Journal of Mathematical Psychology ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

A framework for Bayesian optimality of psychophysical laws

John Z. Sun [a], Grace I. Wang [a,b], Vivek K Goyal [a], Lav R. Varshney [a,c,*]

[a] Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[b] Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Boston, MA 02114, USA
[c] IBM Thomas J. Watson Research Center, Hawthorne, NY 10532, USA

ARTICLE INFO

Article history:
Received 26 March 2012
Received in revised form
26 August 2012
Available online xxxx

Keywords:
Relative Error
Bayesian Quantization
Weber–Fechner
Psychophysical Scale

ABSTRACT

The Weber–Fechner law states that perceived intensity is proportional to physical stimuli on a logarithmic scale. In this work, we formulate a Bayesian framework for the scaling of perception and find logarithmic and related scalings are optimal under expected relative error fidelity. Therefore, the Weber–Fechner law arises as being information theoretically efficient under the constraint of limited representability. An even stronger connection is drawn between the Weber–Fechner law and a Bayesian framework when neural storage or communication is the dominant concern, such as for numerosity. Theoretical results and experimental verification for perception of sound intensity are both presented.

1. Introduction

Psychophysical scales relate stimulus intensity to perceptual intensity and are central to understanding how the external world is internally represented. For example, the logarithmic scale described by the Weber–Fechner law is consistent with observations in audition, vision, haptics and numerosity (Dehaene, 2003; Fechner, 1860; Moore, 1992; Shen, 2003). Alternatives to the Weber–Fechner scaling have also been proposed, e.g. a power law (Fullerton & Cattell, 1892; Stevens, 1957, 1961). Although there is some debate over the validity of various psychophysical laws for different perceptual modalities (Krueger, 1989), many researchers suggest the psychophysical predictions of these models are essentially equivalent (Dehaene, 2003; MacKay, 1963).

A psychophysical scale is described by an increasing function $C(s)$ such that $P = C(S)$, where S and P are random variables corresponding to stimulus and perceptual intensities respectively. The Weber–Fechner law specifies $C(s)$ as $P \propto \ln(S/s_0)$, where s_0 is the threshold below which a stimulus is not perceived (making P a nonnegative quantity). Thus under the Weber–Fechner law, a multiplicative increase in stimulus intensity leads to an additive increase in perceived intensity.

Several principles have been advanced to explain psychophysical scales, but these are formulated purely at the implementational or algorithmic levels (Marr, 1982) without consideration of computational purpose. In particular, arguments based on the physical chemistry of sensory receptors (Cope, 1976) and based on the informational properties of individual neurons (MacKay, 1963) also yielded the Weber–Fechner law, but these arguments did not consider perceptual fidelity. On the other hand, Fechner (1860) solved a differential equation arising from Weber's 'just noticeable difference' experimental procedure, yielding a logarithmic scale, but did not relate it to neurobiology.

In this paper, we propose that psychophysical scales arise at the computational level as optimizations under neurobiological constraints (at the implementational and algorithmic levels). Two threads of theoretical work in neuroscience have emerged that attempt to connect physical properties and constraints of the nervous system to psychological and behavioral properties. The first argues that the physical substrate of perception in the nervous system is algorithmically well-matched to the statistical properties of the natural environment (Olshausen & Field, 2004) and that therefore operation of the brain is probabilistically efficient, i.e. Bayes-optimal (Friston, 2010; Jacobs & Kruschke, 2011; Knill & Pouget, 2004). The second thread argues that internally, brains have remarkable biophysical efficiency when performing information and communication tasks (Laughlin & Sejnowski, 2003) and therefore achieve information theoretically-optimal signal transmission (Borst & Theunissen, 1999). In both threads, design principles governing the nervous system are said to be similar to those in optimized electronic information systems.

Building on these neurobiological insights and therefore adopting the view of optimal information processing and transmission, this paper provides a mathematical framework for understanding psychophysical scales as Bayes-optimal and information

* Corresponding author at: IBM Thomas J. Watson Research Center, Hawthorne, NY 10532, USA.
E-mail address: varshney@alum.mit.edu (L.R. Varshney).

0022-2496/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2012.08.002

theoretically-optimal representations of time-sampled continuous-valued stimuli. Specifically, for statistical distributions that correspond to natural stimuli, the Weber–Fechner law and related scales minimize the expected relative error of representing the stimulus intensity under two models, each motivated by informational limitations. In the first model, each representation of a stimulus takes one of a finite set of values, corresponding to digital signaling. We also discuss an analog information system that has an equivalent formulation. The second model extends the first by allowing for compressed representations and may be more suitable when neural storage or communication has a high cost.

The discretization of a continuous scalar is called *quantization* and is well-studied under a Bayesian framework (Gray & Neuhoff, 1998). Quantization leads to stability in information representation and robustness to additive noise and mismatch while allowing for efficient reduction in communication rate or storage cost (Sarpeshkar, 1998). The choice of the discrete set representing continuous stimuli is called the quantization mapping and is optimized by minimizing some error criterion that takes into account the statistical distribution of the stimulus intensity.

In measuring quantization error for stimuli that span orders of magnitude, expected relative error (ERE) is a suitable measure of accuracy. For a particular quantization mapping Q between a random variable S and its quantized value $\hat{S}$, ERE is defined as

$$\mathrm{ERE}(Q) = \mathbf{E}\left[\frac{|S-\hat{S}|^2}{S^2}\right], \tag{1}$$

where the expectation is taken with respect to the probability density function (pdf) of S, denoted $f_S(s)$. Since ERE corresponds to squared error divided by the energy of S, the resulting error term is scale-invariant. Relative error has been used for perceptual coding in media applications (Gersho & Gray, 1992; Motta, Rizzo, & Storer, 2006). More recently, we have generalized the design of Bayes-optimal quantization mappings for ERE (Sun & Goyal, 2011).

