

Application of Psychophysical Techniques to Haptic Research

Lynette A. Jones, *Senior Member, IEEE*, and Hong Z. Tan, *Senior Member, IEEE*

Abstract—Various psychophysical methods have been used to study human haptic perception, although the selection of a particular method is often based on convention, rather than an analysis of which technique is optimal for the question being addressed. In this review, classical psychophysical techniques used to measure sensory thresholds are described as well as more modern methods such as adaptive procedures and those associated with signal detection theory. Details are provided as to how these techniques should be implemented to measure absolute and difference thresholds and factors that influence subjects' responses are noted. In addition to the methods used to measure sensory thresholds, the techniques available for measuring the perception of suprathreshold stimuli are presented. These scaling methods are reviewed in the context of the various stimulus and response biases that influence how subjects respond to stimuli. The importance of understanding the factors that influence perceptual processing is highlighted throughout the review with reference to experimental studies of haptic perception.

Index Terms—Evaluation/methodology, haptic I/O, input devices and strategies, user interfaces

1 INTRODUCTION

THE discipline of psychophysics is focused on determining the relation between physical stimuli and the perception of those stimuli. The development of techniques that enable the measurement of sensory processes and of statistical models that characterize the performance of the human operator has therefore been an essential part of psychophysical research. In addition, psychophysical techniques have been used to analyze the capacities of different sensory modalities, to study the underlying biological mechanisms that account for these sensory abilities and to understand the processes involved in human decision making. The distinction between a focus on linking sensory magnitudes to stimulus magnitudes in psychophysical research and on understanding the transformation of the input that impinges on the peripheral sense organs and is transmitted to the cerebral cortex is one that dates back to the 19th century and is referred to as outer and inner psychophysics, respectively.

Any sensation that the human organism responds to can vary along at least four basic dimensions—intensity, quality, extension, and duration—each of which can be studied experimentally. Quality refers to the fact that different forms of stimulation evoke different sensations, and so, for example, light falling on the retina is perceptually a very different experience from mechanical stimulation of the skin. Within a sensory modality sensations may

also vary in quality such as a thermal stimulus of 24°C that is perceived as being cold when in contact with the skin but is perceived as being painful if it decreases to 14°C. The attribute of extension refers to the spatial aspects of a stimulus such as its size, location on the body, and separation between points of stimulation, whereas duration refers to the temporal features of the stimulus. Psychophysical techniques have been used to analyze each of these four dimensions, all of which are essential components of human sensory perception.

This review is focused on the application of psychophysical techniques to haptic research. It begins with a description of the unique features of the haptic system that influence how psychophysical methods are implemented. A brief historical perspective on classical psychophysical methods used to measure sensory thresholds is presented followed by more detailed descriptions of these methods. Modern psychophysical techniques used to measure sensory thresholds including adaptive procedures and those associated with the theory of signal detection are then described. In addition to the measurement of thresholds, a number of psychophysical methods, called sensory scaling techniques, are used to measure how suprathreshold stimuli are perceived. In this review, psychophysical techniques are presented in the context of published experimental studies in haptics. There are a number of excellent resources available to the reader who is interested in additional information on psychophysical methods [1], [2], [3].

2 HUMAN HAPTIC SYSTEM

There are several unique characteristics of the haptic sensory system that impact how psychophysical methods can be implemented experimentally. For many stimuli of interest, the process by which observers perceive the stimuli can be either passive or active and the information extracted can be fundamentally different in these two situations. Under passive conditions only tactile cues are available to

• L.A. Jones is with the Department of Mechanical Engineering, Massachusetts Institute of Technology, Room 3-137, 77 Massachusetts Ave., Cambridge, MA 02139. E-mail: ljones@mit.edu.

• H.Z. Tan is with the Human Computer Interaction Group, Microsoft Research Asia, Beijing, P.R. China and the School of Electrical and Computer Engineering at Purdue University, West Lafayette, Indiana. E-mail: hongtan@purdue.edu.

Manuscript received 24 Nov. 2012; accepted 28 Nov. 2012; published online 12 Dec. 2012.

Recommended for acceptance by J.E. Colgate.

For information on obtaining reprints of this article, please send e-mail to: toh@computer.org, and reference IEEECS Log Number TH-2012-11-0098. Digital Object Identifier no. 10.1109/TOH.2012.74.

the observer whose stationary hand makes contact with a surface or object that may or may not be moving. In contrast, during active haptic sensing the observer's hand moves voluntarily over a surface or object and both tactile and kinesthetic cues are available regarding the object being explored [4], [5]. In the latter situation, the sensory information used to identify objects and extract information about their properties arises from stimulation of receptors embedded in skin, muscles, tendons, and joints. For some tasks such as the perception of roughness or the softness of deformable materials, there is little difference between tactile and haptic sensing of surfaces provided there is relative motion between the finger and the object [6], [7], [8]. For other tasks such as the perception of weight, actively lifting an object is superior to sensing its weight only via tactile stimulation of the passive hand [9]. The importance of the mode of exploration on the sensory information available and hence the perceptual performance of subjects means that it is essential to control precisely how subjects interact with the stimuli being presented.

In contrast to the large number of stimuli and associated short interstimulus intervals that are typical of psychophysical studies of the visual and auditory systems, in general the number of haptic stimuli that can be presented during a psychophysical experiment is small. This reflects the delays inherent in feedback-controlled electromechanical systems that are often used to deliver the stimuli, and the time required for subjects to explore the stimuli. In haptic experiments, subjects manually interact with a device or an object and may explore two or three stimuli sequentially before making a judgment. The exposure time required to perceive the stimuli and the duration of a single trial will, therefore, be much longer (3-10 s) than the presentation times in visual or auditory psychophysical experiments in which stimuli can be presented briefly (50-100 ms) and responded to quickly. The sequential nature of manual exploration means that judgments of properties such as numerosity (i.e., how many objects are being grasped) [10] or whether a particular haptic stimulus is present in an array [11] take considerably longer than equivalent judgments made using visual stimuli.

Another consideration in designing psychophysical studies in haptics is that many stimuli are multidimensional and even during a simple act such as grasping an object, information is extracted about its global shape, local surface features, texture, temperature, and compliance [11]. It is, therefore, extremely important to consider which cues participants may use to derive their judgments and to appreciate that most human perception involves judging stimuli in a context. It has occasionally been determined that the information that participants use to perform a task is not necessarily that which the experimenter intended. This has clearly been shown in studies of tactile spatial acuity involving two-point discrimination in which subjects are required to indicate whether one or two points of an aesthesiometer are in contact with the skin. Johnson et al. [12], [13] have shown that subjects can discriminate between one and two 0.5 mm diameter points accurately even when there is no separation between the points based solely on intensive rather than spatial cues—a single point

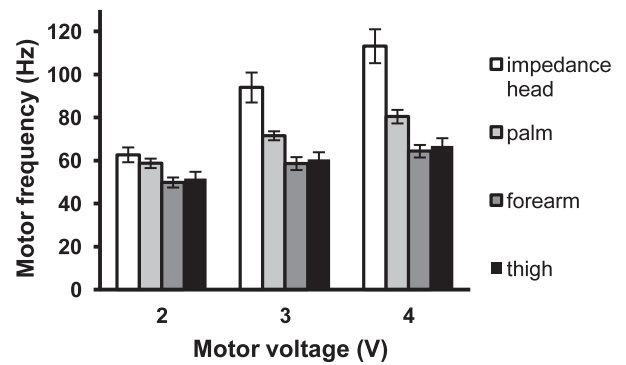


Fig. 1. Mean motor frequency (\pm SEM) recorded from pancake motors mounted on an impedance head and on the palm, forearm and thigh as a function of input voltage. From [16] with permission of the IEEE Computer Society.

feels sharper than two points. The use of this test to measure tactile spatial acuity can obviously lead to an overestimation of the spatial resolution of the skin. More precise instruments and analytical measures have, therefore, been developed to measure two-point thresholds [12]. In a similar vein, Tan et al. [14] built an electromechanical system to measure manual discrimination of compliance in which the distance the thumb moved toward the index finger could be either fixed or randomly varied. They found that participants used the terminal force cues in the fixed distance condition to perceive the compliance of the system and that when these cues were greatly reduced, compliance thresholds were significantly greater, reflecting the poorer resolving capacity for compliance as compared to force. One approach to minimizing the contribution of spurious cues to perceptual judgments is to make the stimuli relatively complex such that a simple unidimensional cue is unable to account for more than a small percentage of correct responses [15].

Finally, in psychophysical experiments involving tactile and haptic stimuli, it is important to characterize the actual stimuli delivered to subjects with precise physical measurements as the properties and dynamics of the human finger/hand/arm affect the output of electromechanical systems. In studies of force or stiffness perception, variations in the forces used to grasp a manipulandum influence the contact area between the finger pads and the device and also the overall stiffness of the hand-device system, both of which affect the mechanical information available to the human observer. In the context of tactile displays, the frequency and amplitude of vibrotactile stimuli delivered by motors mounted on the body can be substantially different from the manufacturers' specifications which are typically based on attaching the motors to rigid structures, as illustrated in Fig. 1.

3 CLASSICAL PSYCHOPHYSICAL METHODS

There are two broad areas of psychophysical research, each of which is associated with a different set of experimental procedures. The first area concerns the measurement of sensory thresholds and in particular absolute and difference thresholds. The second area is focused on the measurement of sensory attributes such as the perceived intensity of suprathreshold stimuli or the ability to identify

TABLE 1
Sensory Resolution and Weber Fractions
for a Range of Tactile and Haptic Stimuli

STIMULUS DIMENSION	RESOLUTION	WEBER FRACTION (%)
Surface texture (roughness)	0.06 μm ^[21]	5-12% ^{[22] [23]}
Curvature	9 μm ^[24]	10% ^[25]
Temperature	0.02-0.09°C ^[26]	0.5-2% ^[27]
Skin indentation	11.2 μm ^[28]	14% ^[29]
Velocity of tactile stimuli		20-25% ^[30]
Vibrotactile frequency (5-200 Hz)	0.3 Hz ^[31]	3-30% ^{[31] [32] [33]}
Vibrotactile amplitude (20-300 Hz)	0.03 μm ^[34]	13-16% ^{[35] [36]}
Pressure	5 gm/mm ² ^[4]	4-16% ^[37]
Force	19 mN ^[38]	7% ^{[39] [40]}
Tangential force		16% ^[41]
Stiffness/compliance		15-22% ^{[14] [42] [43]}
Viscosity		19-29% ^{[44] [45]}
Friction		10-27% ^{[46] [47]}
Electric current	0.75 mA ^[48]	3% ^[49]
Moment of inertia		10-113% ^{[50] [51]}

or categorize stimuli. A persistent theme throughout the history of psychophysics is that there must be a relation between these two aspects of perception, that is the discrimination of differences in intensity must be related in some way to how the apparent magnitude of stimuli is scaled, an endeavor that is characterized as the search for a unifying psychophysical law [17], [18], [19].

The absolute threshold, which is sometimes referred to as RL (from the German Reiz Limen), is defined as the smallest amount of stimulus energy necessary to produce a sensation and may be thought of as the resolution of the sensory system (see Table 1). An absolute threshold is, therefore, detected against a background of the null stimulus, or the noise of the sensory system under study. For most sensory systems, the absolute threshold depends upon the experimental conditions under which it is measured and so these must be specified precisely if comparisons are to be made across experiments. For example, the absolute threshold for detecting a vibratory stimulus on the skin depends on the size of the stimulated area, the duration of the stimulus, the frequency of vibration and the locus of stimulation [20]. As any one of these variables changes so too does the absolute threshold measured.

When a stimulus above absolute threshold is applied to a sensory organ the intensity of the stimulus must increase or decrease by some amount before an individual reports a change in sensation. The difference threshold (DL for Differenz Limen) is defined as the amount of change in a stimulus required to produce a just noticeable difference (JND) in sensation. It was determined by the German physiologist E.H. Weber in 1834 that for many sensory modalities the change in stimulus intensity ($\Delta\phi$) that can be discriminated is a constant fraction (c) of the intensity of the stimulus (ϕ):

$$\Delta\phi/\phi = c. \quad (1)$$

This became known as Weber's law and has been found to hold over a fairly wide range of stimulus intensities [52]. The Weber fraction has been determined for a number of tactile and haptic stimuli as shown in Table 1. It is often used as an index of discrimination and because it is independent of units can be compared across sensory dimensions. The fraction typically increases at extremely low stimulus intensities so a modification to Weber's law is often made with the addition of a small constant to approximate better the empirical data. Not all sensory dimensions follow Weber's law, for example, the JND does not increase systematically with increasing finger span [53] or joint angle [54] as predicted by Weber's law.

