ELSEVIER

BioSystems 43 (1997) 189–197

BioSystems

Simulation of human sensory performance

Willy Wong [a], Kenneth H. Norwich [b,*]

[a] *Faculty of Engineering, Toyama Prefectural University, Toyama, Japan*

[b] *Institute of Biomedical Engineering, Departments of Physiology and Physics, University of Toronto, 4 Taddle Creek Rd., Toronto, Ontario M5S 3G9, Canada*

Received 31 March 1997; accepted 31 May 1997

Abstract

The capacity of human sensory systems for transmitting information has been approximated in the past by using statistical estimators. However, a substantial margin of error remained. The problem is that the error can be reduced to a negligible level only by increasing the number of human trials or tests to the order of about 10^4. Since a human subject can perform at peak only in the order of 10^2 trials per day, the requisite total number of trials could be obtained realistically only by pooling of data from several subjects. Following Houtsma, we have overcome this problem to a large extent by the use of computer simulation. By introducing parameters characteristic of a given subject into the simulation program, we are able to reproduce the subject's performance (say for 500 trials), and to extrapolate his or her performance using the simulation program to 30 000 trials. In this way we can establish limits to the capacity of a single human being to transmit information. © 1997 Elsevier Science Ireland Ltd.

Keywords: Information; Absolute judgment; Statistical bias; Sensory simulation; Theory of senses

1. Introduction

Information theory may be applied to a sensory continuum to measure the amount of information transmitted from a sensory stimulus. However, the number of experimental trials required to produce a result of statistical significance is extraordinarily high, in the order of 10^4. In our experience, a human subject can retain peak concentration long enough to produce only about 100–200 trials per day (although some investigators do press for as many as 500). Various suggestions (which are discussed below) have been advanced to overcome this problem, but only one is really practicable. We have followed the process of Houtsma (1983) in utilizing a computer simulation to provide the missing data. Our simulator models the subject and provides responses that he or she would have made had it been possible to continue the human testing over a period of months or years.

* Corresponding author. Tel.: +1 416 9786698.

PII S0303-2647(97)00040-3

In 1951, Garner and Hake (1951) published a now-celebrated paper applying Shannon's methods of information theory to psychology. While Shannon's work was derived as a means of analyzing the transmission of signals across noisy channels, Garner and Hake were able to apply his theory in the context of human absolute judgements. The absolute judgements considered involve the classification of stimuli into their assigned categories. For example, the range of sound intensities from 1 to 90 decibels may be divided into ten equally spaced categories, 1–10 dB (category 1), 11–20 dB (category 2)… A subject may then be presented with a tone and asked to identify the category to which the tone belongs. If there are only two or three categories from which to choose, the subject is usually able to identify the tone without error. When the number of categories is increased, say to four or five, the subject will make mistakes. The frequency of error increases as the number of categories is increased. Similar experiments have been carried out with many sensory modalities such as vision and taste.

Garner and Hake gave an informational interpretation to absolute judgments. If m equally probable events can occur, the Shannon information content of the event is given by the well-known formula

$$\text{information} = \log(m). \tag{1}$$

Garner and Hake showed that the sensory information transmitted to the subject is less than $\log(m)$ due to the presence of noise (equivocation/confusion) in the system. With use of a stimulus/response matrix, a complete compendium of the number of times a stimulus in category j was identified as belonging to category k, they were able to calculate the amount of information received per stimulus. Experimentally, it was determined that the information received or transmitted could be increased by increasing the number of available categories, m. However, the transmitted information could not be increased beyond 1.5–2 natural units of information, corresponding to a 'virtual' value of $e^{1.7} \sim 6$ categories that could be identified without error. This remarkable fact lead Miller (1956) to call it the 'magical number 7 ± 2'. For example, Garner (1953) obtained a maximum of approximately six categories for loudness. This natural limit to the quantity of information that can be transmitted per stimulus was approximately constant for many different sensory modalities and category types. Many of these concepts are discussed at length by Garner (1962).

