except that the fit parameter K^{-1} , which was shown to represent the noise, is replaced by the high-pass filtered velocity threshold motion. To provide a physical explanation, this simply indicates that the high-pass filtered signal equals the standard deviation of the noise at threshold ($(\Omega_T(t; f) * \text{hpf}\{\delta; f\})_{\text{peak}} = \sigma$).

The primary advantage of the third and the fourth models is that, unlike the first two models, these latter models explicitly include the full representation of dynamic influences, including transient response dynamics, and hence, can be applied to non-sinusoidal stimuli. Furthermore, the fitted parameters will represent overall response dynamics to the extent that the form of the model accurately captures the form of the system being fitted. A notable difference is in utilization of parameters σ and K^{-1} in fitting processes. In the fourth model, σ is not a free parameter added to the fit; rather, it interprets residuals as a physiological property (noise). For comparison, K^{-1} is an additional free parameter that can also be interpreted as the same physiological property (noise). Table 1 summarizes the features of new and pre-existing models.

Fitting methods – algorithms and cost functions

As described above, fits can be accomplished via two different approaches: (1) by performing psychometric fits at each frequency to obtain $T_{ML}(f)$ and then fitting a model to find α and f (“conventional” method); and (2) by obtaining $\beta(f; f)$ that scales the motion stimulus amplitude vector Ω to fit a single psychometric function across frequencies to yield α (the new “across-frequency” fit method).

For the “conventional” method, ML thresholds $T_{ML}(f)$ were first determined at each of several frequencies, and then a function that captures the relevant dynamics was fitted to thresholds using optimization algorithms. We used cost functions that included mean squared log error (MSLE) as opposed to mean squared error or absolute error. For models #1, #2, and #3, the cost function errors we minimized had a general form $mean([(\log T_{ML}(f) - \log T(f))^2]$, where $T_{ML}(f)$ and $T(f)$ are ML thresholds and model fit thresholds, respectively. There are several justifications for using squared log differences to form the cost functions. The primary justification is “practical” as simulations showed that this form of the cost function converged well – better than mean square error, for example. The log can also be justified since for some cost functions this more evenly emphasizes all frequencies independent of whether the calculations are done using velocity, acceleration, or displacement as the input (Soyka et al. 2012). The cost function for model #4 similarly uses MSLE as a cost function, but the optimization algorithm looks for minimum log variance in the peak of the high-pass filtered ML threshold motions, i.e., $mean([mean(\Omega_{ML}(t; f) * \text{hpf}\{\delta; f\})_{\text{peak}} - (\Omega_{ML}(t; f) * \text{hpf}\{\delta; f\})_{\text{peak}}]^2)$, where $\Omega_{ML}(t; f)$ is angular velocity time trace having a peak $T_{ML}(f)$ at each frequency.

A shortcoming of the conventional approach is that the number of trials at each frequency must be great enough to ensure adequate accuracy and precision. With the staircase procedure used for this study there were approximately 55 trials on average per frequency. Fitting a maximum likelihood fit with this number of trials yields threshold underestimation due to the small number of available data points (McCullagh and Nelder 1989). The

conventional approach also has the disadvantage of having a large number of free parameters. In addition to free parameters α and $\tilde{f}$ in a model fit, PSE(μ) and spread (σ) have to be fit at each frequency in order to calculate a ML threshold. Thus, for m frequencies (six for simulations, and seven for experimental data (Grabherr et al. 2008)), the resulting number of free parameter is $2m+k$, where k is either one or two depending on one or both of α and $\tilde{f}$ being free parameter(s). For simulations, the first three models have 14 free parameters while the fourth model has 13, one less than the others, because α in the last model is not a free parameter added to the model.

The second “across-frequency” approach is a novel method that utilizes the relationship $T(\tilde{f}) \cdot \beta(\tilde{f}; \tilde{f}) = \alpha$ to scale stimulus vectors Ω across all frequencies before fitting a single dimensionless psychometric function. To help understand the rationale underlying this second approach, recognize that if the brain is attenuating (e.g., filtering) the signal, then we should attenuate the signal by the same amount to yield the magnitude of the neural signal being discriminated. Such filtering, when it matches the applicable neural filtering would yield the actual neural signal (Figure 1). Therefore, we applied the attenuation represented by each model (equations 3, 4, 5 or 6) to the angular velocity motion to scale the signal magnitude to yield the neural signal being discriminated. A single psychometric function is then fitted to these scaled individual trial data simultaneously across all frequencies, yielding a fitted PSE $\mu(\hat{=}\tilde{\mu})$ and spread $\sigma(\hat{=}\tilde{\sigma})$. Here the values on the x-axis of the psychometric function no longer correspond to actual stimulus magnitude, but instead represent theoretical magnitude of the neural signal. Hence threshold then corresponds to a given magnitude of the neural signal regardless of frequency.

For this approach, deviance is the cost function that is minimized since deviance is the standard “natural” cost function for GLM fits of binominal data (Dobson and Barnett 2008; McCullagh and Nelder 1989) and for fitting psychometric functions using maximum-likelihood methods (Wichmann and Hill 2001a). Deviance is defined in terms of negative log likelihood such that

$$D = -2 \sum_{j=1}^m \sum_{i=1}^{n_j} \log[\phi_{j,i}(\tilde{\Omega}_{j,i}^{y=1}; \tilde{\mu}, \tilde{\sigma})] - 2 \sum_{j=1}^m \sum_{i=1}^{n_j} \log[1 - \phi_{j,i}(\tilde{\Omega}_{j,i}^{y=0}; \tilde{\mu}, \tilde{\sigma})], \quad \text{equation (7)}$$

where i represents the i^{th} trial, n_j is the total number of trials at the j^{th} frequency, and $\tilde{\Omega}$ is the motion stimuli amplitude scaled by the fitted model. For instance, with y being the response vector (0=negative direction, 1=positive direction), $\phi_{j,i}(\tilde{\Omega}_{j,i}^{y=1}; \tilde{\mu}, \tilde{\sigma})$ indicates the probability of the response being positive for the i^{th} trial at the j^{th} frequency where $\tilde{\Omega}_{j,i}^{y=0,1} = \Omega_{j,i}^{y=0,1} \cdot \beta(\tilde{f}_j; \tilde{f})$. Here the probability is determined based on the normalized psychometric function defined with $\tilde{\sigma}$ being equivalent to the free parameter α from Table 1, and $\tilde{\mu}$ being the frequency-independent scaled PSE.

Because there is only one psychometric function fit, there are only three free parameters, $\tilde{\mu}$, $\tilde{\sigma}$, and $\tilde{f}$ for each of the four models. In fact, due to their similarity, models #3 and #4 yield nearly identical parameter estimations when used with the new fit method. Therefore, we only proceed with two models (#1 and #4) using the new fit.

Furthermore, this approach also potentially relieves the underestimation posed by the conventional approach since the number of data points used in the GLM fit is much greater than the average number of trials at each frequency, summing to 333 trials on average for the second approach as opposed to an average of 55 trials at each frequency in the case of the first approach.