From Weber's work on the difference threshold, the German physicist and philosopher Gustav Fechner attempted to establish a more general relation between physical events and perceptual experience. He proposed that sensation magnitude could be quantified indirectly by relating the physical changes in stimulus intensity ($\Delta\phi$) to the JNDs in sensation. Fechner assumed that the JND was a standard unit of sensation and that all JNDs were subjectively equal in intensity. On the basis of these assumptions and the constancy of Weber's fractions, he proposed that there should be a logarithmic relation between physical and perceived intensity in what became known as Fechner's Law:

$$\Psi = k \log \Phi, \quad (2)$$

where Ψ is the sensation magnitude, ϕ is the intensity of the stimulus and k is a constant multiplier. Although it was subsequently shown that JNDs are not perceptually equivalent, Fechner's research and the publication of his *Elemente der Psychophysik* in 1860 were highly significant in that they established the scientific basis for studying the human mind.

Since the time of Fechner, thresholds have generally been considered as being statistical in nature in that there is not a single value above which a stimulus will always be detected and below which it is never detected. On any given trial, the responses of individuals will be variable depending on their state and the environment. Fechner devised three methods for determining absolute and difference thresholds that are still in use today: the method of constant stimuli, the method of limits, and the method of adjustment. All three methods use stimulus values that are fixed prior to the experiment in contrast to *adaptive techniques* in which the stimuli presented depend on the responses of the subject.

In classical psychophysical methods, the standard stimulus refers to a stimulus with a fixed physical value for which a threshold is being determined. The comparison stimulus has a range of values that change from trial to trial, sometimes being greater than and sometimes less than the standard stimulus. These methods seek to create a *psychometric function* in which the detection of a stimulus (absolute threshold) or the discriminability of two stimuli (difference threshold) increases from near zero or chance to perfect performance as a function of the stimulus magnitude or the difference between stimuli. In detection experiments, a specific point on the psychometric function

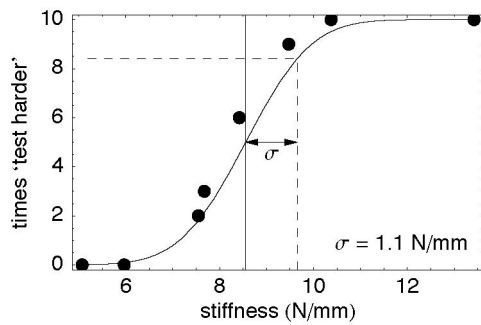


Fig. 2. Psychometric curve with fitted function for stiffness discrimination experiment using the method of constant stimuli. The reference stimulus is indicated by the solid vertical line and the dashed line indicates the position of the threshold, defined at the 84 percent level. From [42] with permission of the IEEE Computer Society.

is defined as the absolute threshold, and in discrimination experiments a parameter associated with the cumulative distribution function, such as the width of a cumulative Gaussian curve fitted to the data, is taken as a measure of the difference threshold.

Fig. 2 shows a psychometric function fitted to the data from a stiffness discrimination experiment in which subjects were required to indicate on each trial which of two stimuli (a reference and comparison) felt harder. From psychometric functions such as these, one or two parameters can be estimated. The most commonly determined parameter is the threshold stimulus value along the stimulus intensity axis, which may be the midpoint of a function that spans the range from chance to perfect performance in a detection experiment (50 percent) or the position of the 84 percent level (width of the cumulative normal curve) in a discrimination experiment (as shown in Fig. 2). A second parameter that can be used to describe performance is the slope of the psychometric function, which is a measure of how rapidly performance changes with a given change in stimulus intensity. In discrimination experiments, an additional parameter that may be of interest is the 0.5 point on the psychometric function that represents the value of the comparison stimulus that over a large number of trials is perceptually equivalent to the standard stimulus. The 0.5 point is known as the *point of subjective equality* (PSE) and the difference between the standard stimulus and the PSE is called the *constant error*.

3.1 Method of Constant Stimuli

For determining absolute thresholds, the method of constant stimuli involves presenting the same set of between five and nine different stimuli that span the range from imperceptible to almost always detected. There are typically 20 trials per stimulus level [55]. The specific stimulus values selected are either based on a priori knowledge about the sensory system or preliminary experiments that indicate the range that satisfies the above conditions. On each trial, the subject indicates whether the stimulus has been perceived and from these data a psychometric function is constructed with the proportion of “yes” responses plotted as a function of stimulus intensity [56]. A distribution function (e.g., logistic, cumulative Gaussian, Weibull, arcsine) is then fitted to the data

and from this the absolute threshold can be defined as the stimulus intensity at which the proportion of “yes” responses is 0.5.

Pairs of stimuli are presented when the method of constant stimuli is used to calculate difference thresholds and with this procedure the participant judges which stimulus (the standard or comparison) is perceived to be greater (e.g., heavier, rougher, larger, and harder). Typically between five and nine values of the comparison stimulus are selected, separated by equal distances along the physical scale, and equally spaced on either side of the standard stimulus. The values of the comparison stimuli are chosen such that the largest stimulus will almost always be judged to be greater than the standard stimulus and the smallest comparison stimulus will almost always be perceived to be less than the standard stimulus (see Fig. 2).

The standard and comparison stimuli may be presented to different receptive areas at the same time (e.g., each hand) and the subject responds left or right on each trial or they may be presented to the same receptive area but consecutively and the subject indicates first or second stimulus. There are errors of measurement associated with each of these conditions (simultaneous versus sequential presentation) which must be controlled for in designing the experiment. In simultaneous stimulus presentation, participants may be influenced by differences in the sensitivity of the receptive areas in addition to differences between the two stimuli presented. This is referred to as the *space error* and is controlled for by presenting the standard stimulus to one site on half the trials and the other site on the other half. When stimuli are presented sequentially, there is a tendency for participants to judge the comparison stimulus as being greater than the standard stimulus when it is presented second as compared to when it is presented first, a finding referred to as the *time error*. This is controlled for by counterbalancing the temporal order of presentation of the comparison and standard stimuli. With this method, the difference threshold can be calculated using the PSE (0.5) and the comparison stimulus values at the 0.25 and 0.75 points on the psychometric function. An upper difference threshold is defined as the difference between the PSE and the 0.75 point and a lower difference threshold is the difference between the PSE and 0.25 point. The mean of these two thresholds is taken as the difference threshold for a particular standard stimulus [1].

3.2 Method of Limits

The method of limits is one of the most frequently used techniques for determining sensory thresholds and although it is considered to be less precise than the method of constant stimuli it is less time consuming and so is considered more efficient [1]. For absolute thresholds, stimuli are presented in a descending and ascending series a number of times beginning with a stimulus either well above or well below the anticipated threshold. On each successive presentation, the intensity of the stimulus is changed by a small amount in the direction of the threshold. In the descending series, the trial stops when the participant reports that the stimulus is no longer perceived and in the ascending series the trial stops when the participant first

indicates the presence of the stimulus. A number of ascending and descending series are presented and the absolute threshold is defined as the mean of the transition points in each of the series presented.

In determining a difference threshold, the method of limits involves presenting standard and comparison stimuli in pairs and on successive presentations the magnitude of the comparison stimulus is changed by a small amount in the direction of the standard stimulus. The comparison stimuli are presented in an ascending and descending series, with the ascending series involving comparison stimuli smaller than the standard stimulus and the descending series involving stimuli greater than the standard stimulus. There are two transition points obtained with this procedure, an upper threshold, which is the point at which “greater” responses change to “equal” responses, and a lower threshold, which is the point at which “less” responses change to “equal” responses. Each of the ascending and descending series is repeated a number of times (typically 3-6). The interval on the standard stimulus dimension over which the participant does not perceive a difference between the standard and comparison stimulus is called the *interval of uncertainty* and is calculated by subtracting the upper threshold from the lower threshold in each series. The difference threshold is defined as half the interval of uncertainty.

The method of limits is associated with two types of errors known as *errors of habituation* and *errors of expectation*. The former refers to the tendency of participants to continue responding with the same response even though they may have detected a change in the stimulus. Errors of expectation refer to the opposite tendency, namely participants anticipate a change in the stimulus and report prematurely that a change in the stimulus has been detected. If these two types of errors were of equal magnitude, they would cancel each other out, but this is unlikely. One strategy to prevent errors of expectation is to vary the starting points in the ascending and descending series, and errors of habituation can be minimized by keeping the length of each series reasonably short.

The up-and-down or staircase method is a variation on the method of limits in which a sequence of stimuli which progressively decrease or increase in value is presented and when the participant's response changes the stimulus value is recorded and the stimulus sequence is immediately reversed [57]. For example, when the participant first says “no” or “equal” in a descending series, the experimenter starts an ascending series which ends when the participant first says “yes” or “equal.” This procedure continues until a sufficient number of transition points have been measured; the mean of these transition points is calculated as the threshold. This method is described in more detail in Section 5.

3.3 Method of Adjustment

The method of adjustment has mainly been used to measure difference thresholds but can be used to measure absolute thresholds. In comparison to the other classical methods for determining absolute thresholds, this method is considered the least precise. As the name suggests, a participant adjusts the intensity of the stimulus, so it is

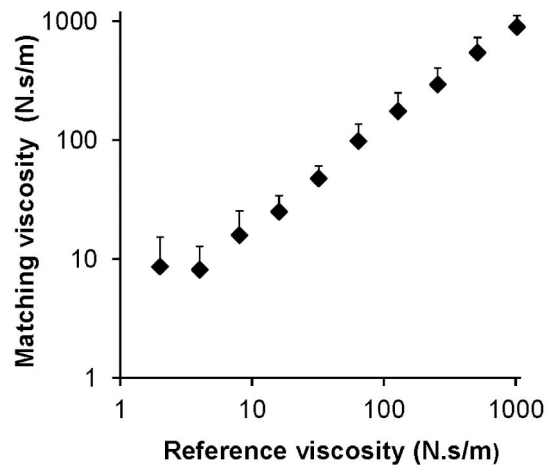


Fig. 3. Relation between the viscosity of the reference motor connected to the right arm and the matching viscosity (set by the subject) of the matching motor connected to the left arm. The standard deviation of each mean is shown. Adapted from [45] with permission of Springer.

mainly used with variables for which there is continuous control over the physical variable (e.g., amplitude of a vibrotactile stimulus). In measuring absolute thresholds with the method of adjustment, the stimulus intensity level is initially set either well above or well below threshold and the participant adjusts the intensity until the stimulus is no longer perceived or is just perceptible. Usually a fairly large number of ascending and descending series are tested, each of which starts at a different point on the physical scale. The mean of the final settings in each series is the absolute threshold.

In measuring difference thresholds with this method, the participant adjusts a comparison stimulus until it is perceived to be equal to a standard stimulus. This method is often referred to as the *method of average error* and it is the difference between the physical intensity of the comparison stimulus set by the participant and that of the standard stimulus that is of interest. The mean of the comparison stimuli is the PSE and the difference between the PSE and standard stimulus is the constant error. Provided there are a sufficient number of trials, the standard deviation of the comparison stimuli can be defined as the difference threshold. This method has been used extensively in studies of force and stiffness perception, in which participants have adjusted the forces produced by two muscle groups (known as the contralateral limb matching procedure) or the elastic or viscous stiffness of two electromechanical systems until they are perceived to be equal (see Fig. 3).

3.4 Summary

From the results of numerous experiments in which classical psychophysical techniques were used, it became very clear that nonsensory factors influenced the ability to detect and discriminate between stimuli. Some people are very careful in their judgments, whereas others may be more lenient and although they may have very similar sensitivities to the stimuli being presented, their thresholds will be different because of these response biases. In the middle of the 20th century, a new era in psychophysical research began with the application of statistical decision

theory to the measurement of thresholds [58]. The theory of signal detection proposed that the person being tested is continuously receiving sensory input and on each trial decides whether the input comes from the presence of a signal (stimulus) or from noise (external or internal background activity). The detection problem is primarily one of detecting signals in noise. The person is assumed to set a response criterion above which she/he will indicate the presence of a stimulus. In this way, a measure of the sensitivity to the stimulus is derived together with an independent estimate of the subject's response criterion. The application of signal detection theory to psychophysics led to the development of explicit predictions about the relations between different experimental procedures. The development of forced-choice and signal detection methodologies provided more objective psychophysical procedures to study perception and computational methods became available to evaluate response bias independent from participants' sensitivities. These techniques are now discussed in detail.