Another quantization model was recently proposed to show that a logarithmic scaling can result when minimizing *worst-case* relative error (Portugal & Svaiter, 2011). However, there are fundamental differences between ERE and worse-case error in their psychological predictions and performance guarantees. For example, it is intuitive that a worst-case relative error criterion yields a logarithmic scaling, but we will show that the logarithmic scaling arises in a Bayesian formulation *only* when the stimulus distribution takes a certain form or when the quantized values are compressed. Moreover, worst-case quantizers generally perform poorly because they do not utilize information about the stimulus distribution and thus have lower information transmission rates. Hence, the theory presented here is more biologically plausible and provides meaningful insights regarding how the brain might exploit the distributions of natural stimuli.

One may wonder if relative error, either expected or worst-case, has psychological significance. Before making the worst-case assumption, Portugal and Svaiter (2011) motivated their work by noting that relative error is prominent in numerical analysis and physics, and hence has significant value in computations that may occur at a cognitive level. Sun and Goyal (2011) formalized the use of *expected* relative error for compression of audio and video, benefiting from decades of psychophysical studies for the purpose of matching data compression to perceptual processes in the brain (Jayant, Johnston, & Safranek, 1993). Often, perception appears to be sensitive to ratios between stimulus intensities rather than absolute stimulus intensities—the outputs of many perceptual processes appear to be independent of scale—hence relative error is the natural fidelity criterion. The quantization models proposed here can be generalized to account for other distortion measures,

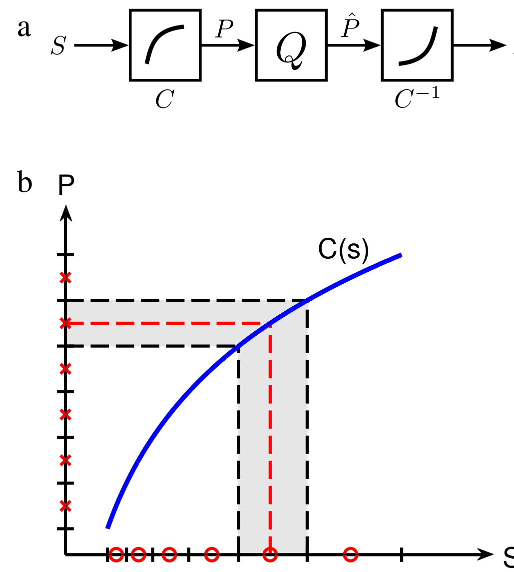

Fig. 1. (a) Block diagram of the quantization model for perception. Stimulus intensity S is transformed by a nonlinear scaling $C(s)$, resulting in perceptual intensity P. Since biological constraints limit perception, only a discrete set of levels $\{\hat{P}\}$ are distinguishable. The corresponding discrete stimulus set has elements $\hat{S} = C^{-1}(\hat{P})$. (b) The discrete mapping induced by quantization. On the vertical axis, quantization reduces the real line to a discrete set of points (indicated by crosses) called the perception dictionary. Because the scaling $C(S)$ is invertible, the quantization of perception also induces quantization in the stimulus intensity on the horizontal axis (indicated by circles). For example, any two stimulus intensities in the gray region are indistinguishable.

cf. Appendix B, and we will briefly discuss how to experimentally test the validity of error criterion assumptions.

The remainder of the paper is organized as follows. First, we introduce the two quantization models in Sections 2 and 3. Next, we present examples under which the Weber–Fechner law is Bayes-optimal in Section 4. Finally, we provide a discussion of the main results in Section 5.

2. A Bayes-optimal model for limited perception

We begin by formulating the first quantization model for perception. Like in earlier work, we abstract away neural circuitry and underlying communication mechanisms at the implementation level. By considering cognitive responses to stimuli, we show that perception fits naturally into a quantization framework.

Recall the psychophysical scale is denoted $P = C(S)$. Since perception lies on a subjective scale, we normalize the minimal and maximal perceptual intensities to be 0 and 1 respectively. For natural stimuli, the space of S is continuous. In most psychophysical studies, the space of P is assumed to have such fine resolution that it can be approximated as continuous. However, assuming that perception is limited, the brain only distinguishes a discrete set of levels. Abstractly, this means that a range of P is mapped to a single representative point $\hat{P}$ (Fig. 1). This discrete mapping, called the quantizer, is denoted $\hat{P} = Q(P)$, with output space the set of percepts $\{\hat{P}\}$, which is the *perception dictionary*. Using the invertible function $C(s)$, the equivalent representative stimulus intensities are $\hat{S} = C^{-1}(\hat{P})$.

Since the spacings between elements in the perception dictionary, called *perceptual uncertainty*, are equidistant, the only flexibility in this model is $C(s)$. The Bayesian optimization is therefore

$$\operatorname*{argmin}_{C(s)} \ \mathrm{ERE}\left(Q\left(C(S), K\right)\right), \tag{2}$$

where Q is the quantization mapping with scale $C(s)$ and perception dictionary size K. For a given stimulus pdf $f_S(s)$, what choice of $C(s)$ minimizes the ERE between S and $\hat{S}$? How does the size of the set of $\hat{P}$ affect this choice?

First we study how the stimulus distribution affects $C(s)$. If $f_S(s)$ is bounded such that $0 < s_0 \leq S \leq s_1 < \infty$, where s_0 and s_1 are constants, then according to quantization theory, to minimize ERE, the optimal relationship between $C(s)$ and $f_S(s)$ satisfies

$$\frac{dC(s)}{ds} \propto s^{-2/3} f_S^{1/3}(s) \quad \text{for } s \in [s_0, s_1], \tag{3}$$

with constraints $C(s_0) = 0$ and $C(s_1) = 1$. The proof of (3) relies on certain approximations for large dictionary sizes and uses calculus to show optimality (Sun & Goyal, 2011); see also Appendix B. We emphasize that (3) holds approximately even for dictionaries of moderate size.