For both approaches α was constrained to be greater than zero, and $\tilde{f}$ was constrained to be within the range of frequencies tested (from 0.09Hz to 5.1Hz for simulations and from 0.04Hz to 5.1Hz for experimental data). If $\tilde{f}$ fell outside the observation range, it was flagged as a fit with one or more of the parameters “out-of-range”; all fit parameters for that data set were excluded from further analysis.

Information criteria as measures of goodness of fit

To assess goodness of fit, we used deviance (D , see previous section), which is the standard quality of fit statistic used for maximum-likelihood estimation method in general and GLM fits in particular (McCullagh and Nelder 1989; Dobson and Barnett 2008). Further more, since the number of free parameters varied depending on models and approaches, we assessed fit quality using standard information criteria. Specifically, we calculated AIC_c (corrected Akaike Information Criterion) and BIC (Bayesian Information Criterion):

$$AIC_c = D + 2N\rho / (N - \rho - 1), \quad \text{equation (8)}$$

$$BIC = D + 2\rho \log N, \quad \text{equation (9)}$$

where N is total number of trials for all frequencies, and ρ is the number of free parameters (Burnham and Anderson 2004; Berg et al. 2004; Celeux et al. 2003).

Information criteria are a standard method to assess goodness of fit. A different form of information criteria calculations based on mean square error were recently utilized in a similar context for evaluating threshold models and the number of free parameters (Roditi and Crane 2012). The primary difference between the two approaches is that we calculate AIC and BIC using deviance, which is the standard cost function for maximum likelihood fits. Roditi and Crane used mean square error, which is the standard cost function for most other approaches.

Simulations

To test both new and existing model fits, one-interval direction-recognition thresholds for an “ideal” subject were simulated using a 3-down 1-up staircase procedure like that used for the “real” subjects. The 3-down 1-up procedure continued until there were nine reversals – five minima and four maxima reversals, including the initial minimum. Step size was log-doubled or log-halved. Initial stimulus magnitude was eight times threshold ($8T$). Staircase threshold $\hat{T}_s$ was calculated by taking the arithmetic mean of the last pair of reversals. Each set of simulated experiments consisted of six sessions that each corresponded to one of six frequencies, 0.1, 0.2, 0.5, 1, 2, and 5Hz.

At each frequency, the simulated responses were determined using three “known” parameters: (1) the PSE (point of subjective equality), which serves as mean for a psychometric function, was set to one of three values, 0°/s, 0.2°/s, and 0.5°/s; (2) the spread of a psychometric function was set to 1°/s; and (3) the characteristic frequency was set to 0.4Hz. It is worth noting that a bias of 0.5°/s is 50% of the threshold, so this represents a large bias for normal subjects - but it is a bias level that we occasionally observe experimentally.

Our “ideal” subject was simulated mimicking the approach taken in earlier studies (Wichmann and Hill 2001a; Wichmann and Hill 2001b). Monte Carlo simulations were used to generate binomial data using a cumulative Gaussian distribution with mean μ and standard deviation σ (Merfeld 2011). Specifically, in MATLAB, for a given stimulus level ω_i , the probability of the “ideal” subject responding positive is $p_i = \text{cdf}(\text{'norm'}, \Omega_i, \mu, \sigma)$. The binary response for each individual trial was then generated by a MATLAB function `binornd(1,  $p_i$ )`, which randomly generates single positive response with probability or single negative response with $1 - p_i$. Because of the random effects, the staircase procedure generates a different stimuli-response set – and hence different fit parameters – for each simulated test session.

1000 staircase procedures were simulated at each frequency with normalized spread of one to create identical normalized data to be used for all underlying models, and then stimuli vectors were scaled using the underlying models. Consequently, all model simulations had the same underlying noise variations, enabling more direct comparisons among and between the models and fit methods.

Experimental setups and subjects

To also test these models and approaches using real data, we applied the methods to published data from earth-vertical yaw rotation perception threshold study. See Grabherr et al. (2008) for methodological details. In brief, subjects were tested at seven different frequencies, 0.05, 0.1, 0.2, 0.5, 1, 2, and 5 Hz. From 0.1 to 5Hz, an adaptive 3-down 1-up single-interval direction-recognition staircase procedure was used to obtain thresholds (Grabherr et al. 2008). At 0.05Hz, a non-adaptive constant motion stimuli procedure was used instead of a staircase procedure because lower frequency motion stimuli require larger displacements for the same magnitude of velocity. Therefore, a full staircase procedure was not possible at 0.05Hz, which was the lowest frequency tested.

Among the nine normal subjects tested in the earlier study, we used data from eight normal subjects (4 males and 4 females, age 37.5 ± 11) – one more than Grabherr et al. (2008) included in their final analysis. (See discussion for information related to this “extra” subject inclusion.)

Fitting models to experimental data

Perception threshold responses were fitted for each subject using the GLM and least squares methods described above. The fit parameters were used to generate threshold curves as a function of frequency for each subject. Because the threshold distribution across subjects

was not normal and was not significantly different from lognormal (Grabherr et al. 2008), the logarithmic mean and the standard error of the log-mean were calculated across subjects to yield pooled results across subjects.

Results

Staircase vs. ML thresholds for individual frequencies - simulations

When the PSE was zero, our simulations showed that on average the staircase thresholds (T_s) yielded 4% overestimation while ML thresholds (T_{ML}) yielded 7% underestimation. When the PSE was non-zero, the staircase threshold overestimation increased substantially. For instance, when the PSE increased to 0.2°/s and 0.5°/s (20% and 50% of σ , respectively), the threshold overestimation increased to 7% and 19%. In comparison, the ML threshold underestimation was unchanged when the PSE equaled 0.2°/s (7%) and increased to 13% when the PSE equaled 0.5°/s. (See discussion for details.)

Furthermore, the staircase threshold's precision was found to be worse than that of ML threshold. For example, for zero PSE, the standard deviation of the staircase thresholds was 0.35, which was 16% greater than that of ML thresholds (0.30).

Maximum-likelihood fitted parameter distributions

Figure 2 shows example histograms of the fitted parameters – PSE (μ), ML thresholds (σ), and characteristic frequency (f) – obtained from the simulations. For simplicity, we show the histograms for a fit of the fourth model using the new across-frequency fitting algorithm, since this fit is fully characterized by these three parameters. (Other fits had more free parameters as noted before.) Similar distributions for these three fit parameters – which were the three primary parameters of interest – were obtained for the other model fits.