A stimulus/response (or confusion) matrix is used to tabulate the results from an experiment in absolute judgments. The elements of the matrix, N_{jk}, show how many times a stimulus from category j (where $1 \leq j \leq m$, for m categories) was identified as a stimulus from category k (where $1 \leq k \leq m$). In Fig. 1, the elements of the confusion matrix are written out explicitly. Using the symbol x_j to define a stimulus in the jth category, and y_k to define a response in the kth category, we obtain an estimate of the probability $p(x_j, y_k)$ by dividing N_{jk} by N, the total number of trials, where

$$N = \sum_{j=1}^{m} \sum_{k=1}^{m} N_{jk}. \tag{2}$$

The number of times stimulus j was presented, irrespective of the response, is given as

$$N_{j\Box} = \sum_{k} N_{jk}. \tag{3}$$

Similarly, the number of times the response k was given, irrespective of the input, is given as

	y_1	y_2	$\cdots$	y_k	$\cdots$	y_m
x_1	N_{11}	N_{12}	$\cdots$	N_{1k}	$\cdots$	N_{1m}
x_2	N_{21}	N_{22}				$\vdots$
$\vdots$	$\vdots$		$\ddots$			
x_j	N_{j1}					
$\vdots$	$\vdots$					
x_m	N_{m1}	$\cdots$				N_{mm}

Fig. 1. The input/output (or confusion) matrix with the elements written out explicitly. The x_j elements represent categories of different stimulus intensities and the y_k elements represent different response categories. m is the total number of categories.

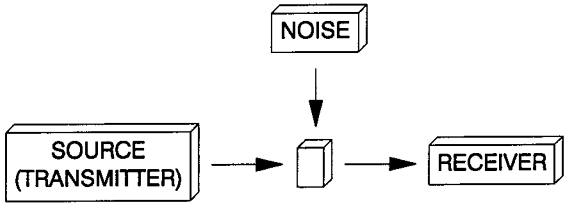

Fig. 2. Schematic diagram showing a channel of communication between the source and the receiver. Signals sent along the channel may be received incorrectly due to the presence of noise.

$$N_{\square k} = \sum_j N_{jk}. \quad (4)$$

With $N_{j\square}$ and $N_{\square k}$, we can easily estimate the corresponding probabilities as

$$p(x_j) = N_{j\square}/N \qquad p(y_k) = N_{\square k}/N. \quad (5)$$

We further define $p(y_k|x_j)$ as the conditional probability of obtaining y_k, given that x_j was transmitted. $p(x_j|y_k)$ is defined analogously. Two fundamental equations link these probabilities:

$$p(x_j, y_k) = p(x_j|y_k)\, p(y_k)$$
$$p(x_j, y_k) = p(y_k|x_j)\, p(x_j). \quad (6)$$

Shannon's information measure is defined as

$$H = -\sum_i p_i \log p_i, \quad (7)$$

where p_i is the discrete probability for the ith occurrence. If the p_i's are all equal, we retrieve the simple Eq. (1). Given $p_i < 1$ for all i, H is an upwards convex function. Shannon applied his H function to calculate the information transmitted across a noisy channel. That is, a source transmits information across a channel to which noise is added (please see Fig. 2). At the other end, the receiver obtains the transmitted information less the information loss due to noise (also called the equivocation). That is,

$$\text{transmitted information} = \text{source information} - \text{equivocation}. \quad (8)$$

Garner and Hake adapted Shannon's mathematical formulation and H function for use with the confusion matrix. The source information is calculated from Eq. (7) using

$$H(X) = -\sum_j p(x_j) \log p(x_j), \quad (9)$$

while the equivocation is taken as

$$H(X|Y) = -\sum_j \sum_k p(x_j, y_k) \log p(x_j|y_k). \quad (10)$$

Hence, the transmitted information, $\mathscr{I}(X|Y)$ or, using the simpler symbol, $\mathscr{I}_t$, is given as

$$\mathscr{I}_t = \mathscr{I}(X|Y) = H(X) - H(X|Y). \quad (11)$$

Using experimental data, the right-hand side of this equation can be evaluated, and it may be demonstrated numerically how transmitted information increases with increasing m when N is held constant.

2. Experimentation and simulation

Subjects, seated in a sound-attenuated room, were presented tones at 1000 Hz, each for a duration of 1.5 s. Tones were presented binaurally through headphones. A time interval of 20 s separated the stimulus tones. The intensity of tones varied from 1 to 90 dB HL or 11–90 dB HL (HL = hearing level). Intensities were randomized uniformly over this range using a standard random number generator. We modified the paradigm by which the confusion matrix has been obtained in the past. Instead of requiring the subject to relearn category assignments each time a different number of categories is used, we asked the subject to respond only with integral decibel values. Experiments conducted in the past tested the combined capacity of the subject to recall category boundaries, and to classify tones into the recalled categories. Presently, our interest is solely in the latter: the ability of a subject to discriminate between tones by placing them in correct categories. Hence, we eliminated the need for relearning category boundaries. 160–200 (tones) trials, were run on a given subject in a single day. 480–500 trials, in total were run on each subject. $\mathscr{I}_t$ was estimated in the usual manner using Eq. (11). These calculated values, as we shall see, systematically overestimated the transmitted information.