Next, we address how the size of the perception dictionary affects $C(s)$ and ERE. If the set $\{\hat{P}\}$ has K equally-spaced elements between 0 and 1, then

$$\hat{P}_k = \frac{k - 1/2}{K}, \quad k = 1, 2, \ldots, K. \tag{4}$$

The equivalent stimulus representation is simply $\hat{S}_k = C^{-1}(\hat{P}_k)$. As K increases, so does the resolution of perception, leading to a more accurate approximation of stimulus intensity and reducing the squared error factor in ERE. As mentioned above, the optimization of $C(s)$ is unaffected by K in the limit of large K, and the convergence is fast enough such that the asymptotic scaling can be considered with little loss of optimality.

Therefore, $C(s)$ is heavily dependent on $f_S(s)$ and less dependent on K. Moreover, $C(s)$ is also explicitly dependent on the error measure to be minimized, which is ERE. Indeed, the optimal scale $C(s)$, stimulus pdf $f_S(s)$, and error measure are intertwined under quantization theory such that knowledge of any two can predict the third.

That the psychophysical scale $C(s)$ adapts to the statistical properties of the stimulus pdf $f_S(s)$ implies that $C(s)$ should change when $f_S(s)$ changes. Such plasticity would allow individuals to adapt to long-term changes in their perceptual environment. For example, this phenomenon has been observed for perception of sound intensity in individuals after long-term use of auditory prostheses that modify $f_S(s)$ (Philibert, Collet, Vesson, & Veuillet, 2002; Thai-Van, Philibert, Veuillet, & Collet, 2009).

We briefly mention an analog model for limited perception; a more formal discussion is given in Appendix A. This analog model leads to the same expression for the psychophysical scale that minimizes ERE, making the conclusions we draw from (3) not contingent on discretization in the encoding of stimuli. Suppose that rather than quantization, zero-mean additive noise corrupts P, leading to $\hat{P}$ (Fig. 2). If this additive noise is independent of the stimulus intensity and has variance of the same order as the perceptual uncertainty, then the Bayes-optimal $C(s)$ has the same form as (3) in the limit of high signal-to-noise ratio (SNR). Note that the notion of SNR here is at the cognitive level, which may be high even when the SNRs of single neurons are substantially lower.

3. A Bayes-optimal model for limited perception with coding

We now formulate the second quantization model, motivated by stimuli for which psychophysical scales are not affected by sensing mechanisms or communication channels from sensory systems. One such example is numerosity, which has been shown experimentally to follow the Weber–Fechner law (Dehaene, 2003; Nieder & Miller, 2003). As an abstract sensation, numerosity is of particular interest since it does not suffer from physical limitations

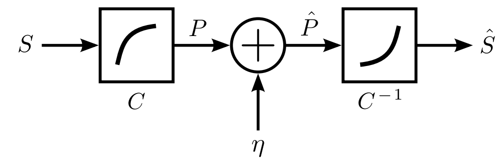

Fig. 2. Block diagram of an analog model for perception that provides equivalent asymptotic results as the model in Fig. 1. Stimulus intensity S is transformed by a nonlinear scaling $C(s)$, resulting in perceptual intensity P. Biological constraints are modeled by limiting the range of $C(s)$ to [0, 1] and by including additive noise η that is independent of S.

like resolution or saturation. Note that small numbers may be qualitatively different in how they are perceived (Le Corre & Carey, 2007). Since numbers can be observed directly with very fine precision, why should numerical perception suffer from the indistinguishability of quantization?

The reason may be *coding*. In the previous model, the quantized values were not represented more efficiently through coding. However, representing more likely values in the perception dictionary with compact representations leads to reduced overall information transmission or storage at the expense of increased computational complexity and delay. An information-theoretic concept called entropy gives the fundamental limits of compressibility through coding (Cover & Thomas, 1991). Efficient entropy-based codes have been suggested for transmission of sensory information (Fairhall, Lewen, Bialek, & de Ruyter van Steveninck, 2001; Ganguli & Simoncelli, 2010), for short-term memory (Brady, Konkle, & Alvarez, 2009), and in the context of learning (Barlow, 1989), where compressed representations may help meet memory capacity constraints (Varshney, Sjöström, & Chklovskii, 2006).

In *variable-rate* quantization, values from the discrete set $\{\hat{P}\}$ (or equivalently, the set $\{\hat{S}\}$) are entropy-coded based on the probabilities of occurrence of the entries in the set (Fig. 3). As the result of coding, the best choice of $C(S)$ is no longer the same as in the previous model (Gish & Pierce, 1968). The Bayesian optimization is now

$$\underset{C(s)}{\operatorname{argmin}} \ \mathrm{ERE}\left(Q(C(S), K)\right) \quad \text{such that } H(Q(S)) < R, \tag{5}$$

where R is the communication/storage rate and $H(\cdot)$ is the entropy function.[1] For a random stimulus S bounded as $0 < s_0 \leq S \leq s_1 < \infty$, the optimal scale for ERE has the property

$$\frac{dC(s)}{ds} \propto 1/s \quad \text{for } s \in [s_0, s_1], \tag{6}$$

with additional normalization constraints $C(s_0) = 0$ and $C(s_1) = 1$ (Sun & Goyal, 2011); see also Appendix B. Unlike in (3), there is no dependence on the pdf $f_S(s)$ in (6). The scaling that satisfies the above conditions is

$$C(s) = \frac{\ln(s/s_0)}{\ln(s_1/s_0)} \quad \text{if } s \in [s_0, s_1], \tag{7}$$

for any pdf $f_S(s)$. In fact, the scale is only dependent on the endpoint values of the distribution. For unbounded random variables, an analogous result holds under mild conditions on the decay of the tail of the pdf. Hence, quantization followed by efficient coding leads naturally to the logarithmic relationship in the Weber–Fechner law. This holds for all well-behaved stimulus distributions and the resolution can be tuned easily by the level of compression.