Simulated thresholds at each frequency had symmetric distributions (skew<2) (Bulmer 1979) as opposed to the right-skewed distribution found experimentally (Grabherr et al. 2008) for 'real' subjects (skew>2). This suggests that the skew reported for experimentally measured thresholds (Benson et al. 1989; Grabherr et al. 2008; Soyka et al. 2011) is due to variations across subjects. On the other hand, for both 'ideal' and 'real' subjects, the PSE was symmetric (skew<2) while the characteristic frequency f was right-skewed (skew>2). To compensate for the right-skewness, f was logarithmically transformed, which is identical to the transformation performed for earlier threshold measures (Benson et al. 1989; Grabherr et al. 2008). Therefore, to estimate the first moments for the simulations, the arithmetic mean was used for threshold and PSE, while the logarithmic mean (geometric mean) was used for the characteristic frequency. On the other hand, for averaging experimental data, the arithmetic mean was used for PSE, and the logarithmic mean was used for threshold and characteristic frequency.

Conventional fitting method – simulations

Figure 3 and table 2 show the fitted threshold curves for the four models using the conventional fitting method with vestibular bias equal to zero. Nearly identical results (not shown) were obtained when vestibular bias varied from zero. All four fits displayed similar

trends. Specifically, estimated μ values were unbiased, having a mean less than 0.01%, whereas estimated σ values were underestimated by 10%. f values were estimated with the mean staying within 1% of the true value. Mean number of total trials was 333 with standard deviation of 21.

Table 3 summarizes goodness of fit for the four models. Goodness of fit scores were calculated only for the simulations (M out of 1,000) that satisfied the condition that the fitted characteristic frequency was within the measured frequency range, from 0.1Hz to 5Hz. All model fits had a convergence failure rate of approximately 2%. Deviances were virtually identical for the four models; while the fourth model gave better information criteria scores because it had one fewer free parameter. Specifically, all but the fourth model have 14 free parameters, 12 of which are PSE and threshold at each of 6 frequencies, and 2 of which are α and f . The last model has 13 free parameters because α (“plateau” level) is a parameter that emerges from the fit and not a free parameter of the fitting procedure.

New across-frequency fitting method – simulations

When the two new models, models 1 and 4, were fitted using the new “across-frequency” fit method, it resulted in several improvements. First, underestimation of σ was reduced to 2% (Figure 4 and Table 4), which is about one fifth that resulting from the conventional fitting method (Table 3). Also, the failure rate dropped to 0.2% (Table 5), which indicates that the novel method is more robust. Furthermore, these new model fits had better goodness of fit scores because these fits had fewer free parameters. Combined effects of these improvements are apparent in Figure 4 and Table 4. Beyond significant reduction in σ underestimation, the other estimated parameters maintained the same accuracy as for the conventional fitting method. Furthermore, the precision (std) for each fit parameter was also somewhat improved by the novel approach – roughly a 5% improvement each for μ and σ and roughly a 20% improvement for f .

Table 5 summarizes goodness of fit scores for the new across-frequency fitting approach. Deviance scores reduced slightly from those obtained using the conventional fitting approach (Table 3), reflecting the improved fits observed in Figure 4. But goodness of fit scores showed a more substantial reduction, especially in BIC, which penalizes the number of free parameters the most.

Staircase vs. ML thresholds for individual frequencies – experimental data

Consistent with simulation results, lower values were observed for mean ML thresholds than for staircase thresholds except at 0.05Hz (Figure 5), where thresholds were measured using a different procedure (Grabherr et al. 2008). Consistent with previous reports (Grabherr et al. 2008; Benson et al. 1989), fitted ML thresholds did not deviate significantly from a lognormal distribution (Anderson-Darling test, $p>0.4$) across subjects. While fitted ML thresholds across subjects also did not deviate significantly from a normal distribution (Anderson-Darling test, $p>0.15$), 4 out of 7 frequencies showed positive skewness ($\text{skew}>2$), which is consistent with lognormal distribution.

Conventional fitting method – experimental data

Figure 6 shows models fitted to ML thresholds, and parameters found from the fits are summarized in Table 6. Shapiro-Wilks and Kolmogorov-Smirnov tests showed that PSE's and characteristic frequencies of eight subjects did not significantly deviate from normal and lognormal distributions, respectively ($p > 0.3$ and $p > 0.6$). Here the PSE's were obtained by first multiplying the scaling function $\beta(f; f)$ to the raw $\mu(f)$'s obtained at each frequency per subject, which were then averaged across frequencies per subject to estimate μ 's. Mean PSE's in Table 6 indicate arithmetic means of PSE's across subjects. All models yielded mean PSE less than 0.01 °/s with standard deviation of 0.05 °/s or less. However, it is important to point out that even though the average PSE's did not significantly deviate from 0 (t-test, $p = 0.205$), three subjects had non-zero PSE's (t-test, $\mu = -0.12^\circ/\text{s}$, $p = 0.018$; $\mu = 0.21^\circ/\text{s}$, $p = 0.018$; $\mu = -0.13^\circ/\text{s}$, $p = 0.033$). For the phenomenological fits (models #1 and #2), the mean plateau occurred at different levels, at 0.38 °/s for the model #1 and at 0.48 °/s for the model #2. Since the characteristic frequency for both was similar being approximately 0.5Hz, the difference in the plateaus is due to the slope of the model #2 changing faster than the model #1. For model #3 and model #4, both attained the same plateau of 0.4 °/s and cutoff frequency of 0.27Hz.

Table 7 showing the goodness of fit scores demonstrates that the fourth model was as accurate as the third model while requiring one less free parameter and thus yielding lower information criteria scores. Across all four models fitted using the conventional method, the deviance was nearly identical. When the number of free parameters was accounted for in goodness of fit information criteria scores, the fourth model again received the best score due to fewer free parameters. In spite of similar deviances, estimated characteristic frequencies were quite different, approximately 0.5Hz for phenomenological models (models #1 and #2) and 0.27Hz for models #3 and #4.

New across-frequency fitting method – Experimental data

Threshold prediction curves and fitted parameters – both averaged across subjects – are shown in Figure 7 and Table 8, respectively. Note that fitted curve (Figure 7) appears to be above the data points on average. This is consistent with individual threshold fits underestimating the actual threshold as demonstrated earlier by the simulations. In fact, when the first and the fourth models were fitted using the new across-frequency fitting approach, plateaus of both models increased by about 25%, which is greater than the 10% increase observed for simulations. This 25% increase is partially due to relief of threshold underestimation and may also relate to the improved fit of other parameters (bias and characteristic frequency). (Also note that simulations were performed with a single known “subject”, while the experimental fits are being performed across a population of subjects.).

Improvements in the goodness of fit yielded by the across frequency fit approach are demonstrated in Table 9. The deviance for these across frequency fits was roughly 5% lower than for the conventional fit approach (Table 7). Considering that the mean raw deviance varied by less than 2 across the 4 models (Table 7) this 13 point drop in deviance, which corresponds to one SEM (standard error of the mean), strongly suggests that the novel approach of scaling then combining all data across frequencies to construct a single

frequency-independent psychometric function yields better parameter estimation. When the number of free parameters was accounted for using information criteria scores, *AICc* decreased by approximately 3 SEM, and *BIC* by almost 13 SEM relative to the conventional approach (Table 7).