4 SIGNAL DETECTION THEORY

Signal detection theory takes a probabilistic approach to modeling the human decision making process in the presence of noise. It takes into account the fact that the same stimulus is not always perceived the same way. By introducing a decision model, independent measures of human sensitivity and response bias can be obtained. This ability to separate decision making (required to make a response) from perception (what we are really interested in) led to the extensive use of signal detection theory in psychophysical experimental design and data analyses. In this section, we present common experimental paradigms and the associated decision models. This section is largely based on unpublished lecture notes by Durlach [59]. Interested readers should refer to Macmillan and Creelman [60] for a detailed treatment of signal detection theory and its use in psychophysical studies.

4.1 One-Interval, Two-Alternatives, Forced-Choice (1I-2AFC) Experiments

The most common signal detection experiment is called a one-interval, two-alternatives, forced-choice (1I-2AFC) paradigm. One interval means that one stimulus is presented on each trial. Two alternatives means that there are two stimulus alternatives. Forced choice means that the participant has to indicate which of the two stimuli was presented, and it is not acceptable to respond "I don't know." This design is also known as the "yes-no" experiment due to the two-alternative nature of the experiment (see [60]). The two stimuli (S_1 and S_2) can be noise and signal in a detection experiment, or a smaller amplitude signal and a larger amplitude signal in a discrimination experiment, respectively. On each trial, $S_i (i = 1, 2)$ is presented with a predetermined a priori probability (usually 0.5). The probabilities $P(S_1)$ and $P(S_2)$ always add up to 1.0. The results of an experiment can be summarized by a two-by-two stimulus-response matrix:

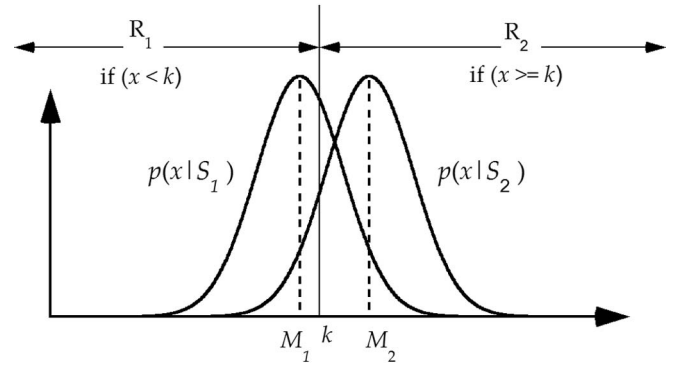


Fig. 4. Decision model for 1I-2AFC experiments.

	R_1	R_2
S_1	N_{11}	N_{12}
S_2	N_{21}	N_{22}

where N_{ij} is the number of times stimulus S_i is called response R_j (assuming, without loss of generality, that R_1 and R_2 are the correct responses for S_1 and S_2 , respectively). Two independent quantities, false-alarm rate (FA) and hit rate (H), are calculated as follows:

$$FA = N_{12}/(N_{11} + N_{12}); H = N_{22}/(N_{21} + N_{22}), \quad (3)$$

from which the sensitivity index and response bias can be derived according to the decision model outlined below.

4.2 Decision Model for 1I-2AFC Paradigm

Fig. 4 illustrates the decision model for the 1I-2AFC paradigm. The variable x is a real random variable and the x -axis is called the "decision axis." Each stimulus presentation determines a value of x . The distributions of x values given S_1 and S_2 are modeled by two conditional probability density functions $p(x|S_1)$ and $p(x|S_2)$, respectively. The statistics of x are independent of all aspects of the experiment except S_1 and S_2 . It is further stipulated that there exists a fixed cut-off value k , called the "criterion," on the x -axis. The participant responds R_1 if and only if $x < k$, and R_2 if and only if $x \geq k$.

The conditional response probabilities $P(R_j|S_i)$ can be expressed in terms of the conditional probability density functions $p(x|S_i)$. The false-alarm and hit rates are, respectively,

$$FA = P(R_2|S_1) = \int_k^{+\infty} p(x|S_1)dx, \quad (4)$$

$$H = P(R_2|S_2) = \int_k^{+\infty} p(x|S_2)dx.$$

Assuming that the density functions are Gaussians with means M_1 and M_2 and an equal variance of σ , we have

$$p(x|S_1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-M_1)^2}{2\sigma^2}} \text{ and } p(x|S_2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-M_2)^2}{2\sigma^2}}. \quad (5)$$

We can now define the two performance metrics, sensitivity index d' and response bias β , as follows:

$$d' = \frac{M_2 - M_1}{\sigma} \quad \text{and} \quad \beta = \frac{1}{\sigma} \left(k - \frac{M_1 + M_2}{2} \right). \quad (6)$$

The sensitivity index d' is the normalized distance along the decision axis between the means of the two conditional probability density functions (see Fig. 4). Its value is dependent on the two stimuli S_1 and S_2 , but not the response criterion k . The d' value, therefore, provides a measure of the participant's sensitivity to the difference between S_1 and S_2 , regardless of the participant's response bias. The ability to estimate detection or discrimination performance independently of response bias is the main reason why signal detection theory experiments are preferred over the classical psychophysical methods outlined in Section 3.

The response bias is defined as the normalized distance between the response criterion k and $(M_1 + M_2)/2$, the average of the means of the two conditional probability density functions (see Fig. 4). A response bias of $\beta = 0$ means that the participant is unbiased in the sense that the response criterion is chosen optimally. Given S_1 and S_2 , d' is constant but β can change with k .

4.3 Data Analysis for 1I-2AFC Paradigm

The false-alarm and hit rates collected from a 1I-2AFC experiment are used to compute the sensitivity index and response bias as follows: The frequencies of occurrence (3) are taken as estimates of the corresponding probabilities (4). The latter are converted to *normal deviates* as follows:

$$FA = \int_{-\infty}^{z(FA)} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx, \quad \text{and} \quad (7)$$

$$H = \int_{-\infty}^{z(H)} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx,$$

where $z(FA)$ and $z(H)$ denote the normal deviates (also called z scores) for the false-alarm and hit rates, respectively. It can then be derived, from (4)-(7), that

$$d' = z(H) - z(FA), \quad \text{and} \quad \beta = -\frac{z(H) + z(FA)}{2}. \quad (8)$$

With this formulation, the receiver operating characteristics curve, also known as the isosensitivity curve, has a simple linear form with a slope of 1 and an intercept of d' :

$$z(H) = z(FA) + d'. \quad (9)$$

Given S_1 and S_2 (hence a constant d'), increasing the hit rate (e.g., by lowering the response criterion k) causes the false-alarm rate to increase as well. Therefore, there exists a tradeoff between the hit and false-alarm rates as one adjusts the response criterion.

To derive a detection or discrimination threshold from a 1I-2AFC paradigm, d' values for several pairs of S_1 and S_2 are obtained. For example, in a texture grating amplitude discrimination experiment [61] the reference amplitude (S_1) was 50 μm and the test amplitudes (S_2) were 55, 60, 65, and 70 μm . The average slope ($\bar{\alpha}$) of the function $d'(\Delta S)$, where $\Delta S = S_1 - S_2$, was calculated by taking the average of the ratios d' over ΔS for the four texture grating pairs. The discrimination threshold (or JND) is then defined as the amplitude difference corresponding to $d' = 1$; or equivalently, the inverse of the average slope, $1/\bar{\alpha}$ (see Section 2.4,

[61]). The criterion of $d' = 1$ corresponds to 69 percent on the psychometric function assuming $\beta = 0$.

Variations of the 1I-2AFC paradigm include the two-interval, two-alternatives, forced-choice (2I-2AFC) paradigm, and the confidence-rating paradigm. Interested readers can find more information in [60].

5 ADAPTIVE TECHNIQUES

Adaptive procedures are efficient ways of estimating thresholds by adapting stimulus intensities based on preceding stimuli and responses. The method of limits is the simplest adaptive procedure in the sense that a reversal in response from yes to no (or vice versa) terminates an experimental run. We review several commonly used adaptive procedures including the simple up-down method (also known as the staircase method) and transformed up-down methods. We also give examples of variations on the up-down methods; for example, by changing the number of stimulus intervals per trial and by interleaving multiple up-down sequences in one experiment.

The simple up-down method is analogous to the method of limits except that an ascending or descending sequence does not terminate after the first reversal. Instead, the experiment continues until many reversals are obtained around the threshold value to be estimated. Although it is not always explicitly stated, the simple up-down method employs the "1-up 1-down" rule. After each "no" response in the detection experiment (stimulus cannot be perceived) or an incorrect response in the discrimination experiment (the larger stimulus level is judged to be the smaller one, or vice versa), the stimulus intensity or the difference between the reference and comparison stimuli is increased, respectively. Conversely, after each "yes" or correct response, the stimulus intensity or the difference is decreased. With this method, the task is made easier or harder based on the subject's performance, rather than following a prescribed sequence of stimulus presentations. The simple up-down method estimates the 50th-percentile point of the psychometric function. Compared to the method of constant stimuli, the simple up-down method is more efficient because it reduces the number of trials at stimulus intensity levels at which the proportion of "yes" responses is close to zero or close to one, where little information is gained about the threshold to be estimated. The threshold level is obtained by averaging the peak and valley values on a plot of stimulus level versus trial number. The first few reversals of a run are usually not included in the final data analysis because the stimulus level has yet to converge to near the threshold level being estimated.

When the step size for changing stimulus level is chosen appropriately, the simple up-down method works well. Typically, a larger step size is used for the first few reversals to quickly bring the stimulus level to close to the threshold being estimated. A smaller step size is used for the remaining reversals to increase the precision in estimating the threshold. Step size can also be gradually decreased during the course of an experiment; for example, a step size can be c/n where c is a constant and n is the trial number [62]. Whenever in doubt, one should aim for a larger initial step size. According to Leek [3], efficiency is reduced by

only 25 percent if the initial step size is twice the optimum value, but by 100 percent if the initial step size is half the optimum value.

With the simple up-down method, only thresholds at the 50 percent point on a psychometric function can be estimated. The threshold value corresponds to either a detection threshold or PSE in a discrimination experiment. Levitt [63] reviews transformed up-down methods that can be used to estimate thresholds at percentile points other than 50 percent on a psychometric function. For example, a transformed one-up, three-down method produces a threshold at the 79.4 percent point on a psychometric function. According to the one-up three-down rule, the stimulus level goes up after one incorrect response and goes down after three *consecutive* correct responses. The point of convergence can be calculated by noting that at the threshold level where stimulus levels converge, the probability of the stimulus level going up or down is equal and must be 0.5. If we let $P(X)$ be the probability of a correct response, then the probability of the stimulus level going down is equal to $P^3(X)$ in a one-up, three down transformed adaptive method. Setting this probability to 0.5 yields $P(X) = 0.794$. Similarly, a one-up, two-down transformed up-down method estimates thresholds at the 70.7 percent point ($\sqrt{0.5}$) on a psychometric function, and a one-up four-down transformed up-down method estimates thresholds at the 84.1 percent point ($\sqrt[4]{0.5}$).

Unlike the transformed up-down methods that adapt stimulus levels after a predetermined sequence of responses, the weighted up-down method adapts the stimulus level after each response [64]. Furthermore, the unforced weighted up-down method allows for an additional “don’t know” response [65]. The changes to stimulus level after each correct response (Δ_{cor}) and incorrect response (Δ_{incor}) are constrained by $\Delta_{\text{cor}} \cdot P(X) + \Delta_{\text{incor}} \cdot (1 - P(X)) = 0$ so that the stimulus level will converge to $P(X)$. The step size for each “don’t know” response (Δ_{unsure}) is taken to be the weighted sum of Δ_{cor} and Δ_{incor} . In the case of a two-alternative detection or discrimination experiment where the probability of making a correct response by chance is 0.5, $\Delta_{\text{unsure}} = 0.5 \cdot \Delta_{\text{cor}} + 0.5 \cdot \Delta_{\text{incor}}$. For $P(X) = 0.75$, $\Delta_{\text{incor}} = -3\Delta_{\text{cor}}$ and $\Delta_{\text{unsure}} = -\Delta_{\text{cor}}$ ($\Delta_{\text{cor}} > 0$ correspond to a decrease in stimulus level). Both simulation and empirical data show the unforced weighted up-down method to be more efficient than the simple up-down method, especially for some naïve participants who have difficulty achieving stable threshold estimates.