We suggest that coding is an essential part of perceiving numbers, which need not have distributions of a particular form.

[1] For a discrete random variable X with probability mass function $f_X(x)$, the entropy function is defined as $H(X) = -\sum_x f_X(x) \log f_X(x)$. Since the output of a quantizer Q can only take a discrete set of values, its output is a discrete random variable if the input S is random.

Please cite this article in press as: Sun, J. Z., et al. A framework for Bayesian optimality of psychophysical laws. Journal of Mathematical Psychology (2012), doi:10.1016/j.jmp.2012.08.002

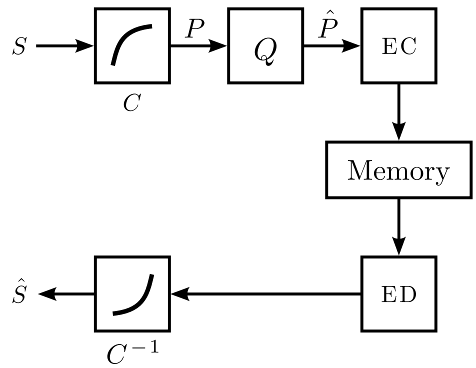

Fig. 3. Block diagram of the quantization model with coding. An entropy coder (EC) is used to code quantized values $\hat{P}$ and is then stored into memory. Later, the perception is fetched from memory, decoded using an entropy decoder (ED). The EC and ED steps include no noise, and no information is lost in these steps.

Furthermore, since the optimal psychophysical scales for entropy-coded representations do not depend on the statistical properties of the source, there is task independence and constancy (Ganguli & Simoncelli, 2010).

4. Examples

In this section, we connect the proposed Bayesian quantization models to the Weber–Fechner law. This relationship is clear in the second model since the logarithmic scale is optimal for all stimulus distributions, as shown in (7). However, in the first model, the optimized scale depends explicitly on the stimulus distribution through (3).

Thus, it is not obvious that the Weber–Fechner law will be realized using the first model. However, the sensations corresponding to many natural phenomena have statistical distributions that obey a power law over a range of intensities that are of behavioral interest (Mandelbrot, 1982; Rocchesso & Fontana, 2003; Zipf, 1949), and we will demonstrate that such distributions do yield a logarithmic scale. As a case study, we present empirical evidence that the computed scalings of natural sounds are well approximated by the Weber–Fechner law.

4.1. Power-law stimulus distributions

We begin by developing psychophysical scales for the Pareto distribution. Given parameters $\alpha > 0$ and $s_0 > 0$, the Pareto pdf is

$$f_S(s) = \frac{\alpha}{s_0}\left(\frac{s}{s_0}\right)^{-\alpha-1} \quad \text{if } s \geq s_0\text{; and 0 otherwise,} \tag{8}$$

where s_0 corresponds to the lowest perceivable stimulus intensity. The pdf decays with an exponent of $-(\alpha+1)$ and intensity is not upper-bounded, i.e. there is no upper threshold to perception.

Using (3), the derivative of $C(s)$ is

$$\frac{dC(s)}{ds} = \frac{\alpha}{3s_0}\left(\frac{s}{s_0}\right)^{-\alpha/3-1} \quad \text{if } s \geq s_0\text{; and 0 otherwise.} \tag{9}$$

The psychophysical scale that results from the above expression and satisfies the boundary conditions is

$$C(s) = 1 - \left(\frac{s}{s_0}\right)^{-\alpha/3} \quad \text{if } s \geq s_0\,. \tag{10}$$

In general, the psychophysical scales generated by (10) are concave on a logarithmic scale and hence are inconsistent with the Weber–Fechner law. However, a bounded pdf is more practical because there are usually lower and upper limits to what is perceivable. With parameters α, s_0 and s_1, the bounded power-law distribution is

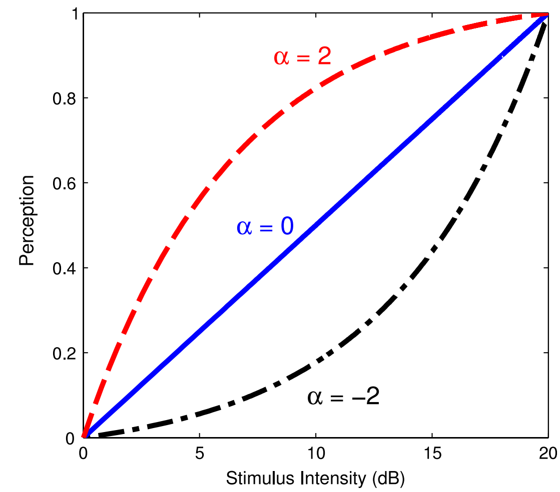

Fig. 4. Sample scaling $C(s)$ for bounded power-law densities with three choices for α. Perception intensity on the vertical axis is normalized to lie between 0 and 1, corresponding to the smallest and largest perceivable stimulus intensities, respectively. Logarithmic scaling results when $\alpha = 0$.

$$f_S(s) \propto s^{-\alpha-1} \quad \text{if } s \in [s_0, s_1]\text{; and 0 otherwise,} \tag{11}$$

normalized to have unit integral. Here, s_0 and s_1 are the lower and upper thresholds of perception, yielding a more psychophysically reasonable model. Note that α is no longer restricted to be positive as in (8). Repeating the same analysis as above, the derivative of $C(s)$ is

$$\frac{dC(s)}{ds} \propto s^{-\alpha/3-1} \quad \text{if } s \in [s_0, s_1]\text{; and 0 otherwise.} \tag{12}$$

For the special case of $\alpha = 0$, or equivalently an exponent of -1 in the decay of the pdf, (12) simplifies to

$$C(s) = \frac{\ln(s/s_0)}{\ln(s_1/s_0)} \quad \text{if } s \in [s_0, s_1], \tag{13}$$

which is precisely the Weber–Fechner law. For other choices of α, the scaling is

$$C(s) = \frac{s^{-\alpha/3} - s_0^{-\alpha/3}}{s_1^{-\alpha/3} - s_0^{-\alpha/3}} \quad \text{if } s \in [s_0, s_1]\,, \tag{14}$$

providing a generalization to logarithmic scaling accounting for a large class of scales. Fig. 4 demonstrates how three choices of α affect the Bayes-optimal $C(s)$.