Discussion

Summary of Findings

Our primary goals in this study were to define and test methods to fit dynamic models that characterize vestibular perception thresholds as a function of frequency and to reanalyze published data. Using the conventional fitting method, the data show that the new high-pass filter model (#4) yielded better information criterion scoring than the other three models due to having fewest free parameters. The new polynomial fit model (#1) yielded performance that roughly matched the other phenomenological fit (#2). The two new models (models #1 and #4) were then tested using the new “across-frequency” fit procedure, which was the primary innovation introduced herein. As we will discuss in more detail below, we found that this new procedure was more robust (i.e., fewer failures to fit), more precise (i.e., lower standard deviation for fit parameters), more accurate (i.e., less biased fit parameters), and more efficient (i.e., lower information criterion scores) than the traditional fitting approach.

Procedure robustness, accuracy, precision, and efficiency

As previously described, fitting psychometric functions to small data sets often yields significantly underestimated thresholds. The new fitting method overcomes this underestimation problem by combining all data across frequencies, which directly increases sample number several fold. Specifically, simulations showed that the new fit procedure was more accurate, yielding a five-fold reduction in underestimation (Table 2 & 4). Simulation results also showed that the new fit procedure (“across frequency fit”) reduced failure rate by ten fold (Table 3 & 5) compared to the conventional fit procedure, demonstrating that the new fit procedure is more robust. In a clinical or lab setting, any test that yielded a fit failure would need to be repeated, so this also improves overall testing efficiency.

While more modest than the fore-mentioned improvements in robustness and accuracy, simulation results also showed that the new procedure yielded more precise parameter estimates. Specifically, the standard deviations for the estimated parameters were 5% to 20% smaller for the new fit method (Table 4) than for the conventional approach (Table 2).

The new procedure was also more efficient, which we demonstrate by referring back to the fact that the new method was more precise, more accurate, and more robust than the conventional method for the same number of trials. Of course, we could have maintained the same accuracy or precision and found that fewer trials were required using the new across frequency fit procedure – demonstrating that the new fitting procedure more efficiently utilized the available data. As a simple demonstration of this, we used 1000 simulations to determine the number of trials required by the new across frequency fit method to maintain the same precision on the estimate of sigma as the conventional method (Table 2) and new

method (Table 4). We found that the new cross-frequency fitting method required about 10% fewer trials to attain the same precision on the estimate of sigma.

Finally, the new procedure yielded better information criterion scores, which were shown to improve between 10% to 30% for the new fit procedure (Table 5) when compared to the conventional fit procedure (Table 3). The lowered information criterion for the new across-frequency fit procedure were the result of both decreased deviance and fewer parameters. In addition, model #4 yielded the lowest information criterion scores for the conventional fit procedure (Table 3), primarily due to the reduced number of free parameters.

Threshold estimation via generalized linear models (GLM)

Consistent with theoretical predictions (McCullagh and Nelder 1989), simulations revealed threshold underestimation for all model fits. This underestimation was caused by the small number of trials available to fit at each frequency, which is consistent with previous studies (Wichmann and Hill 2001a). This effect is readily remedied by simply increasing the number of trials since the underestimation decreases as n increases but this takes additional time and effort. As an alternative that is more efficient, we found that this underestimation became five times smaller when all trials across all frequencies were scaled and then combined to construct a single frequency-independent psychometric function. This effectively increases the number of trials fit by the GLM model fit without requiring any additional testing at any frequency.

While we are aware of several exceptions (Zupan and Merfeld 2008; Knoblauch and Maloney 2008; Yssaad-Fesselier and Knoblauch 2006) that utilized GLM fits, psychometric fits are often performed using numerical optimization methods like Nelder Mead simplex to maximize the likelihood. Putting numerical limits (e.g., round off errors) aside, both methods usually calculate the same answer, but GLM fits are guaranteed to converge to a unique global minimum (Wedderburn 1976) – most other numerical methods only guarantee local convergence. Furthermore, psychometric fits using GLM methods are robust, simple to perform (e.g., requiring just one command line in MATLAB), and “provide a unifying framework for many commonly used statistic techniques” (Dobson and Barnett 2008) from which additional specific benefits can accrue. Examples of specific insights include: (1) GLM fits provide a way to calculate the expected parameter estimation bias associated with small (circa 100 trials or less) samples (McCullagh and Nelder 1989); and (2) GLM fits provide a covariance matrix for the estimated fit parameters.

We maximize likelihood herein by using generalized linear models to fit both parameters of a Gaussian CDF to simulated data obtained using adaptive paradigms and experimental data obtained using adaptive paradigms. By doing so, we combine the accuracy and precision of maximum likelihood fits with the efficiency of adaptive paradigms. We showed that such fits of data obtained using adaptive test paradigms yield more accurate and more precise threshold estimates than those estimated using traditional staircase analyses (e.g., averaging track reversals).

Perception Thresholds as a function of frequency

Rotation perception thresholds have been found to display high-pass dynamic characteristics (Grabherr et al. 2008). Models #1 and #2 are phenomenological approximations of such features, but these models lack direct physiological implications. On the other hand, models #3 and #4 capture high-pass filter dynamics by providing a possible explanation for the measured frequency dependence of perceptual thresholds.

Though the thresholds are certainly influenced by the peripheral dynamics, we chose not to constrain the perceptual neural dynamics in models #3 and #4 to have a form identical to the periphery because the frequency range of our data is limited between 0.05 and 5 Hz. Therefore, we do not have adequate data to fit the influence of all known dynamic effects measured in animals, like peripheral lead dynamics (Goldberg and Fernandez 1971b). Furthermore, since human afferent response dynamics probably do not precisely match animal response dynamics and given the limited two-decade range of data, we chose instead to fit the simplest possible model and found that a simple first-order high-pass filter fit the data well.

Other semicircular canal responses show qualitatively similar high-pass dynamic characteristics but the cut-off frequency for other rotation responses is well below the threshold cut-off frequency (0.26 Hz). For example, the semicircular canal itself is characterized by high-pass dynamics in the velocity domain (Ifediba et al. 2007); afferent recordings in the squirrel monkey show a time constant of 5.7 s (Goldberg and Fernandez 1971a), which corresponds to a cut-off frequency of 0.03 Hz. The VOR also shows high-pass dynamic characteristics with the time constant ranging between roughly 15 and 30 s (Dimitri et al. 2001; Cohen et al. 1981; Raphan et al. 1979; Peterka et al. 1990; Leigh and Zee 1983), which corresponds to a cut-off frequency between 0.01 and 0.005 Hz. Inclusion of additional canal decay time constant and/or additional velocity storage time constant from VOR did not have a substantial impact on the fit values because our data did not extend below 0.05 Hz – the frequency range that is significantly affected by these longer time constants. Finally, and perhaps most pertinent, these threshold dynamics do not match the dynamics of human yaw rotation perception measured using magnitude estimation (Bertolini et al. 2011; Bronstein et al. 2008; Okada et al. 1999; Sinha et al. 2008), which show decay time constants between 10 and 25 s - corresponding to cut-off frequencies between 0.006 and 0.016 Hz. An earlier paper (Haburcakova et al. 2012) has suggested that this apparent discrepancy might be explained by a high-pass filter included as part of the decision making process.