One may argue that with the simple up-down method, a participant can fully anticipate the change in stimulus level and even be able to manipulate the stimulus sequence by responding differently. This problem can be avoided by interleaving two or more simple up-down sequences in one experiment. Consider two simple up-down sequences (one ascending A and one descending B; see left panel of Fig. 5). On each trial, one of the two sequences is randomly selected with an equal a priori probability of 0.5 (see right panel of Fig. 5). If sequence A is chosen, then the intensity level of the next trial is determined by the previous trials belonging to sequence A. This way, the participant can no longer expect the stimulus intensity to always increase after an

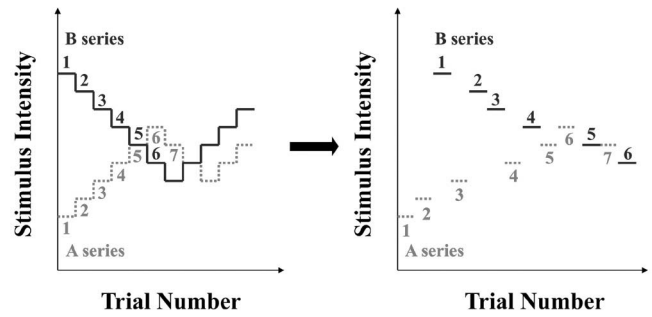


Fig. 5. Illustration of the double-random staircase method. The left panel shows two independently run up-down adaptive sequences, with the numbers indicating the trial number within each sequence. The right panel shows the two sequences interleaved in one experiment in which one of the two sequences is randomly chosen on each trial.

incorrect response, or vice versa. The experiment ends when both sequences have been completed. An example of interleaving five adaptive sequences in one experiment is illustrated later in this subsection.

Another variation of the adaptive procedure is the number of stimuli (“intervals”) presented per trial. In a one-interval experiment, one of two stimuli (noise or signal in a detection experiment; reference or comparison stimulus in a discrimination experiment) is presented per trial. In a two-interval experiment, both stimulus alternatives are presented in a trial, and the participant’s task is to identify the interval (first or second) during which a test stimulus (in the case of a discrimination experiment) is presented. A three-interval experiment can also be devised where the test stimulus is presented at a randomly selected interval and the other two remaining intervals contain the reference stimulus. The three-interval experiment is particularly advisable in an experiment in which it is difficult to label the stimulus alternatives. For example, it might be difficult for a participant to recognize a surface grating as following a sinusoidal or a square-wave profile [66]. It is, however, possible to use the three-interval design in a discrimination experiment in which the participant has to identify only the interval during which the surface grating felt *different* from the gratings presented during the other two intervals. As the number of intervals increases per trial, there is an increase in the total experimental time. Therefore, few studies go beyond three intervals.

The first study to employ a three-interval one-up three-down adaptive procedure was probably Brisben et al. [34]. The most elaborate extension of this method was by Barbagli et al. [67] who interleaved five adaptive sequences, each following a three-interval, one-up three-down procedure. The experiment involved measuring force-direction discrimination thresholds along five reference force directions. To eliminate possible learning effects, the adaptive sequences corresponding to the five reference directions were interleaved. For each reference force direction, a three-interval, one-up three-down procedure was used to estimate the discrimination threshold corresponding to the 79.4 percent point on the psychometric function. On each trial, one of the five reference force directions was randomly chosen with an a priori probability of 0.2. For each reference force direction, the initial difference in force direction ($\Delta\alpha$) was set to

8 degrees (for faster convergence of the stimulus level toward the threshold being estimated). After the first five reversals, $\Delta\alpha$ was reduced to 2 degrees for better accuracy in estimating the threshold. Each of the five sequences was terminated after 12 reversals at 2 degrees. The threshold corresponding to each reference force direction was computed as the average of the peaks and valleys of $\Delta\alpha$ values for the last 12 reversals (see [67] for further details). It is interesting to note that people are generally unable to verbalize the 3D angle difference between two force vectors. By employing the three-interval procedure in which two of the intervals contained the reference force direction and one randomly selected interval contained the test force direction, the participants were able to simply identify the interval during which the force direction was perceived to be different (i.e., “odd one out”).

6 INFORMATION TRANSFER

The psychophysical methods discussed so far are mainly concerned with the characterization of *peripheral* haptic sensory mechanisms. We now switch our discussion to central limitations. From an information theoretical framework, humans can be viewed as a noisy communication channel that receives and sends information through our sensorimotor systems [68]. Performance can then be characterized in terms of information transfer and information rate. Unlike threshold estimates that are mainly limited by the resolution of peripheral sensors (e.g., mechanoreceptors), information measures reflect central limitations such as memory. In this section, we first review concepts and formulas in information theory and the absolute identification paradigm that is typically used to estimate mutual information between stimuli and responses. We then cover a few practical issues in obtaining reliable estimates of information transfer. We end the section with a discussion of how to estimate information transfer associated with multidimensional stimulus sets.

6.1 Information Transfer and Information Rate

A typical way to measure information transfer is to run an absolute identification experiment. First, a set of K stimuli ($S_i, i = 1, \dots, K$) is constructed. A corresponding set of K responses ($R_j, j = 1, \dots, K$) is then assigned so that R_j is the correct response to S_i when $j = i$. On each trial, the participant is presented with a stimulus randomly selected from the stimulus set. The participant chooses a response from the response set after each stimulus presentation. The experimental results are tabulated in the form of a stimulus-response confusion matrix with rows corresponding to stimuli and columns responses. Information transfer (IT) is then calculated from the confusion matrix. The quantity IT measures the increase in information about the signal transmitted from knowledge of the signal received. For a particular stimulus-response pair (S_i, R_j), the quantity IT is given by

$$IT(S_i, R_j) = \log_2 \frac{P(S_i|R_j)}{P(S_i)}, \quad (10)$$

where $P(S_i|R_j)$ is the conditional probability of S_i given R_j , and $P(S_i)$ is the a priori probability of S_i . The average

information transfer is given by the weighted sum of individual ITs:

$$\begin{aligned} IT &= \sum_{j=1}^K \sum_{i=1}^K P(S_i, R_j) \log_2 \left(\frac{P(S_i|R_j)}{P(S_i)} \right), \\ &= \sum_{j=1}^K \sum_{i=1}^K P(S_i, R_j) \log_2 \left(\frac{P(S_i, R_j)}{P(S_i)P(R_j)} \right), \end{aligned} \quad (11)$$

where $P(S_i, R_j)$ is the joint probability of stimulus S_i and response R_j , and $P(R_j)$ is the probability of R_j . The probabilities in the IT equation can be approximated by the frequency of occurrence to obtain the maximum likelihood estimate IT_{est} :

$$IT_{est} = \sum_{j=1}^K \sum_{i=1}^K \frac{n_{ij}}{n} \log_2 \left(\frac{n_{ij} \cdot n}{n_i \cdot n_j} \right), \quad (12)$$

where n_{ij} is the number of times the joint event (S_i, R_j) occurs, $n_i = \sum_{j=1}^K n_{ij}$ and $n_j = \sum_{i=1}^K n_{ij}$ are the row and column sums, respectively, and $n = \sum_{j=1}^K \sum_{i=1}^K n_{ij} = \sum_{j=1}^K n_j = \sum_{i=1}^K n_i$ is the total number of trials collected. A related measure, $2^{IT_{est}}$, is interpreted as the number of stimulus levels that can be correctly identified without error.

One way to contrast an absolute identification experiment with a discrimination experiment is to consider the case of listening to piano notes. During a discrimination experiment, the participant is asked to judge whether a reference note (e.g., middle C) and another note (e.g., C#) have the same pitch. Any two keys on a piano are perfectly discriminable. During an absolute identification experiment, the participant is asked to listen to one note and identify its pitch. Only people with absolute pitch can identify the correct key when a single piano key is played in isolation. Therefore, whereas a discrimination task requires the participant to tell two similar signals apart by comparing them, an absolute identification task requires perfect memory of all signals to correctly identify one signal presented in isolation.

Although many studies employing the absolute identification paradigm report the results only in terms of percent-correct scores, others also report information transfer. In haptics research, the absolute identification paradigm is often used to study identification of factor location [69], [70], [71], [72], [73]. Whenever possible, it is preferred that information transfer be calculated and compared across different studies that use varying numbers of alternatives in the stimulus set. The example confusion matrices shown in Fig. 6 explain why information transfer is preferred to percent-correct scores as the performance measure for identification tasks. Matrices (A) and (B) both represent total confusion, as indicated by 0-bit information transfer in both cases. The percent-correct scores of 50 and 25 percent, respectively, are misleading in two respects. First, both scores may indicate some competence of the participants in identifying stimulus alternatives but it is clear that the participant was performing at chance level in both cases. Second, the lower percent-correct score for (B) may give the appearance of worse performance when in fact performance was equally poor in both cases. A comparison of the matrices (C) and

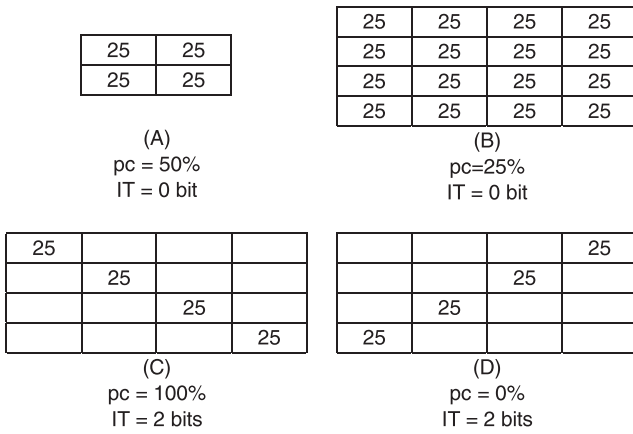


Fig. 6. Four example stimulus-response confusion matrices illustrating the advantage of using an information-transfer measure over percent-correct scores.

(D) shows that participants were able to correlate stimuli and responses perfectly as indicated by the information transfer of 2 bits in both cases, but the percent-correct scores are drastically different. It is true that the participant in case (D) used the wrong responses for all stimuli. However, it is also apparent that the participant was able to identify each stimulus alternative as a distinct signal. The problem of assigning wrong response labels can be easily mitigated, if the main purpose of the study is to find out whether participants can correctly identify stimuli presented in isolation. In this regard, the information transfer of 2 bits in case (D) indicates perfect correlation between stimuli and responses despite the participant's use of wrong response labels.

An important concept associated with the absolute identification experiment is channel capacity, the upper limit for transmitted information. In his classic paper, Miller [74] noted, with numerous (mostly auditory and visual) examples, that transmitted information with stimuli that vary along only one dimension (e.g., pitch or loudness of a tone, a point on a line, saltiness, etc.) is limited to 2.3 to 3.2 bits; or equivalently, 5-7 perfectly identifiable levels. In haptics research, the unidimensional channel capacity appears to be lower: up to four levels for finger-span length and close to three levels for force magnitude and stiffness identification [70], [75]. Experimentally, as the number of alternatives in a stimulus set increases, estimated information transfer will initially increase linearly and then reach an asymptotic value that represents the maximum information transfer that can be obtained with the stimulus set, that is channel capacity (see [76] for data on size identification). It is, therefore, important when designing an absolute identification experiment to ensure that the number of alternatives in a stimulus set exceeds that of the anticipated channel capacity ($\log_2 K > IT_{est}$, where K is the number of stimulus alternatives). This can be simply accomplished by choosing a K that is large enough so that the percent-correct score is < 100 percent.

There is, however, a cost to choosing a K that is too large. It has been shown that IT_{est} is a statistically biased estimate of IT and that the bias generally decreases as the total number of trials increases [77]. A general rule-of-thumb is to collect $> 5K^2$ trials to minimize the statistical bias in

estimated information transfer. Therefore, the larger the K , the greater the number of trials that needs to be presented per participant, thereby increasing the experimental time. In practice, pilot tests should be conducted with one or two participants to get a rough estimate of channel capacity. Then the number of stimulus alternatives should be chosen such that the task is difficult but not too challenging for the participants. We recommend choosing a K such that $\log_2 K$ exceeds IT_{est} by 1 to 2 bits. With this approach, there is no longer the need to conduct multiple experiments to find the maximum number of stimuli, for example, the number of tactor locations on a belt that one can identify without error. Once the channel capacity for tactor localization is known, then the total number of identifiable tactor locations can be calculated by $2^{IT_{est}}$. One can then redesign the placement of tactors using up to $2^{IT_{est}}$ number of tactors (see [69] for an example of this approach).