Thus, there is an intimate match between the Weber–Fechner law and a bounded power-law distribution. Indeed, such a distribution with $\alpha = 0$ matches precisely with logarithmic scaling. However, other exponents yield minor deviations which may also be observed experimentally.

4.2. Natural sounds

The above results predict experimentally falsifiable psychophysical scales based on power-law stimulus distributions. In general, while natural stimuli may not be easily identified as exactly power-law distributed, many are approximately power-law over relevant ranges. One such example is the intensity of speech, which is often modeled as Gaussian-distributed on a dB scale (lognormal). Indeed, lognormal and power-law distributions are often empirically indistinguishable (Mitzenmacher, 2004). We test datasets of animal vocalizations and human speech and find the optimal psychophysical scale to be well-approximated by a logarithmic relationship where the intensity is most probable.

Animal vocalizations and human speech comprise complex harmonic and transient components that convey behaviorally-relevant meaning. For example, many animals vocalize to convey information related to mating rituals, predator warnings, or the locations of food sources. The individuals that best process these sounds may be those that are most likely to survive and reproduce. In this way, the auditory system may have evolved to optimally process natural sounds like vocalizations in order to efficiently extract relevant acoustic cues. One such cue is the perceived intensity of a sound, or its loudness. In particular, the normal human ear perceives sound intensities with roughly 1 dB JND across nearly the entire range of perceivable levels (Fay, 1988), with only slight variation near the extremes, which is consistent with the logarithmic relationship in the Weber–Fechner law (Florentine, Buus, & Mason, 1987; Viemeister, 1983).

We employ two datasets comprising animal vocalizations and human speech sounds. The animal vocalization data (DS1) includes 55 rain forest mammals (33 min) taken from commercially available CDs (Emmons, Whitney, & Ross, 1997). The human speech data (DS2) corresponds to a male speaker reciting a corpus of 280 English sentences (8 min) (Wen, Wang, Dean, & Delgutte, 2009).

Silence intervals, defined as intervals of 50 ms in which the signal did not exceed 10% of the maximum magnitude of the recording, were removed from the recordings. The resulting sound files were broken into successive intervals of 100 ms, and the root mean square (rms) was computed for each interval. The empirical sound level distributions of the rms values were used to compute $C(s)$.

For both DS1 and DS2, the predicted psychophysical scales are well-approximated by a straight line where the intensity is most probable (Fig. 5); since the horizontal axis is logarithmic, this indicates a logarithmic relationship. Moreover, the deviation from a logarithmic scaling is most prominent at the extremes of the stimulus pdf, where experimental studies also demonstrate breakdown of the Weber–Fechner law (Atkinson, 1982; McBride, 1983).

We also varied several parameters that affect the empirical distribution. These parameters included the thresholds and interval lengths used to calculate the silence intervals, lengths of the interval over which the rms values were computed, and histogram bin widths. To account for gain control mechanisms, we also used an rms gain parameter to horizontally shift the empirical distribution. We found that the scales induced by these parameter changes did not vary greatly and had similar goodness-of-fit characteristics on a linear regression.

To summarize, sound intensity perception scales determined from animal vocalization data and our optimality principles are consistent with the basic Weber–Fechner law.

5. Discussion

Through quantization frameworks for perception, we have determined that scaling laws observed in psychophysical studies are Bayes-optimal for expected relative error. Sensations that are measured by sensory mechanisms in the periphery and ones that are abstract are both considered. Although they have different costs, both situations have optimized scalings that include the Weber–Fechner law. Moreover, other scaling laws may be consistent under the first model, and this theory provides an experimentally testable hypothesis for them.

There are several key assumptions that anchor this framework. The first and most fundamental assumption is that the acquisition of information in the stimulus is Bayes-optimal at a computational level (Marr, 1982). This is well-motivated since numerous studies suggest that there exist neural mechanisms for adaptation to

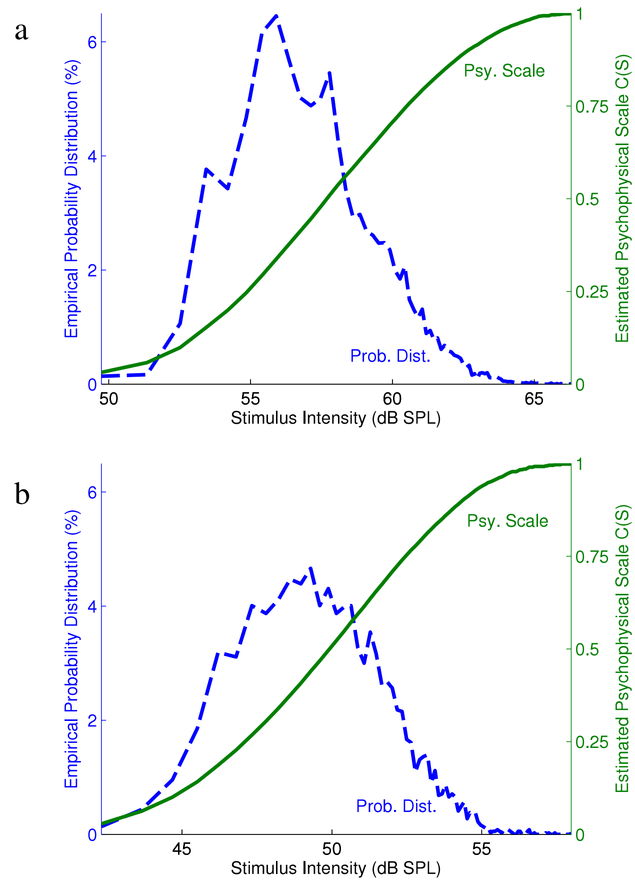

Fig. 5. Empirical distribution of stimulus intensities (blue dashed line, left axis) and corresponding psychophysical scaling function, based on (3) (green solid line, right axis) for (a) mammalian vocalizations (DS1) and (b) human speech (DS2). The near linearity of the scaling functions in log scale indicates an approximation to the Weber–Fechner law. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the stimulus distribution as well as numerous feedback and feedforward channels (Lamme, Supèr, & Spekreijse, 1998; Wen et al., 2009).