Effects of bias

Effects of vestibular bias on threshold estimation will be explored in more depth in a subsequent paper, but we briefly explore the effects of bias and present the assumptions made herein. In short, the presence of a non-zero vestibular bias had only a small impact on the precision or accuracy of any of the three fit parameters when cross-frequency fits were performed. The primary explanation is that combining trials across frequency yields a large enough data set – 333 trials on average - to estimate all of the fit parameters both accurately and precisely. While detailed results are not included herein, we performed fits using the

new cross-frequency approach when vestibular biases were present. For these cross-frequency fits, we found that the accuracy and precision for all three fit parameters were about the same with and without vestibular bias. Specifically, there was no impact on the precision of σ and f . There was a small decrease in μ precision from 0.093 ($\mu = 0$) to 0.103 and 0.131 for 0.2 and 0.5 biases, respectively. The change in accuracy was less than 3% for all three parameters.

Figure 8 illustrates the flow of information from physical stimulus $\bar{\Omega}(f)$ to subject response (Left or Right). $\bar{\Omega}(f)$ is a noise-free time-varying angular velocity signal from the semicircular canals. Our threshold model attenuates this angular velocity signal (e.g., using a high-pass filter for models #3 and #4) and then determines the peak velocity of this attenuated velocity signal. Noise (“jitter”) characterized by a Gaussian distribution - having both a bias and variance – is added to this sampled peak. This is the signal utilized by the probabilistic decision processes discussed in detail throughout this paper. Specifically, this signal is compared to the subjective decision boundary, which is zero for our direction recognition task, to decide whether the motion was rightward or leftward.

The jitter variance and bias is assumed constant - independent of frequency or even type of motion. The variance represents the influence of noise accumulated along all neural pathways (including afferent noise) associated with this specific motion signal's discrimination processes. In comparison, bias accumulated before the high-pass filter would be eliminated, so bias can only accumulate after the high-pass filter. Because the bias is assumed constant after filtering, this would be observed experimentally as an increasing bias with decreasing frequency when data at individual frequencies are fit individually. This assumption is not fundamental to the models or fit methods, since models and methods would be adjusted if the assumption were shown to be incorrect, but the assumption is nonetheless clearly delineated as it would affect numerical results.

Experimental Data Fit Comparison

One goal was to reanalyze previously published (Grabherr et al. 2008) dynamic yaw rotation threshold data. The data were reanalyzed with three changes: 1) 79.4% thresholds were converted to 84.3% thresholds (one sigma); 2) thresholds were recalculated using ML method; and 3) models were fitted using the new across-frequency fitting method.

The published report suggested high-pass characteristics with a threshold plateau of 0.71°/s and a cut-off frequency of 0.23 Hz at 79.4% correct criterion set by 3-down 1-up testing procedure. The reanalysis included herein presents these thresholds as the one-sigma thresholds (84.3% correct) (Merfeld 2011), which yields a one-sigma threshold plateau of 0.86°/s.

The revised analysis performed herein similarly finds high-pass characteristics, though the fit parameters differ. Specifically, when modeled using the across frequency fit, the data yielded a) a plateau threshold value of 0.48°/s for model #1 and 0.49°/s for model #4, and b) a characteristic frequency of 0.47 Hz for model #1 and 0.26 Hz for model #4.

There are several reasons for the difference from the Grabherr et. al. (2008) results, which deserve elaboration. Part of the difference is probably due to variability simply due to the small number of data points (i.e., subjects) being averaged. Another part of the difference is that one subject was excluded from the final analysis in the published paper because a threshold could not be obtained at 0.05Hz. However, inclusion of this subject's dataset was made possible by using the across frequency fit approach, including the 0.05 Hz data. In other words, the ML fit allowed us to recover the entire data set for this subject.

Another explanation is that the data were originally analyzed for each frequency by averaging staircase minima and maxima. As shown herein – both via simulations and analytically in an appendix – such staircase threshold estimates overestimate the actual threshold. Furthermore, the earlier study used mean squared error (MSE) as the cost function, which put more weight on high thresholds, whereas the simulations performed herein used mean squared log error (MSLE), which changes the weighting. This change of cost function to better account for the lognormal distribution of thresholds has dual effects of lowering both characteristic frequency and plateau value.

Furthermore, the difference between the characteristic fit frequencies for models 1 and 4 was expected since the model fit is different. For example, when we created a transfer function model and simulated data that had a cut-off frequency of 0.26 Hz and then fit those data with model #1, the characteristic frequency converged to a value of 0.46Hz on average, which is consistent with the difference found between models #1 and #4 in the real data. In other words, both model fits converged to an appropriate fit value – even though the specific values are different due to the different shape of the different models.

This distinction deserves a little elaboration. There is a fundamental difference between a polynomial model (model #1) and a transfer function model (model #4). While bias and threshold for all four model fits can be compared directly, the characteristic frequency for the two phenomenological models (models #1 and #2) cannot be quantitatively directly related to one another, nor can the characteristic frequency from either of the phenomenological fits be quantitatively directly related to the characteristic frequency from either of the transfer function models (#3 and #4). Since models #3 and #4 both involve simulating the dynamic response of a transfer function, the characteristic frequency for these two models can be directly compared to one another and should, as observed, converge to roughly the same values when fitting data.

Potential contribution of vibration to experimentally measured thresholds

Since yaw rotation perception threshold was defined as a proportion of the physiologic noise, we decided to quantify the amount of external vibration so that we could assess the likelihood that such vibrations contribute to observed direction recognition thresholds. We recorded vibration during yaw rotation on a MOOG motion platform using a 6DOF motion sensor (MicroStrain 3DM-GX2). Motion was a 1Hz single-cycle sinusoid acceleration with a peak velocity of 1°/s, which is near threshold. Across 100 repeated measurements, noise rms (root mean square) was 0.269°/s near peak. Given that the measured electronic rms sensor noise at rest was 0.232°/s, the motion platform vibration component is 0.136°/s if we assume that the noise variances add as would be the case for uncorrelated vibration and

electrical sensor noise. This noise level is well below the observed perception noise of 0.5°/s. Therefore, it seems very unlikely that external vibrations make a predominant contribution to such vestibular direction recognition thresholds; rather the internal physiological noise appears likely to yield the predominant influence.

Algorithm choice

Our simulations suggest that the new approach of fitting simultaneously across frequencies yields more accurate, more precise, more robust, and/or more efficient parameter estimates than the traditional approach of fitting the responses for each frequency individually and then fitting across frequency. Therefore, we suggest that this approach be utilized. We also suggest continuing to perform psychometric fits at individual frequencies and plotting the resultant fitted thresholds as a function of frequency (e.g., Figure 7).