How can one overcome the 'bias' induced by a small sample size, N? Returning to Eq. (9), $N_{j\square}/N$, for a finite sample size, is an estimator of $p(x_j)$. Thus,

$$\hat{H}(X) = -\sum_j \frac{N_{j\square}}{N} \log \frac{N_{j\square}}{N} \tag{12}$$

is the estimator for

$$H(X) = -\sum_j p(x_j) \log p(x_j). \tag{13}$$

The difference $H - \hat{H}$ has been termed the bias. We can write analogous expressions for the estimators and biases of $\hat{H}(Y)$, $\hat{H}(X|Y)$ and $\hat{H}(Y|X)$. Since the absolute value of the bias will decrease with increasing N, it can be seen that $\mathscr{I}_t$, as evaluated from Eq. (11) using estimators, is properly regarded as a function not only of m, but also of N. Several investigators have derived theoretical expressions estimating the bias in $\mathscr{I}_t$ for a sample of finite size and from these estimates, one can predict how quickly the estimated value will converge to the true value. However, their results tended to be either too inaccurate due to the approximations made (Miller, 1955) or too hard to evaluate (Carlton, 1969). Miller's equation is quite simple, but requires the assumption that Np_i is large. Rogers and Green (1955) examined the moments of H in order to estimate its variance; their result is considerably more general, making fewer assumptions. However, the final expression for the calculation of $H(X)$, which is quite complicated, must be generalized to work for the two-dimensional categorical matrix. The most general result is found in the paper by Carlton, but the evaluation of the result requires knowledge of the actual probabilities ($p(x_j, y_k)$, $p(x_j)$, etc., which are not known).

Hence, we followed the lead of Houtsma (1983), who utilized a computer simulation to generate missing data for the categorical matrix. For a given input X, Houtsma simulated the human response by adding a uniformly distributed random noise R to the input X. In other words, the output Y is obtained from $Y = X + R$. A parameter s was introduced to characterize the width of the uniform distribution. He then plotted $\mathscr{I}_t$ as a function of the number of trials, with different values of s. Houtsma found monotonically decreasing values of $\mathscr{I}_t$ for increasing number of trials as was expected from the overestimation of information. Later, Houtsma's work was extended by Mori (1991) and Mori and Ward (1996) to multivariate information analysis in absolute identification experiments. They used a series of Monte Carlo simulations to investigate the overestimation of information in the study of sequential dependencies. However, unlike Houtsma's approach, employing a linear relationship ($Y = X + R$) to simulate the response, they used a matrix manipulation scheme to generate response sequences.

Our method was to employ the computer to simulate each subject's performance for large values of N, in this way reducing the bias to negligible levels and obviating the need for an impossibly large number of human trials. Of course, it was necessary that the parameters determining each subject's performance could be obtained with requisite precision from a limited number of human trials in order to be reasonably certain that the simulation would emulate the subject's performance accurately. Our method differed from Houtsma's in that we assumed that R is normally distributed. We proceeded as follows.

It can easily be shown that

$$\mathscr{I}(X|Y) = \mathscr{I}(Y|X). \tag{14}$$

Writing out $\mathscr{I}(Y|X)$ in full we have

$$\begin{aligned}\mathscr{I}(Y|X) &= H(Y) - H(Y|X) \\ &= -\sum_k p(y_k) \log p(y_k) \\ &\quad + \sum_j p(x_j) \sum_k p(y_k|x_j) \log p(y_k|x_j).\end{aligned} \tag{15}$$