The second assumption is that perception is algorithmically discrete and the mapping is simply a deterministic function of the stimulus intensity, as illustrated in Fig. 1. This framework is inspired by engineering systems and information-theoretic results on Bayes-optimal data acquisition. Although it is debatable whether this type of discretization occurs in individual neurons, quantization is plausible at the cognitive level. Moreover, we have described how equivalent asymptotic scalings occur with an analog coding scheme. In the context of this framework, the adaptation mechanisms discussed above better estimate the distribution shape and thresholds to precisely predict $C(s)$.

A third assumption is that the accuracy measure to minimize is ERE. We have motivated this choice through related research in numerical computation and perceptual coding, as well as the general perceptual principle of scale-invariance. However, ERE may be too simple to fully capture the goals of information acquisition and may provide only a crude first-order approximation. In Appendix B, we show that the psychophysical implications of this framework are robust within a large class of error criteria. More generally, the true cost may be a combination of several error measures weighted differently depending on behavioral context. In this case, our framework is still useful since the error measure, stimulus distribution and optimizing psychophysical scale are interrelated such that the observation of any two predicts the third.

Please cite this article in press as: Sun, J. Z., et al. A framework for Bayesian optimality of psychophysical laws. Journal of Mathematical Psychology (2012), doi:10.1016/j.jmp.2012.08.002

Therefore one can predict the error measure optimized by the system using the experimentally observed stimulus distribution and psychophysical scale.

Finally, although the analysis and optimality results presented apply to any stimulus signal that is stationary in time (see Theorems 1 and 2), the scalar system model does not exploit the memory or correlation in time of natural stimuli. Our system model takes acquisition to be independent across intensity samples. Indeed, one could code over blocks of samples or apply a causal filter to improve compression, which may change the optimal choice of $C(s)$ (Gray & Neuhoff, 1998). However, such coding will increase signal acquisition delay. For external stimuli such as sound or vision, speed of perception may be more important than improved compression. For this reason, the models presented here are psychologically reasonable under latency constraints.

An interesting question is whether either of the two quantization models proposed here can be applied to the stimulus class of the other. For example, can coding occur for stimuli sensed by the periphery? Alternatively, many numerical quantities, such as salaries or populations have been proposed to approximately follow Pareto distributions. Could numerosity scaling laws be logarithmic without the need for coding? We feel affirmative answers to these questions are less plausible than what is proposed here but they remain topics of worthwhile exploration.

In conclusion, we show that the Weber–Fechner law arises as the Bayes-optimal mapping in two quantization frameworks. In the first model, when stimulus intensity is simply discretized, Weber–Fechner is intimately tied to stimulus distributions that decay as power-law functions. In the second, when discrete percepts are coded, the Weber–Fechner law becomes more general and is optimal for *all* statistical distributions. These results are dependent on several assumptions which are well-supported by neuroscientific and information scientific principles. This work points out at least three psychophysical ideas: the importance of stimulus statistics in interpreting scaling laws, the necessity of adaptation to stimulus distributions in neural circuitry, and the possibility of information theoretically-optimal acquisition structures at a cognitive level.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants 0643836 and 1115159. The authors thank the reviewers, the associate editor, Bo Wen, Dmitri B. Chklovskii, Barbara Shinn-Cunningham, Katherine Heller, and Jeff Beck for their insightful comments.

Appendix A. Mathematical methods for analog model

In Section 2, we introduced a quantization model (QM) and determined the psychophysical scale that minimizes ERE for a given stimulus distribution. Here, we summarize an analog model (AM) based on suppression of additive noise and relate the optimization of the scale in the two models (Fig. 2). Most of the details below on QM were originally derived in Sun and Goyal (2011); the parts related to AM are new.

We first give some insight into how (3) was derived. We use a concept called the *point density function* $\lambda(s)$, defined to be the derivative of the scale:

$$\lambda(s) = C'(s). \tag{A.1}$$

Since $C(s)$ is strictly monotonic and has specified boundary conditions, $C(s)$ and $\lambda(s)$ have a one-to-one correspondence. One interpretation of the point density for large K is that $\lambda(s)\delta$ is the approximate fraction of perception dictionary entries in a range centered at s with width δ. The point density function facilitates understanding of the dependence of ERE on $C(s)$. For stimulus distribution $f_S(s)$ and quantizer Q with point density $\lambda(s)$ and dictionary size K, the ERE satisfies

$$\lim_{K\to\infty} \mathrm{ERE}(Q)\cdot K^2 = \frac{1}{12}\,\mathbf{E}[S^{-2}\lambda^{-2}(S)]. \tag{A.2}$$

Thus, ERE decreases approximately as the square of the increase in dictionary size, and this approximation becomes more precise as K increases. Experimental results in quantization theory suggest that (A.2) is accurate even for moderate values of K (Neuhoff, 1993). The expectation on the right-hand side of (A.2) is a simple function of the point density and stimulus distribution; one can optimize over $\lambda(s)$ to find the asymptotically best scaling $C(s)$ (Gray & Neuhoff, 1998). Direct application of Hölder's inequality yields (3).