With respect to models, both a phenomenological model and a model that fully considers dynamic characteristics (e.g., a transfer function implementation) can be justified. We found that the polynomial model (# 1) yielded performance that matched the performance achieved by fitting a steady-state transfer function gain model (# 2) but offers the advantage of being more general as the polynomial model can be replaced by any function that captures a relationship between thresholds and the stimulus. We found that the new unity-gain transfer function model (model #4) yielded incrementally better information criteria scores (showing equivalent deviance with one less free parameter) than the approach of fitting a transfer function model that included a variable gain (model #3).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Appendix

Threshold estimations via traditional staircase procedure

Our simulation results showed that staircase thresholds yielded overestimation even when no bias was present ($\mu = 0$). Such overestimation of staircase thresholds is innate and cannot be avoided. A brief derivation of this fact follows. One traditional way to fit psychometric functions utilizes a z-score $z = (\omega - \hat{\mu})/\hat{\sigma}$, where ω is stimulus level and $\hat{\mu}$ and $\hat{\sigma}$ are estimations of the bias and the spread of a psychometric function ($\phi = \Phi$), respectively. Fitting a line $z = b_1 + b_2\omega$ to such z-transformed data, it can be shown (e.g., **Gescheider, 1985) that:

$$b_1 = \frac{(\sum \omega^2)(\sum z) - (\sum \omega)(\sum \omega z)}{n(\sum \omega^2) - (\sum \omega)^2} = \text{mean}(z) - \text{mean}(\omega) \frac{\text{cov}(\omega, z)}{\text{var}(\omega)}, \quad \text{equation (10-1)}$$

$$b_2 = \frac{n(\sum \omega z) - (\sum \omega)(\sum z)}{n(\sum \omega^2) - (\sum \omega)^2} = \frac{\text{cov}(\omega, z)}{\text{var}(\omega)}. \quad \text{equation (10-2)}$$

Using the above definition of the z-score and the line equation, it is straightforward to show that equals b_1 , and equals $-\hat{\mu}/\hat{\sigma}$, and b_2 equals $1/\hat{\sigma}$. Now combined with equations (10-1) and (10-2), $\hat{\mu}$ can be rewritten as follow:

$$\hat{\mu} = \text{mean}(\omega) - \text{mean}(z) \frac{\text{var}(\omega)}{\text{cov}(\omega, z)}, \quad \text{equation (11-1)}$$

Since $\text{cov}(\omega, z)$ is equal to $\text{cov}\left(\omega, \frac{\omega - \hat{\mu}}{\hat{\sigma}}\right) = \frac{1}{\hat{\sigma}} \text{var}(\omega)$, we can solve for $\hat{\sigma}$ in terms of $\hat{\mu}$, which is then

$$\hat{\sigma} = \frac{\text{mean}(\omega) - \hat{\mu}}{\text{mean}(z)}. \quad \text{equation (11-2)}$$

However, because the staircase threshold estimation procedure ignores any vestibular bias that may be present, it essentially assumes that $\hat{\mu} = 0$. Therefore, $\hat{\sigma}$ is approximated as follows by the staircase procedure

$$\hat{\sigma} = \frac{\text{mean}(\omega)}{\text{mean}(z)}. \quad \text{equation (12)}$$

This inevitably leads to overestimation on average by $\hat{\mu}/\text{mean}(z)$, where μ is the underlying vestibular bias.

Glossary: Acronyms

AICc	corrected Akaike Information Criterion
BIC	Bayesian Information Criterion
GLM	generalized linear model
ML	maximum-likelihood
MSE	mean squared error
MSLE	mean squared log error
SEM	standard error of the mean
PSE	point of subjective equality
VOR	vestibulo-ocular reflex

References

- Benson AJ, Hutt EC, Brown SF. Thresholds for the perception of whole body angular movement about a vertical axis. *Aviat Space Environ Med.* 1989; 60:205–213. [PubMed: 2712798]
- Benson AJ, Spencer MB, Stott JR. Thresholds for the detection of the direction of whole-body, linear movement in the horizontal plane. *Aviat Space Environ Med.* 1986; 57:1088–1096. [PubMed: 3790028]
- Berg A, Meyer R, Yu J. Deviance Information Criterion for Comparing Stochastic Volatility Models. *Journal of Business and Economic Statistics.* 2004; 22:107–120.

- Bertolini G, Ramat S, Laurens J, Bockisch CJ, Marti S, Straumann D, Palla A. Velocity storage contribution to vestibular self-motion perception in healthy human subjects. *J Neurophysiol.* 2011; 105:209–223. [PubMed: 21068266]
- Bronstein AM, Grunfeld EA, Faldon M, Okada T. Reduced self-motion perception in patients with midline cerebellar lesions. *Neuroreport.* 2008; 19:691–693. [PubMed: 18382289]
- Bulmer, MG. Principles of statistics. New York: Dover Publications; 1979.
- Burnham KP, Anderson DR. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research.* 2004; 33:261–304.
- Celeux, G.; Forbes, F.; Robert, CP.; Titterton, DM. Deviance information criteria for missing data models. Paris: INSEE; 2003. Centre de recherche en Économie et,statistique.
- Clark B, Graybiel A. Perception of the postural vertical in normals and subjects with labyrinthine defects. *J Exp Psychol.* 1963; 65:490–494. [PubMed: 14021495]
- Cohen B, Henn V, Raphan T, Dennett D. Velocity storage, nystagmus, and visual-vestibular interactions in humans. *Ann N Y Acad Sci.* 1981; 374:421–433. [PubMed: 6978639]
- Dimitri PS, Wall C 3rd, Rauch SD. Multivariate vestibular testing: laterality of unilateral Meniere's disease. *J Vestib Res.* 2001; 11:405–412. [PubMed: 12446965]
- Dobson, AJ.; Barnett, AG. An introduction to generalized linear models. Boca Raton: CRC Press; 2008.
- Goldberg JM, Fernandez C. Physiology of peripheral neurons innervating semicircular canals of the squirrel monkey. 3. Variations among units in their discharge properties. *J Neurophysiol.* 1971a; 34:676–684. [PubMed: 5000364]
- Goldberg JM, Fernandez C. Physiology of peripheral neurons innervating semicircular canals of the squirrel monkey. I. Resting discharge and response to constant angular accelerations. *J Neurophysiol.* 1971b; 34:635–660. [PubMed: 5000362]
- Grabherr L, Nicoucar K, Mast FW, Merfeld DM. Vestibular thresholds for yaw rotation about an earth-vertical axis as a function of frequency. *Exp Brain Res.* 2008; 186:677–681. [PubMed: 18350283]
- Graybiel A, Hupp DI, Patterson JL Jr. The law of the otolith organs. *Fed Proc.* 1946; 5:35. [PubMed: 21065950]
- Green, DM.; Swets, JA. Signal detection theory and psychophysics. New York: Wiley; 1966.
- Haburcakova C, Lewis RF, Merfeld DM. Frequency dependence of vestibuloocular reflex thresholds. *J Neurophysiol.* 2012; 107:973–983. [PubMed: 22072512]
- Hall JL. Hybrid adaptive procedure for estimation of psychometric functions. *J Acoust Soc Am.* 1981; 69:1763–1769. [PubMed: 7240589]
- Ifediba MA, Rajguru SM, Hullar TE, Rabbitt RD. The role of 3-canal biomechanics in angular motion transduction by the human vestibular labyrinth. *Ann Biomed Eng.* 2007; 35:1247–1263. [PubMed: 17377842]
- Jaynes, ET.; Bretthorst, GL. Probability theory: the logic of science. Cambridge, UK ; New York: Cambridge University Press; 2003.
- Knoblauch K, Maloney LT. Estimating classification images with generalized linear and additive models. *J Vis.* 2008; 8:10.1–10.19. [PubMed: 19146276]
- Leek MR. Adaptive procedures in psychophysical research. *Percept Psychophys.* 2001; 63:1279–1292. [PubMed: 11800457]
- Leigh RJ, Zee DS. The neurology of eye movements. 1983
- Macmillan, NA.; Creelman, CD. Detection theory : a user's guide. Mahwah, N.J.: Lawrence Erlbaum Associates; 2005.
- Mann CW, Berthelot-Berry NH, Dauterive HJ Jr. The perception of the vertical; visual and non-labyrinthine cues. *J Exp Psychol.* 1949; 39:538–547. [PubMed: 18140137]
- McCullagh, P.; Nelder, JA. Generalized linear models. Vol. 37. London ; New York: Chapman and Hall; 1989.
- Merfeld DM. Signal detection theory and vestibular thresholds: I. Basic theory and practical considerations. *Exp Brain Res.* 2011; 210:389–405. [PubMed: 21359662]