Appealing to the results we had obtained experimentally, we can estimate the probability distributions in Eq. (15). A sample of the data obtained from our experiments is found in Fig. 3. $p(x_j)$ is constant, since it is controlled by the experimenter who imposes this condition. We notice that $p(y_k)$ is also nearly constant, indicating a lack of preference on the part of the subject for any particular response category, except perhaps for those near the margins. These so-called 'edge effects' or 'anchor effects' have been observed in almost all category experiments. $p(y_k)$ can be estimated by the convolution of two probability density functions (please see Section 4), but it can be approximately simulated using a standard generator for pseudorandom numbers with the uniform distribution. That is, $p(y_k) = 1/m$, the reciprocal of the number of categories. Finally, $p(y_k|x_j)$ is the distribution governing the probability of response y_k, given input x_j. That is, $p(y_k|x_j)$, which describes the process of human error, is given by the distribution of points along each row of the matrix. Inspection of the matrix in Fig. 3 shows that these probability functions which are discrete can be approximated by the (continuous) normal distribution, centered about the element on the main diagonal. For the first two rows and the final row, the normal distribution is a poor approximation for the skewed distributions observed near the left and right edges (anchor effects).

The variance of each row does not vary a great deal numerically across the different rows. Moreover, the mean row variance was observed to be approximately constant for each subject with constant value of m regardless of the number of trials. For example, in one subject (Subject B) for $N = 160$, $m = 20$, the geometric mean variance was calculated to be 2.15 (date of measurement 02/04/95); for $N = 320$, $m = 20$, variance = 2.63 (02/04/95 + 08/13/95); and for $N = 480$, $m = 20$, variance = 2.50 (02/04/95 + 08/13/95 + 08/20/95).

In our simulations, we have taken a constant variance for all rows, obtained by taking a geometric mean of the value calculated for all the rows of the matrix of a given subject, with N equal to the cumulative number of trials for that subject. Using the standard Box-Müller algorithm for generating pseudorandom numbers with the normal distribution, we simulated categorical matrices using the average row variance obtained from an experiment we had previously conducted.

To allow in our simulations of the confusion matrix for the skewness produced by anchor effects, we take a reflection of the distribution at each side of the matrix. For example, if we are dealing with the first row, the distribution is centered about element ($j = 1$, $k = 1$). Hence, any response which, in the absence of edges, would have normally been placed in element (1, 0), are lumped into element (1, 1); correspondingly, any response in (1, −1) is placed into (1, 2), etc. In this way, we can, at least approximately, account for the asymmetrical distributions near the edges.

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
x_1	36	4	1	1	0	0	0	0	0	0
x_2	10	17	21	3	0	0	0	0	0	0
x_3	2	12	13	17	4	0	0	0	0	0
x_4	1	5	6	9	11	2	0	0	0	0
x_5	0	0	1	16	22	20	0	0	0	0
x_6	0	0	0	2	8	25	13	0	0	0
x_7	0	0	0	3	5	11	27	5	0	0
x_8	0	0	0	0	0	5	14	30	4	0
x_9	0	0	0	0	0	0	2	15	29	4
x_{10}	0	0	0	0	0	0	0	0	11	33
y_k^{total}	49	38	42	51	50	63	56	50	44	37

Fig. 3. Typical experimental results for the categorical matrix obtained from a single subject *B*. Total number of points is $N = 480$ and the number of categories is $m = 20$. The stimulus intensity ranged from 11–90 dB. Notice that the points clustered about the main diagonal (the N_{ii} elements for perfect response). Farther from the edges, one can see the tendency toward the normal distribution of the point densities in each row. Furthermore, the distribution of responses, $\tilde{p}(y_k)$, is approximately uniform, as indicated by the tabulated values of y_k^{total}, listed at the bottom.

3. Results

3.1. Phase 1: Validating the simulation over a range of values of N for which experimental values of $\mathscr{I}_t$ were available

We simulated for $N = q$ trials and compared with experimental $\mathscr{I}_t$ for $N = q$. In Fig. 4, the comparison was made for $1 \leq q \leq 500$. Since the experimental data were analyzed with 90 categories, we took $m = 90$ in the simulation. The geometric mean row variance was calculated to be 14.1. The simulated $\mathscr{I}_t$ curve is in good agreement with the measured data curve as demonstrated in Fig. 4.