Now consider the error arising from AM when η is bounded noise independent of the stimulus and has variance $\sigma^2(\eta)$. The noisy stimulus intensities take the form $\hat{S} = C^{-1}(C(S) + \eta)$. Because $C(s)$ is continuous and strictly monotonic (hence differentiable), we use Taylor's theorem to describe $\hat{S}$ through a linear approximation. Taylor's theorem states a function $g(x)$ that is $n+1$ times continuously differentiable on a closed interval $[a, x]$ takes the form

$$g(x) = g(a) + \left\{\sum_{k=1}^{n} \frac{g^{(k)}(a)}{k!}(x-a)^k\right\} + R_n(x, a), \tag{A.3}$$

with a Taylor remainder term

$$R_n(x, a) = \frac{g^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1} \tag{A.4}$$

for some $\xi \in [a, x]$. Therefore, $\hat{S} \approx W(C(S)) + W'(C(S))\eta$ for small η, where $W(s) = C^{-1}(s)$. By the definition of $W(s)$, $W(C(s)) = s$ and $W'(C(s)) = 1/C'(s)$, so we can simplify the above equation to $\hat{S} \approx S + \eta/\lambda(S)$.

Using (1) while noting the expectation is taken with respect to both S and η, the ERE in the AM is

$$\mathrm{ERE} \approx \mathbf{E}[S^{-2}\lambda^{-2}(S)]\,\sigma^2(\eta), \tag{A.5}$$

with approximation error corresponding to the first-order Taylor remainder term. Since the remainder decays as η^2 while the ERE decays as η, we can be more mathematically precise about the behavior of the ERE.

Theorem 1. *Consider a stimulus intensity S following a stationary random process with probability density $f_S(s)$ that is smooth and positive on $\mathbb{R}$ (or a compact interval). The intensity is scaled through function $C(s)$ and then perturbed by independent additive noise η that is bounded and has variance $\sigma^2(\eta)$. The expected relative error between the stimulus and its noisy version satisfies*

$$\lim_{\sigma^2(\eta)\to 0} \mathrm{ERE}\cdot\sigma^2(\eta) = \mathbf{E}[S^{-2}\lambda^{-2}(S)]. \tag{A.6}$$

Noting that the right-hand sides of both (A.2) and (A.6) have the same form, the choice of $\lambda(s)$ that minimizes ERE is the same for both. Hence, QM and AM have equivalent asymptotic psychophysical scales.

Appendix B. Robustness of the ERE criterion

In (2) and (5), the optimal psychophysical scales are found with respect to the expected relative error criterion as defined in (1). In this supplementary discussion, we investigate the sensitivity of the psychophysical implications to the choice of error criterion. In particular, we find that similar optimal psychophysical scales arise for a wide class of error criteria, demonstrating the Bayesian framework is robust to this choice.

For a particular quantization mapping Q between a random variable S and its quantized value $\hat{S}$, we define the *rth-power expected relative error* to be

$$\mathrm{rERE}(Q) = \mathbf{E}\left[\frac{|S-\hat{S}|^r}{S^r}\right], \tag{B.1}$$

where the expectation is taken with respect to S. Using similar analysis to the derivation of (A.2) and the theorems of Cambanis and Gerr (1983), we can show the following:

Theorem 2. *Consider a stimulus intensity S following a stationary random process with probability density $f_S(s)$ that is smooth and positive on $\mathbb{R}$ (or a compact interval). The intensity is scaled through function $C(s)$ and then discretized using a scalar quantizer Q defined by a point density $\lambda(s)$ and codebook size K. The rth-power expected relative error between the stimulus and its quantized version satisfies*

$$\lim_{K\to\infty} \mathrm{rERE}(Q)\cdot K^r = \frac{1}{(r+1)2^r}\,\mathbf{E}[S^{-r}\lambda^{-r}(S)]. \tag{B.2}$$

Utilizing Hölder's inequality, we can show the solution to

$$\operatorname*{argmin}_{C(s)}\ \mathrm{rERE}\left(Q\left(C(S),K\right)\right), \tag{B.3}$$

satisfies

$$\frac{dC(s)}{ds} \propto s^{-r/(r+1)} f_S^{1/(r+1)}(s) \quad \text{for } s\in[s_0,s_1] \tag{B.4}$$

for K large if S bounded as $0 < s_0 \le S \le s_1 < \infty$. In Section 4.1, we considered the best scale for the class of bounded power-law distributions. In contrast to (12), the optimal scale has the property

$$\frac{dC(s)}{ds} \propto s^{-\alpha/(r+1)-1} \quad \text{if } s\in[s_0,s_1]; \text{ and 0 otherwise.} \tag{B.5}$$

However, we can see that the choice of $\alpha = 0$, corresponding to $f_S(s) \propto s^{-1}$ still yields the logarithmic scale. Hence, this distribution is invariant to the choice of r.

Similarly, for a random variable S bounded as $0 < s_0 \le S \le s_1 < \infty$, the solution to

$$\operatorname*{argmin}_{C(s)}\ \mathrm{rERE}\left(Q\left(C(S),K\right)\right) \quad \text{such that } H(Q(S)) < R, \tag{B.6}$$

satisfies

$$\frac{dC(s)}{ds} \propto 1/s \quad \text{for } s\in[s_0,s_1] \tag{B.7}$$

for K large just like (6). This suggests that when entropy coding is allowed, the optimal psychophysical scale is again logarithmic regardless of r.