- Merfeld DM, Park S, Gianna-Poulin C, Black FO, Wood S. Vestibular perception and action employ qualitatively different mechanisms. I. Frequency response of VOR and perceptual responses during Translation and Tilt. *J Neurophysiol.* 2005a; 94:186–198. [PubMed: 15728767]
- Merfeld DM, Park S, Gianna-Poulin C, Black FO, Wood S. Vestibular perception and action employ qualitatively different mechanisms. II. VOR and perceptual responses during combined Tilt&Translation. *J Neurophysiol.* 2005b; 94:199–205. [PubMed: 15730979]
- Okada T, Grunfeld E, Shallo-Hoffmann J, Bronstein AM. Vestibular perception of angular velocity in normal subjects and in patients with congenital nystagmus. *Brain.* 1999; 122(Pt 7):1293–1303. [PubMed: 10388795]
- Peterka RJ, Black FO, Schoenhoff MB. Age-related changes in human vestibulo-ocular and optokinetic reflexes: pseudorandom rotation tests. *J Vestib Res.* 1990; 1:61–71. [PubMed: 1670138]
- Raphan T, Matsuo V, Cohen B. Velocity storage in the vestibulo-ocular reflex arc (VOR). *Exp Brain Res.* 1979; 35:229–248. [PubMed: 108122]
- Roditi RE, Crane BT. Suprathreshold asymmetries in human motion perception. *Exp Brain Res.* 2012; 219:369–379. [PubMed: 22562587]
- Sinha N, Zaher N, Shaikh AG, Lasker AG, Zee DS, Tarnutzer AA. Perception of self motion during and after passive rotation of the body around an earth-vertical axis. *Prog Brain Res.* 2008; 171:277–281. [PubMed: 18718313]
- Soyka F, Giordano PR, Barnett-Cowan M, Bulthoff HH. Modeling direction discrimination thresholds for yaw rotations around an earth-vertical axis for arbitrary motion profiles. *Exp Brain Res.* 2012; 220:89–99. [PubMed: 22623095]
- Soyka F, Robuffo Giordano P, Beykirch K, Bulthoff HH. Predicting direction detection thresholds for arbitrary translational acceleration profiles in the horizontal plane. *Exp Brain Res.* 2011; 209:95–107. [PubMed: 21234751]
- Walsh EG. Role of the vestibular apparatus in the perception of motion on a parallel swing. *J Physiol.* 1961; 155:506–513. [PubMed: 13782902]
- Wedderburn RWM. On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models. *Biometrika.* 1976; 63:27–32.
- Wichmann FA, Hill NJ. The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept Psychophys.* 2001a; 63:1293–1313. [PubMed: 11800458]
- Wichmann FA, Hill NJ. The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Percept Psychophys.* 2001b; 63:1314–1329. [PubMed: 11800459]
- Yssaad-Fesselier R, Knoblauch K. Modeling psychometric functions in R. *Behav Res Methods.* 2006; 38:28–41. [PubMed: 16817511]
- Zupan LH, Merfeld DM. Interaural self-motion linear velocity thresholds are shifted by rollvection. *Exp Brain Res.* 2008; 191:505–511. [PubMed: 18843487]

Fig. 1.

Each model includes one non-dynamic component α and one dynamic component $\beta(f; \tilde{f})$ where $\tilde{f}$ is the fitting parameter. $T(f)$ represents estimated velocity thresholds as a function of frequency. From the rearrangement of equation (2) so that $T(f) \cdot \beta(f; \tilde{f}) = \alpha$, it can be seen that $\beta(f; \tilde{f})$ is a scaling function that behaves much like a high-pass filter, scaling thresholds to a constant value. Consistent with this view, $\tilde{f}$ is a characteristic frequency that is equivalent to a cutoff frequency in a high-pass filter, where $\beta(f; \tilde{f})$ approaches 0 as f decreases, and $\beta(f; \tilde{f})$ approaches 1 for large f . This implies that thresholds at higher frequencies asymptotes to α , which represents the physiological noise level.

Fig. 2.

Example histograms for the fitted parameters for simulated results. “True” simulated values for the three free parameters are shown in the respective panel. Simulated motion direction-recognition data were fitted to the fourth model using the new (second) fitting method. The x-axis of the characteristic frequency histogram is drawn using a logarithmic scale to compensate for the right-skewness. Dashed vertical lines indicate the true values. Total sample number in each histogram is 1,000.

Fig. 3.

Simulated threshold curves obtained from the four models using the conventional fitting method. Grey solid lines indicate the true thresholds. Black solid lines indicate the average (arithmetic mean) of estimated threshold. Dotted black lines represent plus and minus one standard deviation of the estimated threshold curves. Because plots are drawn in logarithmic scale, error lines are asymmetric (shorter upper bound) with respect to the means. All prediction curves resulted in underestimation as expected from the underestimation of ML thresholds reported in the first paragraph of the results.

Fig. 4. Simulated threshold curves obtained from the first and the fourth models using the novel across-frequency fitting method. Grey solid lines with circles indicate the true thresholds. Black solid lines indicate the average (arithmetic mean) fitted threshold curves. Dotted black lines represent plus and minus one standard deviation of the fitted threshold curves. Average fitted threshold curves nearly perfectly match the true values.

Fig. 5.

Comparison of thresholds for individual subjects using different threshold estimation methods. Black squares indicate mean staircase thresholds originally reported in Grabherr et al, black circles indicate mean ML thresholds, and grey circles indicate individual subject's thresholds. Logarithmic means and standard errors of the log-means are shown.