The channel capacity of 7 ± 2 in Miller's [74] paper is for unidimensional stimuli only. With stimuli that vary along multiple dimensions, the overall channel capacity can be greatly increased. It is a challenge, however, to measure experimentally multidimensional channel capacity because of the thousands of trials needed to obtain an unbiased estimate of IT_{est} . For example, Rabinowitz et al. [78] studied identification of vibratory intensity, frequency, and contactor area with five levels per parameter, resulting in a total of $K = 125$ stimulus alternatives. This required a total of 78,125 trials according to the $5K^2$ rule! Durlach et al. [75] proposed a *general additivity law* that allows the prediction of multidimensional IT from ITs estimated with unidimensional stimuli. For example, let us consider the case of a 2D experiment where the stimuli vary along two dimensions (e.g., frequency and amplitude of a vibration). It has been shown by Ashby and Townsend [79] that

$$IT(F, A) \leq IT(F, \text{fixed } A) + IT(A, \text{fixed } F), \quad (13)$$

where $IT(F, A)$ denotes the information transfer estimated from a 2D absolute identification experiment in which the values of both frequency (F) and amplitude (A) are varied and have to be identified, $IT(F)$ denotes the information transfer estimated from a unidimensional absolute identification experiment in which only the value of F is varied and has to be identified, and $IT(A)$ denotes a similar unidimensional absolute identification experiment on amplitude in which the value of F is held constant. The equality only holds when the two dimensions are perceptually independent, which rarely happens. Therefore, the sum of unidimensional ITs usually results in an overestimate of the multidimensional IT. The general additivity law states that

$$IT(F, A) = IT(F, \text{roving } A) + IT(A, \text{roving } F), \quad (14)$$

always holds, provided that during the unidimensional experiment on F , the values of A are randomly varied from trial to trial, and similarly with the experiment on A . The 1D ITs so obtained are, therefore, appropriately reduced to account for the perceptual interaction between F and A . This general additivity law has been empirically verified in several studies [75], [80], [81], [82]. It can be easily extended to more than two dimensions in the form:

$$\begin{aligned}
IT(A, B, C, \dots) &= IT(A, \text{rov } B \& C \& \dots) \\
&+ IT(B, \text{rov } A \& C \& \dots) + IT(C, \text{rov } A \& B \& \dots) + \dots,
\end{aligned} \tag{15}$$

where A, B, C, \dots denote dimensions along which a multi-dimensional stimulus set can vary.

It may appear surprising that humans can correctly identify only a small number of stimulus levels even though the stimulus range contains a large number of JNDs. For example, within a force range of 0.1-5.0 N, there are 41 JNDs (assuming a force-magnitude JND of 10 percent) but the information transfer is 1.54 bits (or equivalently, only three perfectly identifiable force magnitude levels) [70]. This is explained by a decision model developed by Durlach et al. [75] stipulating that whereas the JND is limited by the “sensory noise” in our peripheral sensory systems, IT is further limited by a “memory noise” due to the need to hold the absolute rather than relative force-magnitude percepts in our memory. Therefore, it is much easier to *discriminate* the relative magnitudes of two forces than to *identify* the absolute magnitude of a force in isolation, or to *discriminate* the relative locations of two factors on a belt than to *identify* the absolute location of a single factor on the belt. Therefore, it is important to determine whether the nature of a research question involves discrimination of haptic stimuli or absolute identification as the results can be drastically different.

7 SCALING TECHNIQUES

The measurement of sensory thresholds provides valuable information about the sensory processes under study and how different stimulus conditions influence human perception. For example, we know that pressure thresholds on the skin are influenced by the speed of indentation, the size of the stimulated area, the presence of preindentation, and the exposure time to the stimulus. In the context of tactile and haptic displays, thresholds are often used to evaluate the performance of the device, perhaps with the objective of determining its optimal configuration [83] or to compare different devices [84]. In addition to sensory thresholds, the measurement of other attributes is required for a more complete understanding of the sensory system. Questions such as how the shape of an object on the skin affects the perception of contact force [25], or whether the perceived roughness of virtual textures is similar to that of real textures [85] require a different set of experimental methods. The procedures developed to measure these sensory attributes are known as scaling techniques and are used to derive an understanding of the relation between changes in the physical stimulus (weight, temperature, compliance, friction, spatial frequency) and the associated sensation (heaviness, warmth/cold, softness, slipperiness, roughness).

Many scaling techniques were first proposed by Stevens [86] who hypothesized that subjects could estimate the perceived intensity of stimuli directly by assigning numbers that corresponded to the sensations or by selecting an equivalent intensity in another modality (e.g., matching the perceived intensity of a force with a sound whose loudness could be varied). He proposed that “equal stimulus ratios produce equal subjective ratios” and determined that the

relation between the physical and perceived intensity of a stimulus was best described by a power function:

$$\Psi = k\Phi^a, \tag{16}$$

where Ψ is sensation magnitude, k is a constant determined by the absolute size of the numbers assigned, and a is the exponent of the power function that depends on the sensory modality and the conditions of stimulus presentation. When these scaling techniques were first used, there was a considerable amount of research and theorizing about the exponent of the power function, whether it was specific for each perceptual continuum and how much it varied for the same continuum as the stimulus conditions changed. Across sensory modalities, the exponent was found to range from a low of about 0.2 for the function relating light intensity to brightness to a high of over 2.5 for the relation between current and electrotactile intensity [87]. The latter exponent reflects the very rapid increase in perceived intensity with small changes in the current delivered to the skin.

Within a modality, the exponent can vary considerably as a function of the conditions of stimulus presentation, range of stimuli presented and instructions given to subjects. For example, the exponent of the power function relating the forces generated by the hand and arm to their perceived magnitude ranges from 0.8 to 2.0 across various experiments (for a review see [88]). The effects of the context in which stimuli are presented are considerable, with response variance in scaling tasks being up to 100 times greater than that typically found in threshold discrimination tasks [89], [90]. To place this in the framework of a measuring instrument such as a scale used to measure weight, it would be as if the measured weight of an object changed as a function of the other objects being weighed, the object that was last weighed in the series and the total number of objects being weighed.

The range of stimuli presented within a modality and the dynamic ranges of different sensory continua, that is, the ratio of the maximum to minimum usable stimulus intensities, also have a profound effect on the exponents on the power function relating physical to perceived intensity. Teghtsoonian [87] showed that for 24 different perceptual continua the power function exponents could be predicted from the stimulus ranges employed, with higher exponents being associated with smaller ranges of stimuli. On the basis of his analyses, Teghtsoonian [87] argued that subjects generate a constant range of judgments in scaling experiments, independent of the range of stimuli presented.

7.1 Magnitude Estimation and Production

Over the years numerous experiments have been conducted using the ratio scaling method of *magnitude estimation* in which subjects make numerical estimates of perceived magnitudes. It is generally agreed that subjects should be free to choose their own range of numbers and should not be constrained by an experimenter-defined standard stimulus and associated modulus (e.g., this stimulus has a value of 100). Subjects are simply instructed to make their judgments reflect how many times greater or less one sensation is than another (the ratio between two sensations). During an experimental session, between 10 and 20 stimuli

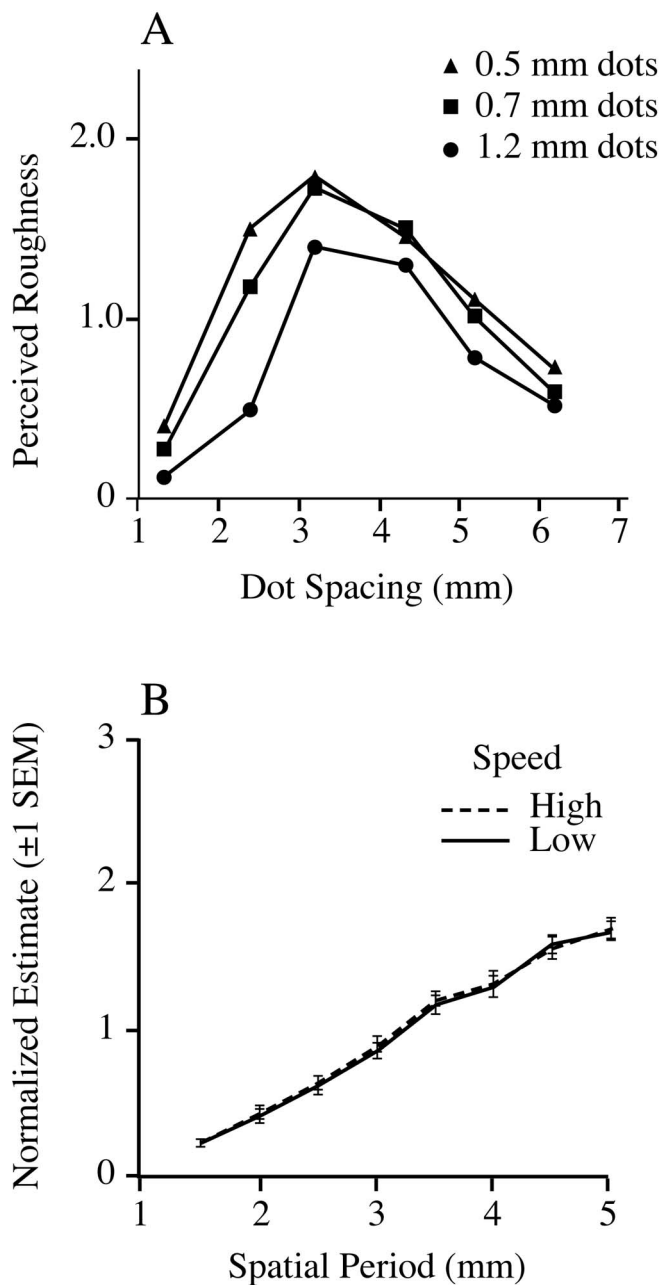


Fig. 7. (A) Normalized roughness magnitude as a function of dot spacing for three different dot diameter sets. Reprinted from [91] with permission of the Society for Neuroscience. (B) Mean normalized roughness estimates as a function of spatial period at low and high scanning speeds. Reprinted from [92] with permission of Springer.

are typically presented in a random order with 5 to 10 repetitions of each stimulus. The geometric mean of the numbers assigned by each individual to each stimulus is calculated and is the perceived magnitude for that stimulus [1]. The data from magnitude estimation experiments can also be normalized by dividing each subject's responses by the grand mean of all the subject's estimates given during the experiment, as illustrated in Fig. 7. In the upper graph (see Fig. 7A), perceived roughness is displayed as a function of dot spacing for three different dot-diameter sets [91]. The lower graph (see Fig. 7B) shows normalized perceived roughness estimates as a function of the temporal frequency

of the scanned surface [92]. Normalization is necessary because subjects are free to choose any numerical range in reporting the perceived magnitude.

One alternative to assigning numbers to represent the perceived magnitude of stimuli is to use a physical scale as the basis for making comparisons. With this method subjects match a comparison stimulus to a reference, and the value of the matching stimulus on the physical scale (e.g., force in N , angle of rotation in degrees) is taken as the perceived magnitude of the reference stimulus. This technique has been used extensively in studies of force and weight perception [39], [88] (see Section 3.3) and in haptic research on the perception of parallelity where for example a comparison bar is rotated to match a reference bar [93].

Another ratio scaling technique that is used less frequently and is the opposite of magnitude estimation is *magnitude production*. In this procedure, the experimenter provides the numerical estimate and the participant adjusts a stimulus to match the number. This method requires that the stimulus is continuously variable and can be under the participant's control (e.g., force applied by a finger, temperature of a Peltier device, frequency of a vibrating motor). The use of both techniques (magnitude estimation and production) to derive psychophysical functions has been advocated as any errors and biases associated with one method are offset by the other. A procedure for combining the functions from the two procedures called the *method of numerical magnitude balance* involves taking the geometric means of the two psychophysical functions [94].

There have been many critics of magnitude estimation who argue that subjects' responses are discrete verbal responses that possess no quantitative magnitude [95], and that biasing factors such as the stimulus context and range of stimuli presented all have a profound effect on the numbers subjects assign to stimuli [96]. It is clear that the exponent of the power function fitted to the data from a magnitude estimation experiment varies with the stimulus set, sequence of stimulus presentation, stimulus range, spacing between stimuli, and feedback to subjects [18]. Some of these variables are described in more detail in Section 7.4. These factors are important because they indicate that psychophysical scaling data are not invariant when experimental procedures change, and that scaling models are often only descriptive for the particular data set from which they are derived.

7.2 Category Scaling

In category scaling experiments, subjects are presented with a large number of stimuli and told to assign them to a specified number of categories which can either be numbers (e.g., 1, 2, and 3) or verbal labels (e.g., no exertion, light exertion, and maximal exertion). The number of categories can range from 3 to 20 [1]. There is an assumption with this method that sensory attributes are measured on an equal interval scale. However, category judgments are not only determined by the sensation magnitude but are strongly influenced by the other stimuli presented and the frequency with which each stimulus is presented (see Section 7.4). There is a marked tendency (or bias) for subjects to assign stimuli such that all categories are used about equally often regardless of whether the stimulus range is narrow or wide.