References

Atkinson, W. H. (1982). A general equation for sensory magnitude. *Perception & Psychophysics*, *31*, 26–40.
Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, *1*, 295–311.
Borst, A., & Theunissen, F. E. (1999). Information theory and neural coding. *Nature Neuroscience*, *2*, 947–957.
Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, *138*, 487–502.
Cambanis, S., & Gerr, N. L. (1983). A simple class of asymptotically optimal quantizers. *IEEE Transactions on Information Theory*, *IT-29*, 664–676.
Cope, F. W. (1976). Derivation of the Weber–Fechner law and the Loewenstein equation as the steady-state response of an Elovich solid state biological system. *Bulletin of Mathematical Biology*, *38*, 111–118.
Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: John Wiley & Sons.
Dehaene, S. (2003). The neural basis of the Weber–Fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences*, *7*, 145–147.
Emmons, L. H., Whitney, B. M., & Ross, D. L., Jr. (1997). *Sounds of neotropical rainforest mammals*. Ithaca, NY: Cornell Laboratory of Ornithology.
Fairhall, A. L., Lewen, G. D., Bialek, W., & de Ruyter van Steveninck, R. R. (2001). Efficiency and ambiguity in an adaptive neural code. *Nature*, *412*, 787–792.
Fay, R. R. (1988). *Hearing in vertebrates: a psychophysics databook*. Hill-Fay Associates.
Fechner, G. T. (1860). *Elemente der psychophysik*. Leipzig: Breitkopf & Härtel.
Florentine, M., Buus, S., & Mason, C. R. (1987). Level discrimination as a function of level for tones from 0.25 to 16 kHz. *Journal of the Acoustical Society of America*, *81*, 1528–1541.
Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138.
Fullerton, G. S., & Cattell, J. M. (1892). *On the perception of small differences*. Philadelphia, PA: U. Penn Press.
Ganguli, D., & Simoncelli, E. P. (2010). Implicit encoding of prior probabilities in optimal neural populations. In *Adv. neural inform. proc. sys.* Vancouver, Canada (pp. 658–666).
Gersho, A., & Gray, R. M. (1992). *Vector quantization and signal compression*. Boston, MA: Kluwer Acad. Pub.
Gish, H., & Pierce, J. P. (1968). Asymptotically efficient quantizing. *IEEE Transactions on Information Theory*, *IT-14*, 676–683.
Gray, R. M., & Neuhoff, D. L. (1998). Quantization. *IEEE Transactions on Information Theory*, *44*, 2325–2383.
Jacobs, R. A., & Kruschke, J. K. (2011). Bayesian learning theory applied to human cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 8–21.
Jayant, N., Johnston, J., & Safranek, R. (1993). Signal compression based on models of human perception. *Proceedings of the IEEE*, *81*, 1385–1422.
Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*, 712–719.
Krueger, L. E. (1989). Reconciling Fechner and Stevens: toward a unified psychophysical law. *Journal of Behavioral and Brain Science*, *12*, 251–267.
Lamme, V. A. F., Supèr, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology*, *8*, 529–535.
Laughlin, S. B., & Sejnowski, T. J. (2003). Communication in neuronal networks. *Science*, *301*, 1870–1874.
Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: an investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*, 395–438.
MacKay, D. M. (1963). Psychophysics of perceived intensity: a theoretical basis for Fechner's and Stevens' laws. *Science*, *139*, 1213–1216.
Mandelbrot, B. B. (1982). *The fractal geometry of nature*. New York, NY: W.H. Freeman.
Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. New York, NY: Freeman.
McBride, R. L. (1983). Category scales of sweetness are consistent with sweetness—matching data. *Perception & Psychophysics*, *34*, 175–179.
Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, *1*, 226–251.
Moore, B. C. J. (1992). *An introduction to the psychology of hearing*. San Diego, CA: Academic Press.
Motta, G., Rizzo, F., & Storer, J. A. (2006). *Hyperspectral data compression*. New York, NY: Springer.
Neuhoff, D. L. (1993). The other asymptotic theory of lossy source coding. In R. Calderbank, G. D. Forney, Jr., & N. Moayeri (Eds.), *DIMACS series in discrete mathematics and theoretical computer science*: *vol. 14*. *Coding and quantization* (pp. 55–65). American Mathematical Society.
Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, *37*, 149–157.
Olshausen, B. A., & Field, D. J. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, *14*, 481–487.
Philibert, B., Collet, L., Vesson, J.-F., & Veuillet, E. (2002). Intensity-related performances are modified by long-term hearing aid use: a functional plasticity? *Hearing Research*, *165*, 142–151.
Portugal, R. D., & Svaiter, B. F. (2011). Weber–Fechner law and the optimality of the logarithmic scale. *Minds & Machines*, *21*, 73–81.
Rocchesso, D., & Fontana, F. (2003). *The sounding object*. Firenze: Mondo Estremo.
Sarpeshkar, R. (1998). Analog versus digital: extrapolating from electronics to neurobiology. *Neural Computation*, *10*, 1601–1638.
Shen, J. (2003). On the foundations of vision modeling: I. Weber's law and Weberized TV restoration. *Physica D: Nonlinear Phenomena*, *175*, 241–251.
Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*, 153–181.
Stevens, S. S. (1961). To honor Fechner and repeal his law. *Science*, *133*, 80–86.
Sun, J. Z., & Goyal, V. K. (2011). Scalar quantization for relative error. In *Proc. IEEE data compression conf.* Snowbird, UT (pp. 293–302).
Thai-Van, H., Philibert, B., Veuillet, E., & Collet, L. (2009). Assessment of auditory plasticity using psychoacoustic and electrophysiological measurements. *Audiological Medicine*, *7*, 55–66.
Varshney, L. R., Sjöström, P. J., & Chklovskii, D. B. (2006). Optimal information storage in noisy synapses under resource constraints. *Neuron*, *52*, 409–423.
Viemeister, N. F. (1983). Auditory intensity discrimination at high frequencies in the presence of noise. *Science*, *221*, 1206–1208.
Wen, B., Wang, G. I., Dean, I., & Delgutte, B. (2009). Dynamic range adaptation to sound level statistics in the auditory nerve. *The Journal of Neuroscience*, *29*, 13797–13808.
Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.

Please cite this article in press as: Sun, J. Z., et al. A framework for Bayesian optimality of psychophysical laws. Journal of Mathematical Psychology (2012), doi:10.1016/j.jmp.2012.08.002