Fig. 6. Threshold curves of the four models fitted to experimental data using the conventional fitting method. Black solid lines indicate the average (geometric mean) of estimated threshold. Dotted black lines represent plus and minus one standard log-error of the estimated threshold curves. Average ML thresholds (geometric mean, circle marker) in each subplot are also shown.

Fig. 7.

Threshold curves of the two new models fitted to experimental data using the new across-frequency approach. Black solid lines indicate the average (geometric mean) of estimated threshold. Dotted black lines represent plus and minus one standard log-error of the estimated threshold curves. Average ML thresholds (geometric mean, circle marker) in subplots a and b are identical to those in Figure 6, serving as references. Note that the mean curve lies above these reference data points. Supplementary plots of residuals from Fig 3, 4, 6, and 7 are available online (ESM1).

Fig. 8.

Decision making model underlying the model fits #3 and #4. $\bar{\hat{\Omega}}(f)$ is a noise-free time-varying angular velocity signal from the semicircular canals; $\beta(f; \tilde{f})$ is scale factor that is equivalent to high-pass filter of model fits #3 and #4; $\eta(\tilde{\mu}, \tilde{\sigma})$ is jitter that is characterized by Gaussian noise with bias $\tilde{\mu}$ and variance $\tilde{\sigma}^2$; and $\hat{\Omega}_p$ is the peak value of the noisy filtered angular velocity cue.

Table 1

Dynamic models of vestibular direction-recognition threshold. Threshold prediction $\hat{T}$ is expressed as a function of frequency f .

	Model	α	$\beta(f; \tilde{f})$	Source
#1	$T_1(\tilde{f}) = \sigma \tilde{f}$	σ	$(1 + \tilde{f}/f)^{-1}$	New
#2	$\hat{T}_2(f) = K^{-1} \sqrt{1 + (\tilde{f}/f)^2}$	K^{-1}	$[1 + (\tilde{f}/f)^2]^{-1/2}$	Grabherr et al. 2008
#3	$\hat{T}_3(f) = \frac{K^{-1}}{(\bar{\Omega}(t; f) * \text{hpf}\{\delta; \tilde{f}\})_{peak}}$	K^{-1}	$(\bar{\Omega}(t; f) * \text{hpf}\{\delta; \tilde{f}\})_{peak}$	Soyka et al. 2011
#4	$\hat{T}_4(f) = \frac{(\Omega_T(t; f) * \text{hpf}\{\delta; \tilde{f}\})_{peak}}{(\bar{\Omega}(t; f) * \text{hpf}\{\delta; \tilde{f}\})_{peak}}$	$(\Omega_T(t; f) * \text{hpf}\{\delta; \tilde{f}\})_{peak}$		New

Table 2

Three fitted parameters, scaled PSE μ and spread σ , and characteristic frequency f for each of the four dynamic models using the conventional fitting method. The true theoretical values are given in the parenthesis.

Model	$\tilde{\mu} (1^\circ / s)$		$\tilde{\sigma} (1^\circ / s)$		$f (\tilde{0.4Hz})$		
	mean	std	mean	std	mean	-std	+std
#1	0.0032	0.10	0.90	0.19	0.40	0.15	0.24
#2	0.0034	0.10	0.90	0.16	0.40	0.12	0.17
#3	0.0034	0.10	0.91	0.21	0.40	0.14	0.21
#4	0.0034	0.10	0.90	0.21	0.40	0.14	0.21

Table 3

Goodness of fit scores for the four models using the conventional fitting method. ρ is the number of free parameters, M is the number of simulations out of 1,000 sets, where the fitted characteristic frequency was between 0.1 and 5Hz. D is deviance, AIC_c is corrected Akaike Information Criterion, and BIC is Bayesian Information criterion. The mean and standard deviation (in parenthesis) are provided.

Model	ρ	M	D (std)	AIC_c (std)	BIC (std)
#1	14	978	242 (24)	272 (24)	405 (25)
#2	14	983	243 (24)	272 (24)	405 (26)
#3	14	976	242 (24)	272 (24)	405 (25)
#4	13	979	242 (24)	270 (24)	393 (25)

Table 4

Three fitted parameters, scaled PSE μ and spread σ , and characteristic frequency f for each of the two new models using the novel across-frequency fitting method. The true theoretical values are given in parenthesis.

Model	$\mu(0^\circ/s)$		$\sigma(1^\circ/s)$		$f(0.4Hz)$		
	mean	std	mean	std	mean	-std	+std
#1	0.0032	0.092	0.98	0.18	0.40	0.13	0.19
#4	0.0033	0.093	0.98	0.20	0.40	0.12	0.17

Table 5

Goodness of fit scores for the two new models fitted via across-frequency approach. Format and abbreviations are the same as in Table 3.

Model	ρ	M	D (std)	$AICc$ (std)	BIC (std)
#1	3	997	240 (21)	246 (21)	275 (21)
#4	3	998	240 (21)	246 (21)	275 (21)

Table 6

Three parameters, scaled PSE μ' and spread σ' , and characteristic frequency f' for the experimental data fitted to the four dynamic models using the conventional fitting method. Units are given in parenthesis.

Model	$\mu' (^{\circ}/s)$		$\sigma' (^{\circ}/s)$			$f' (Hz)$		
	mean	sem	mean	-sem	+sem	mean	-sem	+sem
#1	-0.004	0.037	0.38	0.057	0.067	0.50	0.12	0.16
#2	0.007	0.051	0.48	0.060	0.069	0.49	0.09	0.11
#3	-0.001	0.040	0.40	0.055	0.064	0.27	0.06	0.08
#4	0.000	0.040	0.40	0.055	0.064	0.27	0.06	0.08

Table 7

Goodness of fit scores for the two new models fitted to experimental data via conventional approach.

Model	ρ	N	D (sem)	AIC_c (sem)	BIC (sem)
#1	16	8	286 (15)	320 (15)	472 (16)
#2	16	8	287 (15)	321 (15)	473 (15)
#3	16	8	286 (15)	320 (15)	473 (16)
#4	15	8	286 (15)	318 (15)	461 (16)

Table 8

Three parameters, scaled PSE μ and spread σ , and characteristic frequency f for the two new models fitted to experimental data using the new across-frequency approach. Units are given in the parenthesis.

Model	μ ($^{\circ}$ / s)		σ ($^{\circ}$ / s)			f (Hz)		
	mean	sem	mean	-sem	+sem	mean	-sem	+sem
#1	-0.007	0.027	0.48	0.07	0.09	0.47	0.13	0.19
#4	0.006	0.028	0.49	0.07	0.09	0.26	0.07	0.10

Table 9

Goodness of fit scores for the two new models fitted to experimental data via the new across-frequency approach.

Model	ρ	N	D (sem)	$AICc$ (sem)	BIC (sem)
#1	3	8	273 (12)	279 (12)	308 (12)
#4	3	8	273(12)	279 (12)	308 (12)