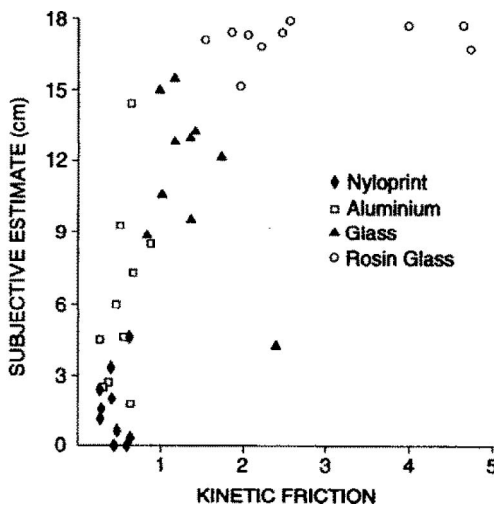


Fig. 8. Relation between subjective estimates of surface friction and the mean measured friction for four surfaces with different coefficients of friction. From [100] with permission of the American Physiological Society.

This means that the spacing of stimuli on the physical dimension and the frequency distribution of stimuli can significantly influence the shape of the psychophysical function. Due to these biases, category scales of the same stimuli judged in different stimulus sets are not linearly related and so category scales cannot be considered a valid interval scale (i.e., one that is invariant to linear transformations). Category scales are generally ordinal scales and so are limited in terms of the analyses that can be performed on the data acquired.

Verbally labeled category scales such as those used to measure pain intensity (visual analog scale (VAS)) or physical exertion (the Borg scale) tend to produce more consistent data than numerical category scales because the labels function as anchors that assist subjects in judging stimuli. These scales appear to be more reliable because they are based on a reference frame derived from everyday experiences rather than on the context of the other stimuli presented. For example, concepts such as slight pain and severe pain are based on accumulated experience rather than on the experimental context. For some scales, such as the Borg scale, it has been found that there is a linear relation between the category ratings given by subjects and other objective measures of physical exertion such as heart rate and oxygen consumption [97]. This is interpreted as indicating that the scale is a valid measure of exertion. One further advantage of verbally labeled category scales is that there is often relatively high agreement across subjects.

7.3 Visual Analog Scales

Another type of rating scale that has been used to estimate stimulus intensity is visual analog scales. The scale is usually a horizontal or vertical line (150–180 mm) presented on a computer screen with two anchors, one at each end, and subjects simply mark the position on the line that corresponds to the perceived intensity of the stimulus under study. Because the scale is continuous, VAS are usually considered superior to category scales such as the Likert Scale. Visual analog scales have been used extensively in studies of pain intensity and have been validated

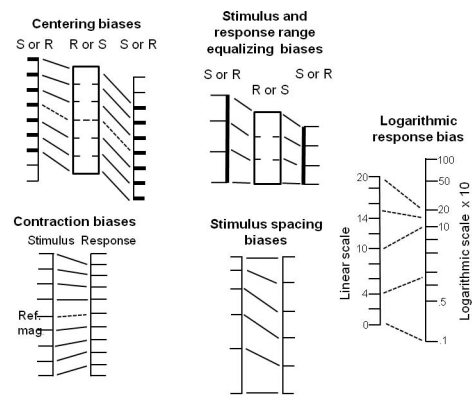


Fig. 9. Graphical representation of biases in quantifying judgments. S is stimulus, and R is response. Redrawn and adapted from [101] with permission of the American Psychological Association.

in that context by examining the relation between ratings and experimentally induced pain [98].

In haptic research, VAS have been used to study the perception of friction of different surfaces [99], [100]. As with the VAS used in pain research, verbal anchors were provided at each end of the line (“most slippery” and “most sticky”) to assist subjects in making judgments. The validity of the VAS used in the above studies was evaluated by determining the relation between subjective estimates of surface friction and measurements of the actual kinetic friction of each surface as it was stroked by the finger. These data are illustrated in Fig. 8.

VAS are subject to the same limitations as the other rating scales described. Their reliability and validity must be demonstrated and not assumed, that is, it needs to be shown that the scale yields consistent results over time and that subjects’ scaling of the phenomena being investigated is consistent with judgments derived using other methods.

7.4 Context Effects—Response Biases

Poulton [96], [101] described a number of biases that influence how subjects respond to stimuli in scaling experiments in which they are required to assign numbers to the perceived magnitude of stimuli or assign stimuli to specific categories. These biases are depicted graphically in Fig. 9. The biases were further classified in terms of effects due to the overall range of stimuli presented (range biases) and those that relate to nonlinearities within the overall range (nonlinear biases). The former includes 1) the centering bias in which subjects center their range of responses on the range of stimuli presented. This means that the same physical stimulus will be assigned different apparent magnitudes in two experiments if it is presented in a stimulus set with a different range of intensities. 2) The stimulus and response equalizing bias refers to the finding that subjects use their full range of responses regardless of the range of stimuli presented. 3) The contraction bias refers to a general property of human responses, namely that large stimuli and differences between stimuli are underestimated and small stimuli and differences are overestimated. This phenomenon is often termed the regression effect.

Nonlinear biases include 1) the local contraction bias in which subjects respond to very high intensity and very low

intensity stimulus ranges as if they had less extreme values than they actually have, and 2) the stimulus spacing bias refers to the finding that subjects respond as if all stimuli in the set being presented were equally probable and were equally spaced geometrically. In category rating experiments this means that all categories are used equally often. 3) The logarithmic bias is a characteristic of how subjects use numbers when estimating the perceived magnitude of a stimulus or assigning a stimulus to a category. Subjects treat 1-, 2-, 3-, ... n -digit numbers as if they were equally frequent which means that numbers are assigned logarithmically rather than linearly.

Various strategies have been proposed to eliminate or minimize the effects of these biases in category rating and magnitude estimation experiments. Some of these solutions such as asking for only a single judgment from each subject are not realistic in the context of psychophysical research [96], [101]. In some experiments, the investigator may be interested only in the relative and not the absolute positions of sensory magnitudes on the rating scale. In this situation, the centering bias would not be important, particularly if the investigator anchors the most intense stimuli to the upper end of the scale and the least intense stimuli to the lower end of the rating scale. Other solutions involve spacing stimuli geometrically and presenting them equally often (to mitigate the effects of the stimulus spacing bias) and using a range of numbers that all have the same number of digits (to avoid the logarithmic bias). Even if it is not feasible to mitigate all the biases that may influence subjects' judgments, it is important to be aware of their impact and to acknowledge that quantitative judgments are likely to be biased.

8 MULTIDIMENSIONAL SCALING

The scaling methods described above are designed to measure a subject's perception of a single stimulus dimension, such as the intensity of vibration or the warmth of a stimulus. In many situations, however, changing one dimension of a stimulus alters the sensory experience of other dimensions as well. For example, when the amplitude of a vibrotactile stimulus increases, the frequency of the stimulus is also perceived to increase even though it remains constant [102]. One approach to studying these phenomena is called multidimensional scaling (MDS), which uses a set of statistical techniques to separate and identify the underlying perceptual dimensions of a group of stimuli [103], [104]. MDS involves having subjects first make judgments of the similarity (or dissimilarity) of stimuli. All possible combinations of stimuli used in the experiment are presented in pairs to subjects who judge the similarity (or dissimilarity). From these measurements, each stimulus is represented as a point in multidimensional space such that the distances between pairs of points represent the degree of similarity among the pairs of stimuli. From these data, the minimal number of underlying dimensions (m) required to represent the results adequately is derived and represented in a two- or three-dimension euclidean space, although the space can be noneuclidean and have more dimensions. Stress values indicating the goodness of fit as a function of m and other properties of

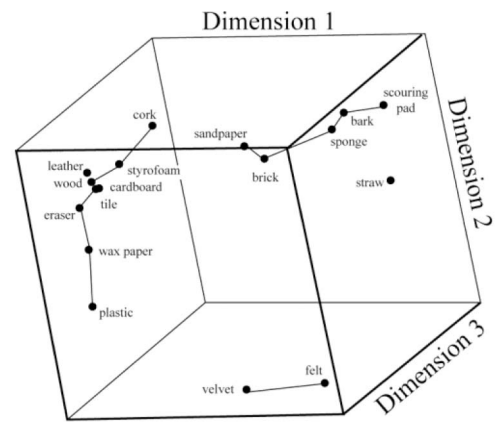


Fig. 10. Cube representing the 3D scaling solution for perceptual texture space. Reprinted from [105] with permission of Springer.

MDS solutions are evaluated to determine the reasonableness of the dimensions in terms of such a space. Every stimulus is located on each of the dimensions derived.

In haptics research, MDS has been used most frequently to study the perception of texture [105], although it has recently been used to analyze the haptic perception of shape [106]. Texture is a multidimensional percept in that surfaces can be perceived along a number of dimensions such as soft/hard, rough/smooth, sticky/slippery. Hollins et al. [105] used a grouping method in which subjects were required to sort the 17 textures presented into at least three and at most seven groups. The MDS analyses revealed that three dimensions were required to account for their subjects' judgments, as depicted in Fig 10. There were two robust and orthogonal dimensions—roughness/smoothness and hardness/softness—and a third dimension that appeared to reflect elasticity or springiness. Using a much larger stimulus set with 124 different samples, Bergmann Tiest and Kappers [107] used MDS to analyze the perception of material parameters. They found that four dimensions were adequate to represent haptic material space but that the dimensions could not be mapped simply to physical properties such as compressibility and roughness which were also measured.

MDS has also been used as a tool to assist in developing tactile and haptic icons [108], [109]. In this context, the objective is to create stimuli that are readily distinguishable and can easily be associated with their intended meaning. Once a set of stimuli have been designed, MDS is used to understand how people perceptually organize this candidate set of icons and to determine which stimulus dimensions are most salient. One of the limitations in using MDS with large stimulus sets, however, is the considerable amount of time and effort required by participants to compare all possible pairs of stimuli or to sort items into similar groups [110]. The former approach is preferred as it provides a more detailed analysis of the perceptual space.

9 CONCLUSIONS

In this review, we have described the range of psychophysical methods available to study haptic perception and have detailed how they should be implemented. The selection of a particular psychophysical method is determined in part

by the question that the researcher is trying to answer. If the sensitivity of two tactile or haptic displays is being compared, then measurement of the difference thresholds of subjects when interacting with the device may be the appropriate metric. In this situation, as in all comparisons between thresholds derived in different experiments, it is important that the investigator specify the point on the psychometric function selected to determine the threshold (e.g., 70.7 percent, 84.1 percent). If the research question entails a comparison between real and simulated stimuli (e.g., real materials and a thermal display simulating the materials), then an absolute identification procedure may be the appropriate method [111]. With this method, in addition to percent correct measures, subjects' responses should be analyzed in terms of the information theoretical framework as it provides an understanding of human channel capacity which can be directly compared across experiments [112].

A further factor that can influence the choice of a particular psychophysical method is the number of trials required to measure the variable of interest precisely. As discussed in Section 1, due to the sequential nature of manual exploration, haptic stimuli take longer to process than their visual or auditory counterparts. It is, therefore, not feasible to conduct experiments requiring thousands of trials which may take hours for subjects to complete. As a result, the haptic researcher has to weigh the time required to conduct the experiment against the precision needed from the results obtained. A method that requires fewer trials may be preferred over one that is theoretically more precise but has other disadvantages associated with a prolonged testing time.

The application of statistical decision theory to the measurement of thresholds has provided more robust psychophysical techniques for studying perception and has enabled models to be developed that evaluate response bias independently from sensitivity. The sensitivity index d' provides a bias-free measure of detection or discrimination performance that is not influenced by a participant's response criterion. By defining a performance criterion of $d' = 1$, thresholds can be estimated from d' values collected from multiple stimuli pairs. Although popular in visual and auditory psychophysical experiments, signal detection theory has yet to be widely applied to psychophysical studies in haptics.

A variety of psychophysical techniques are available to measure the perception of suprathreshold stimuli. Most of these methods are susceptible to a number of well-documented stimulus and response biases that must be taken into consideration when interpreting participants' data [96], [101]. It is clear that sensory stimuli are evaluated in relation to the context in which they are presented, which means that the same stimulus can be rated quite differently when the set of stimuli to which it belongs changes. A further important issue to consider when using rating scales in psychophysical research is the need to show evidence that the scale is reliable and valid. It is not sufficient to assume that numbers assigned on a categorical scale such as a five-point Likert scale possess quantitative magnitude.