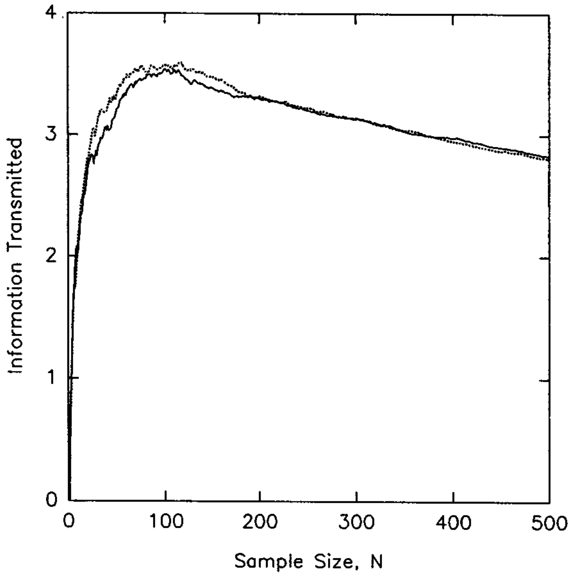

Fig. 4. Results of a simulation are compared to measured values. The simulated curve (solid curve) was generated using the average row variance obtained from a subject's responses (subject J). The dashed curve indicates the information as evaluated from the measured data of the same subject. Please see also Fig. 5.

3.2. Phase 2: Utilizing the simulation to overcome the bias of small N

We now increase N to values of 10^5 or greater. This process would require months or years to complete experimentally. The simulator, however, is ideally suited for such a task. In Fig. 5, $\mathscr{I}_t$ is plotted for a large range of N using the variance 14.1 for all rows. The same variance was used to generate the graphs in both Fig. 4. and Fig. 5.

In agreement with the theoretical results as obtained, for example, by Miller (1955), $\hat{\mathscr{I}}_t$ overestimates $\mathscr{I}_t$. If we were to now require an experimental determination of $\hat{\mathscr{I}}_t$, say, within 5% of the asymptotic value, we would require at least a sample size of 10^4 observations, as shown in Fig. 5. Clearly, the simulator becomes a viable and useful tool for exploratory work in absolute judgements.

4. Discussion

To understand why the simulator works, we return to $H(Y|X)$ in Eq. (15):

$$H(Y|X) = -\sum_j p(x_j) \sum_k p(y_k|x_j) \log p(y_k|x_j). \tag{16}$$

Recall that $p(x_j)$ is a constant $= 1/m$. Hence, the first summation can be approximated to obtain

$$H(Y|X) = -\frac{1}{m} \sum_j \sum_k p(y_k|x_j) \log p(y_k|x_j). \tag{17}$$

The sum over k represents a discrete entropy of $p(y_k|x_j)$. Since the errors made by the subject are assumed to lie on a normal distribution, $p(y_k|x_j)$ represents a discrete approximation to the normal distribution.

It is easily shown (Norwich, 1993) that the entropy of such a discrete distribution will differ from the entropy of the continuous normal distribution only by a constant which is determined by Δx, the width of each category. A partial proof is provided in the appendix.

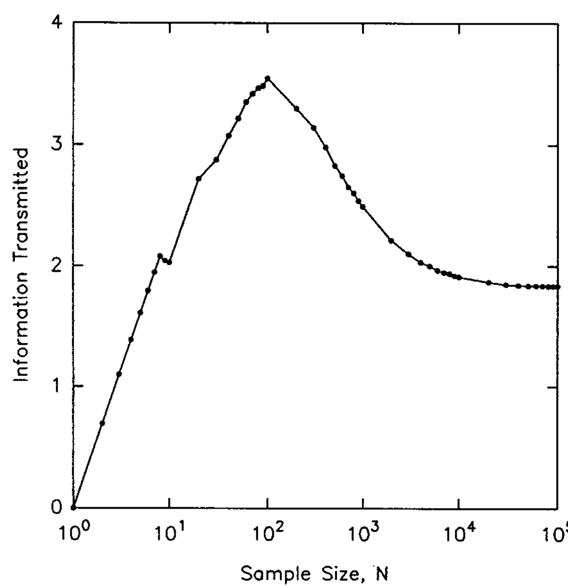

Fig. 5. Graph showing simulated values (●) of information transmitted for large values of N. The same variance was used in both simulations for Fig. 4 and Fig. 5.

When $p(y_k|x_j)$ is approximated with the continuous normal distribution we find the particularly simple result

$$H(Y|X) = \frac{1}{m}\sum_j (1/2)\ln(2\pi e\sigma_j^2) - \ln \Delta x, \tag{18}$$

where σ_j^2 is the variance of the normal distribution in row j. For the remainder of the discussion, we shall continue to work with natural logarithms. The entropies will be expressed in natural units.