In conclusion, the haptic modality presents a number of unique challenges when designing psychophysical experiments due to the nature of the stimuli under study and the time taken by participants to judge them. It is

clear that provided the conditions of stimulus presentation are well controlled and the task assigned to participants is precisely specified, then reliable and valid psychophysical data can be obtained. The use of these techniques has contributed to advancing our understanding of human haptic perception.

ACKNOWLEDGMENTS

This work was supported in part by a grant to Lynette A. Jones from the US National Science Foundation (NSF) under grant no. IIS-1016998, and in part by a grant to Hong Z. Tan from the NSF under grant no. IIS-0904423.

REFERENCES

- [1] G.A. Gescheider, *Psychophysics: The Fundamentals*, third ed. Lawrence Erlbaum, 1997.
- [2] S.A. Klein, "Measuring, Estimating, and Understanding the Psychometric Function: A Commentary," *Perception and Psychophysics*, vol. 63, pp. 1421-1455, 2001.
- [3] M.R. Leek, "Adaptive Procedures in Psychophysical Research," *Perception and Psychophysics*, vol. 63, pp. 1279-1292, 2001.
- [4] L.A. Jones and S.J. Lederman, *Human Hand Function*. Oxford Univ. Press, 2006.
- [5] S.J. Lederman and R.L. Klatzky, "Haptic Perception: A tutorial," *Attention, Perception and Psychophysics*, vol. 71, pp. 1439-1459, 2009.
- [6] R.M. Friedman, K.D. Hester, B.G. Green, and R.H. LaMotte, "Magnitude Estimation of Softness," *Experimental Brain Research*, vol. 191, pp. 131-142, 2008.
- [7] S.J. Lederman, "The Perception of Surface Roughness by Active and Passive Touch," *Bull. Psychonomic Soc.*, vol. 18, pp. 253-255, 1981.
- [8] R.T. Verrillo, S.J. Bolanowski, and F.P. McGlone, "Subjective Magnitude Estimate of Tactile Roughness," *Somatosensory and Motor Research*, vol. 16, pp. 352-360, 1999.
- [9] E.E. Brodie and H.E. Ross, "Jiggling a Lifted Weight Does Aid Discrimination," *Am. J. Psychology*, vol. 98, pp. 469-471, 1985.
- [10] M.A. Plaisier, W.M. Bergmann Tiest, and A.M.L. Kappers, "Haptic Object Individuation," *IEEE Trans. Haptics*, vol. 3, no. 4, pp. 257-265, Oct.-Dec. 2010.
- [11] S.J. Lederman and R.L. Klatzky, "Relative Availability of Surface and Object Properties during Early Haptic Processing," *J. Experimental Psychology: Human Perception and Performance*, vol. 23, pp. 1680-1707, 1997.
- [12] J.C. Craig and K.O. Johnson, "The Two-Point Threshold: Not a Measure of Tactile Spatial Resolution," *Current Directions in Psychological Science*, vol. 9, pp. 29-32, 2000.
- [13] K.O. Johnson and J.R. Phillips, "Tactile Spatial Resolution: 1. Two-Point Discrimination, Gap Detection, Grating Resolution, and Letter Recognition," *J. Neurophysiology*, vol. 46, pp. 1177-1191, 1981.
- [14] H.Z. Tan, N.I. Durlach, G.L. Beauregard, and M.A. Srinivasan, "Manual Discrimination of Compliance Using Active Pinch Grasp: The Roles of Force and Work Cues," *Perception and Psychophysics*, vol. 57, pp. 495-510, 1995.
- [15] K.O. Johnson, T. Yoshioka, and F. Vega-Bermudez, "Tactile Functions of Mechanoreceptive Afferents Innervating the Hand," *J. Clinical Neurophysiology*, vol. 17, pp. 539-558, 2000.
- [16] K.O. Sofia and L.A. Jones, "Mechanical and Psychophysical Studies of Surface Wave Propagation during Vibrotactile Stimulation," to be published in *IEEE Trans. Haptics*.
- [17] L.E. Krueger, "Reconciling Fechner and Stevens: Toward a Unified Psychophysical Law," *Behavioral and Brain Sciences*, vol. 12, pp. 251-320, 1989.
- [18] G.R. Lockhead, "Psychophysical Scaling: Judgments of Attributes or Objects," *Behavioral and Brain Sciences*, vol. 15, pp. 543-601, 1992.
- [19] H.E. Ross, "On the Relations between Discriminability and Apparent Magnitude," *British J. Math. and Statistical Psychology*, vol. 50, pp. 187-203, 1997.
- [20] G.A. Gescheider, J.H. Wright, and R.T. Verrillo, *Information-Processing Channels in the Tactile Sensory System*. Taylor and Francis, 2009.

- [21] R.H. LaMotte and M.A. Srinivasan, "Surface Microgeometry: Tactile Perception and Neural Encoding," *Information Processing in the Somatosensory System*, O. Franzen and J. Westman, eds., pp. 49-58, Macmillan, 1991.
- [22] J.W. Morley, A.W. Goodwin, and I. Darian-Smith, "Tactile Discrimination of Gratings," *Experimental Brain Research*, vol. 49, pp. 291-299, 1983.
- [23] R.J. Sinclair and H. Burton, "Tactile Discrimination of Gratings: Psychophysical and Neural Correlates in Human and Monkey," *Somatosensory and Motor Research*, vol. 8, pp. 241-248, 1991.
- [24] I.E. Gordon and V. Morison, "The Haptic Perception of Curvature," *Perception and Psychophysics*, vol. 31, pp. 446-450, 1982.
- [25] A.W. Goodwin, K.T. John, and A.H. Marceglia, "Tactile Discrimination of Curvature by Humans Using Only Cutaneous Information from the Fingertips," *Experimental Brain Research*, vol. 86, pp. 663-672, 1991.
- [26] K.O. Johnson, I. Darian-Smith, C. LaMotte, B. Johnson, and S. Oldfield, "Coding of Incremental Changes in Skin Temperature by a Population of Warm Fibers in the Monkey: Correlation with Intensity Discrimination in Man," *J. Neurophysiology*, vol. 42, pp. 1332-1353, 1979.
- [27] K.O. Johnson, I. Darian-Smith, and C. LaMotte, "Peripheral Neural Determinants of Temperature Discrimination in Man: A Correlative Study of Responses to Cooling Skin," *J. Neurophysiology*, vol. 36, pp. 347-370, 1973.
- [28] R.S. Johansson and A.B. Vallbo, "Detection of Tactile Stimuli. Thresholds of Afferent Units Related to Psychophysical Thresholds in the Human Hand," *J. Physiology*, vol. 297, pp. 405-422, 1979.
- [29] J. Mei, R.P. Tuckett, D.A. Poulos, K.W. Horch, J.Y. Wei, and P.R. Burgess, "The Neural Signal for Skin Indentation Depth. II Steady Indentations," *J. Neuroscience*, vol. 3, pp. 2652-2659, 1983.
- [30] G.K. Essick, O. Franzen, and B.L. Whitsel, "Discrimination and Scaling of Velocity of Stimulus Motion across the Skin," *Somatosensory and Motor Research*, vol. 6, pp. 21-40, 1988.
- [31] O. Franzen and J. Nordmark, "Vibrotactile Frequency Discrimination," *Perception and Psychophysics*, vol. 17, pp. 480-484, 1975.
- [32] D.A. Mahns, N.M. Perkins, V. Sahai, L. Robinson, and M.J. Rowe, "Vibrotactile Frequency Discrimination in Human Hairy Skin," *J. Neurophysiology*, vol. 95, pp. 1442-1450, 2006.
- [33] M. Rothenberg, R.T. Verrillo, S.A. Zahorian, M.L. Brachman, and S.J. Bolanowski, "Vibrotactile Frequency for Encoding a Speech Parameter," *J. Acoustical Soc. of Am.*, vol. 62, pp. 1003-1012, 1977.
- [34] A.J. Brisben, S.S. Hsiao, and K.O. Johnson, "Detection of Vibration Transmitted through an Object Grasped in the Hand," *J. Neurophysiology*, vol. 81, pp. 1548-1558, 1999.
- [35] J.C. Craig, "Difference Threshold for Intensity of Tactile Stimuli," *Perception and Psychophysics*, vol. 11, pp. 150-152, 1972.
- [36] E. Francisco, V. Tannan, Z. Zhang, J. Holden, and M. Tommerdahl, "Vibrotactile Amplitude Discrimination Capacity Parallels Magnitude Changes in Somatosensory Cortex and Follows Weber's Law," *Experimental Brain Research*, vol. 191, pp. 49-56, 2008.
- [37] H.Z. Tan, B. Eberman, M.A. Srinivasan, and B. Cheng, "Human Factors for the Design of Force-Reflecting Haptic Interfaces," *Proc. Am. Soc. of Mechanical Eng.: Dynamic Systems and Control Conf.*, vol. 55-1, pp. 353-359, 1994.
- [38] E.E. Brodie and H.E. Ross, "Degrees of Freedom of Movement in Weight Discrimination: Effect of Stimulus Intensity," *Int'l Soc. for Psychophysics Fechner Day 86*, pp. 15-20, 1986.
- [39] L.A. Jones, "Matching Forces: Constant Errors and Differential Thresholds," *Perception*, vol. 18, pp. 681-687, 1989.
- [40] X.D. Pang, H.Z. Tan, and N.I. Durlach, "Manual Discrimination of Force Using Active Finger Motion," *Perception and Psychophysics*, vol. 49, pp. 531-540, 1991.
- [41] H.E. Wheat, L.M. Salo, and A.W. Goodwin, "Human Ability to Scale and Discriminate Forces Typical of Those Occurring during Grasp and Manipulation," *J. Neuroscience*, vol. 24, pp. 3394-3401, 2004.
- [42] W. Bergmann Tiest and A.M.L. Kappers, "Cues for Haptic Perception of Compliance," *IEEE Trans. Haptics*, vol. 2, pp. 189-199, 2009.
- [43] L.A. Jones and I.W. Hunter, "A Perceptual Analysis of Stiffness," *Experimental Brain Research*, vol. 79, pp. 150-156, 1990.
- [44] W. Bergmann Tiest, A.C.L. Vrijling, and A.M.L. Kappers, "Haptic Perception of Viscosity," *Proc. Int'l Conf. Haptics: Generating and Perceiving Tangible Sensations (EuroHaptics '10)*, pp. 29-34, 2010.
- [45] L.A. Jones and I.W. Hunter, "A Perceptual Analysis of Viscosity," *Experimental Brain Research*, vol. 94, pp. 343-351, 1993.
- [46] W.B. Knowles and T.B. Sheridan, "The 'Feel' of Rotary Controls: Friction and Inertia," *Human Factors*, vol. 8, pp. 209-215, 1966.
- [47] W.R. Provancher and N.D. Sylvester, "Fingerpad Skin Stretch Increases the Perception of Virtual Friction," *IEEE Trans. Haptics*, vol. 2, no. 4, pp. 212-223, Oct.-Dec. 2009.
- [48] G.B. Rollman, "Electrocutaneous Stimulation: Psychometric Functions and Temporal Integration," *Perception and Psychophysics*, vol. 5, pp. 289-293, 1969.
- [49] G.R. Hawkes, "Information Transmitted via Electrical Cutaneous Stimulus Duration," *J. Psychology*, vol. 51, pp. 293-298, 1961.
- [50] J.G. Kreifeldt and M.-C. Chuang, "Moment of Inertia: A Psychophysical Study of an Overlooked Sensation," *Science*, vol. 206, pp. 588-590, 1979.
- [51] H.E. Ross and A.J. Benson, "The Weber Fraction for Moment of Inertia," *Int'l Soc. for Psychophysics, Fechner Day 86*, pp. 71-76, 1986.
- [52] E.H. Weber, *The Sense of Touch*. Academic Press, 1834/1978.
- [53] N.I. Durlach, L.A. Delhorne, A. Wong, W.Y. Ko, W.M. Rabinowitz, and J. Hollerbach, "Manual Discrimination and Identification of Length by the Finger-Span Method," *Perception and Psychophysics*, vol. 46, pp. 29-38, 1989.
- [54] H.Z. Tan, M.A. Srinivasan, C.M. Reed, and N.I. Durlach, "Discrimination and Identification of Finger Joint-Angle Position Using Active Motion," *ACM Trans. Applied Perception*, vol. 4, 2-1-2-14, 2007.
- [55] W.A. Simpson, "The Method of Constant Stimuli Is Efficient," *Perception and Psychophysics*, vol. 44, pp. 433-436, 1988.
- [56] F.A. Wichman and N.J. Hill, "The Psychometric Function: 1. Fitting, Sampling and Goodness of Fit," *Perception and Psychophysics*, vol. 63, pp. 1293-1313, 2001.
- [57] T.N. Cornsweet, "The Staircase Method in Psychophysics," *Am. J. Psychology*, vol. 75, pp. 485-491, 1962.
- [58] W.P. Tanner and J.A. Swets, "A Decision-Making Theory of Visual Detection," *Psychological Rev.*, vol. 61, pp. 401-409, 1954.
- [59] N.I. Durlach, "A Decision Model for Psychophysics," *Communication Biophysics Group, Research Laboratory of Electronics, MIT*, Oct. 1968.
- [60] N.A. Macmillan and C.D. Creelman, *Detection Theory: A User's Guide*, second ed. Lawrence Erlbaum, 2004.
- [61] M. Kocsis, S. Cholewiak, R. Traylor, B. Adelstein, E.D. Hirlman, and H.Z. Tan, "Discrimination of Real and Virtual Surfaces with Sinusoidal and Triangular Gratings Using the Fingertip and Stylus," to be published in *IEEE Trans. Haptics*.
- [62] H. Robbins and S. Monro, "A Stochastic Approximation Method," *Annals of Math. Statistics*, vol. 22, pp. 400-407, 1951.
- [63] H. Levitt, "Transformed Up-Down Methods in Psychoacoustics," *J. Acoustical Soc. of Am.*, vol. 49, pp. 467-477, 1971.
- [64] C. Kaernbach, "Simple Adaptive Testing with the Weighted up-down Method," *Perception and Psychophysics*, vol. 49, pp. 227-229, 1991.
- [65] C. Kaernbach, "Adaptive Threshold Estimation with Unforced-Choice Tasks," *Perception and Psychophysics*, vol. 63, pp. 1377-1388, 2001.
- [66] S.A. Cholewiak, K. Kim, H.Z. Tan, and B.D. Adelstein, "A Frequency-Domain Analysis of Haptic Gratings," *IEEE Trans. Haptics*, vol. 3, no. 1, pp. 3-14, Jan.-Mar. 2010.
- [67] F. Barbagli, K. Salisbury, C. Ho, C. Spence, and H.Z. Tan, "Haptic Discrimination of Force Direction and the Influence of Visual Information," *ACM Trans. Applied Perception*, vol. 3, pp. 125-135, 2006.
- [68] W.R. Garner, *Uncertainty and Structure as Psychological Concepts*. Wiley, 1962.
- [69] H.-Y. Chen, J. Santos, M. Graves, K. Kim, and H.Z. Tan, "Tactor Localization at the Wrist," *Proc. Sixth Int'l Conf. Haptics: Perception, Devices and Scenarios (EuroHaptics '08)*, pp. 209-218, 2008.
- [70] S.A. Cholewiak, H.Z. Tan, and D.S. Ebert, "Haptic Identification of Stiffness and Force Magnitude," *Proc. Symp. Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 87-91, 2008.
- [71] L.A. Jones and K. Ray, "Localization and Pattern Recognition with Tactile Displays," *Proc. Symp. Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 33-39, 2008.
- [72] I. Oakley, Y. Kim, J. Lee, and J. Ryu, "Determining the Feasibility of Forearm Mounted Vibrotactile Displays," *Proc. Symp. Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 27-34, 2006.

- [73] J.B.F. Van Erp, A.H.C. van Veen, C. Jansen, and T. Dobbins, "Waypoint Navigation with a Vibrotactile Waist Belt," *ACM Trans. Applied Perception*, vol. 2, pp. 106-117, 2005.
- [74] G.A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Rev.*, vol. 63, pp. 81-97, 1956.
- [75] N.I. Durlach, H.Z. Tan, N.A. Macmillan, W.M. Rabinowitz, and L.D. Braid, "Resolution in One Dimension with Random Variations in Background Dimensions," *Perception and Psychophysics*, vol. 46, pp. 293-296, 1989.
- [76] H.Z. Tan, "Identification of Size Using the PHANToM: Towards a Set of Building Blocks for Rendering Haptic Environments," *Proc. Sixth Int'l Symp. Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 197-203, 1997.
- [77] G.A. Miller, "Note on the Bias of Information Estimates," *Information Theory in Psychology*, H. Quastler, ed., pp. 95-100, Free Press, 1954.
- [78] W.M. Rabinowitz, A.J.M. Houtsma, N.I. Durlach, and L.A. Delhorne, "Multidimensional Tactile Displays: Identification of Vibratory Intensity, Frequency, and Contactor Area," *J. Acoustical Soc. of Am.*, vol. 82, pp. 1243-1252, 1987.
- [79] F.G. Ashby and J.T. Townsend, "Varieties of Perceptual Independence," *Psychological Rev.*, vol. 93, pp. 154-179, 1986.
- [80] S.L. Campbell, "Uni- and Multidimensional Identification of Rise Time, Spectral Slope, and Bandwidth," PhD dissertation, Dept. of Speech and Hearing Sciences, Univ. of Washington, 1993.
- [81] G. Denesvich, "Identification of Frequency and Amplitude through Cutaneous Stimulation," BS thesis, Dept. of Electrical Eng. and Computer Science, MIT, 1995.
- [82] H.Y. Chen, J. Park, S. Dai, and H.Z. Tan, "Design and Evaluation of Identifiable Key-Click Signals for Mobile Devices," *IEEE Trans. Haptics*, vol. 4, no. 4, pp. 229-241, July/Aug. 2011.
- [83] N. Garcia-Hernandez, N.G. Tsagarakis, and D.G. Caldwell, "Feeling through Tactile Displays: A Study on the Effect of the Array Density and Size on the Discrimination of Tactile Patterns," *IEEE Trans. Haptics*, vol. 4, no. 2, pp. 100-110, Mar./Apr. 2011.
- [84] C.M. Salisbury, R.B. Gillespie, H.Z. Tan, F. Barbagli, and J.K. Salisbury, "What You Can't Feel Won't Hurt You: Evaluating Haptic Hardware Using a Haptic Contrast Function," *IEEE Trans. Haptics*, vol. 4, no. 2, pp. 134-146, Mar./Apr. 2011.
- [85] B. Unger, R. Hollis, and R. Klatzky, "Roughness Perception in Virtual Textures," *IEEE Trans. Haptics*, vol. 4, no. 2, pp. 122-133, Mar./Apr. 2011.
- [86] S.S. Stevens, "On the Psychophysical Law," *Psychological Rev.*, vol. 64, pp. 153-181, 1957.
- [87] R. Teghtsoonian, "On the Exponents in Stevens' Law and the Constant in Ekman's Law," *Psychological Rev.*, vol. 78, pp. 71-80, 1971.
- [88] L.A. Jones, "Perception of Force and Weight: Theory and Research," *Psychological Bull.*, vol. 100, pp. 29-42, 1986.
- [89] D. Laming, *Sensory Analysis*. Academic Press, 1986.
- [90] D. Laming, *The Measurement of Sensation*. Oxford Univ. Press, 1997.
- [91] C.E. Connor, S.S. Hsiao, J.R. Phillips, and K.O. Johnson, "Tactile Roughness: Neural Codes That Account for Psychophysical Magnitude Estimates," *J. Neuroscience*, vol. 10, pp. 3823-3836, 1990.
- [92] E.M. Meftah, L. Belingard, and C.E. Chapman, "Relative Effects of the Spatial and Temporal Characteristics of Scanned Surfaces on Human Perception of Tactile Roughness Using Passive Touch," *Experimental Brain Research*, vol. 132, pp. 351-361, 2000.
- [93] A.M. Kappers, "Large Systematic Deviations in the Haptic Perception of Parallelity," *Perception*, vol. 28, pp. 1001-1012, 1999.
- [94] R.P. Hellman and J.J. Zwislowski, "Monaural Loudness Function at 1000 cps Tone and Internal Summation," *J. Acoustical Soc. of Am.*, vol. 35, pp. 856-865, 1963.
- [95] R.N. Shepherd, "Psychological Relations and Psychophysical Scales: On the Status of 'Direct' Psychophysical Measurement," *J. Math. Psychology*, vol. 24, pp. 21-57, 1981.
- [96] E.C. Poulton, *Bias in Quantifying Judgments*. Lawrence Erlbaum, 1989.
- [97] G. Borg, "Psychophysical Bases of Perceived Exertion," *Medicine and Science in Sports and Exercise*, vol. 14, pp. 377-381, 1982.
- [98] D.P. Price, P.A. McGrath, A. Rafii, and B. Buckingham, "The Validation of Visual Analogue Scales as Ratio Scale Measures for Chronic and Experimental Pain," *Pain*, vol. 17, pp. 45-56, 1983.
- [99] L.M. Grierson and H. Carnahan, "Manual Exploration and the Perception of Slipperiness," *Perception and Psychophysics*, vol. 68, pp. 1070-1081, 2006.
- [100] A.M. Smith and S.H. Scott, "Subjective Scaling of Smooth Surface Friction," *J. Neurophysiology*, vol. 75, pp. 1957-1962, 1996.
- [101] E.C. Poulton, "Models for Biases in Judging Sensory Magnitude," *Psychological Bull.*, vol. 86, pp. 777-803, 1979.
- [102] J.W. Morley and M.J. Rowe, "Perceived Pitch of Vibrotactile Stimuli: Effects of Vibration Amplitude, and Implications for Vibration Frequency Coding," *J. Physiology*, vol. 431, pp. 403-416, 1990.
- [103] W.S. Torgerson, *Theory and Methods of Scaling*. Wiley, 1958.
- [104] F.W. Young and R.M. Hamer, *Theory and Applications of Multidimensional Scaling*. Lawrence Erlbaum, 1994.
- [105] M. Hollins, R. Faldowski, S. Rao, and F. Young, "Perceptual Dimensions of Tactile Surface Texture: A Multidimensional Scaling Analysis," *Perception and Psychophysics*, vol. 54, pp. 697-705, 1993.
- [106] N. Gaißert, S. Waterkamp, R.W. Fleming, and I. Bühlhoff, "Haptic Categorical Perception of Shape," *PLOS ONE*, vol. 7, no. 8, article e43063, 2012.
- [107] W. Bergmann Tiest and A.M.L. Kappers, "Analysis of Haptic Perception of Materials by Multidimensional Scaling and Physical Measurements of Roughness and Compressibility," *Acta Psychologica*, vol. 121, pp. 1-20, 2006.
- [108] J. Luk, J. Pasquero, S. Little, K. MacLean, V. Levesque, and V. Hayward, "A Role for Haptics in Mobile Interaction: Initial Design Using a Handheld Tactile Display Prototype," *Proc. SIGCHI Conf. Computer Human Interactions Conf.*, pp. 171-180, 2006.
- [109] K. MacLean and M. Enriquez, "Perceptual Design of Haptic Icons," *Proc. EuroHaptics*, pp. 351-363, 2003.
- [110] J. Pasquero, J. Luk, S. Little, and K. MacLean, "Perceptual Analysis of Haptic Icons: An Investigation into the Validity of Cluster Sorted MDS," *Proc. Symp. Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 437-444, 2006.
- [111] H.N. Ho and L.A. Jones, "Development and Evaluation of a Thermal Display for Material Identification and Discrimination," *ACM Trans. Applied Perception*, vol. 4, pp. 13-1-13-24, 2007.
- [112] H.Z. Tan, C.M. Reed, and N.I. Durlach, "Optimum Information-Transfer Rates for Communication through Haptic and Other Sensory Modalities," *IEEE Trans. Haptics*, vol. 3, no. 2, pp. 98-108, Apr.-June 2010.



Lynette A. Jones received the PhD degree from McGill University. She is a senior research scientist in the Department of Mechanical Engineering at the Massachusetts Institute of Technology. Her research interests include tactile communication systems and the development of wearable vibrotactile displays that can be used for navigation and communication in real and simulated environments. An additional research area she is interested in is the development of thermal displays that can be used to facilitate object identification in virtual environments. She is a senior member of the IEEE.



Hong Z. Tan received the PhD degree from Massachusetts Institute of Technology. She is a senior researcher and research manager of the Human Computer Interaction Group at Microsoft Research Asia and a professor of electrical and computer engineering at Purdue University. Her research interests include haptic human-machine interfaces in the areas of haptic perception, rendering, and multimodal performance. She is a senior member of the IEEE.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.