Since we assumed that the variance stays approximately constant over all rows, we replace σ_j^2 with the geometric average σ^2 to obtain

$$H(Y|X) = (1/2)\ln(2\pi e\sigma^2) - \ln \Delta x \tag{19}$$

As N, the total number of trials increases, $p(y_k|x_j)$ tends more closely toward the normal distribution while retaining nearly the same variance. Using the calculus of variations, Shannon and Weaver (1949) were able to demonstrate that, for a given variance, the normal distribution gives the largest entropy. Thus, the equivocation, $H(Y|X)$ approaches its maximum value as $p(y_k|x_j)$ approaches the continuous normal distribution.

Since $Y = X + R$ (Shannon and Weaver, 1949), therefore $p(y_k)$ the probability of response k, is obtained by the convolution of $p(x_j)$ (which is approximated by the uniform density) with the normal distribution centered about x_j. However, as the number of categories increases, the convolution can be approximated by the simple function $p(y_k) \simeq 1/m$ (ignoring anchor effects). Therefore, $H(Y) = \ln m - \ln(\Delta x)$, and we see immediately from Eq. (15) that

$$\mathscr{I}(Y|X) = \ln m - \left(\frac{1}{2}\right)\ln(2\pi e\sigma^2) \tag{20}$$

Thus, this equation will predict the final unbiased value of $\mathscr{I}_t$ under approximate conditions. In our case, we have $\Delta x = 1$ because the categories always increment in units of one. As a numerical example, if we take the parameters used to generate the simulated curve in Fig. 5 ($m = 90$ and $\sigma^2 = 14.1$), Eq. (20) predicts that $\mathscr{I}(Y|X)$ is approximately 1.76 n.u. or 2.54 bits. The asymptotic value as found in Fig. 5 for 10^5 trials is 1.82 n.u. or 2.63 bits. The discrepancy is less than 4%.

If $\Re$ is the range of stimuli (in dB for the matrices we used), and Δx the difference in stimulus value for adjacent categories, then

$$\Re = m\Delta x. \tag{21}$$

Substituting this value for m in Eq. (20), we obtain

$$\mathscr{I}_t = \ln(\Re/\sigma\Delta x) - (\tfrac{1}{2})\ln(2\pi e). \tag{22}$$

Since both the range of hearing and the average row variance are independent of m, we see that $\mathscr{I}_t$ is constant as well. Thus, we have demonstrated a theoretical expression for the channel capacity in absolute judgments. It is instructive to observe that apart from the constant (1/2)ln(2πe), $\mathscr{I}_t$ is essentially a logarithmic measure of the ratio of the range of hearing to the standard deviation in response.

5. Conclusions

We may conclude based on these simulations that the order of 20–30 000 experimental trials with a human subject are needed in order to determine the maximum quantity of information transmitted by perceiving a tone whose intensity may span the complete range of audibility. Since in our experience a subject can perform only about 200 trials per day without undue fatigue, the required number of trials could require as many as 100–150 days of experimentation. Other investigators have pressed subjects to obtain several times as many trials per day without apparent deterioration of performance, but many days of work are still required. In the past, this problem has been overcome in part by pooling of data from several subjects, resulting in a blurring of the data from individual subjects. Attempts have also been made in the past to calculate analytically the bias expected from a restricted number of trials performed on a single subject, and in this way correct a calculation of transmitted information based on a small number of trials. However, such analytical mathematical solutions have been inadequate due to the complexity of the problem.

We have utilized a computer simulation to perform the majority of the 30 000 trials required to calculate information transmission for a single subject. The primary parameter used in the simulation is the subject's measured variance in each category of sound intensity.

It has been observed that the variance does not change systematically when evaluated for increasing numbers of trials. The simulated values for information transmitted match the measured quantities of information over the range of N, the number of trials, in which human subjects partook. For example, if a human subject took part in 500 trials, transmitted information simulated for $1 \leq N \leq 500$ matched the measured values well. We expect that the simulated values for $501 \leq N \leq 10^5$ are those which would have been obtained had the human subject continued.

In the process of devising this computer simulation, we gained further insight into the reason why restricted numbers of trials tend to overestimate the transmitted information. The equivocation, or loss of information, increases as the histogram of human error approaches the normal distribution of errors. The source entropy also increases, but tends to equilibrate more rapidly. The difference gives rise to the usual overestimation.

Following Houtsma, we have found that the computer simulation provides a simple approach to a rather complex problem that has lingered for some years. It has enabled us to make more accurate measurements of the limits of human perception.

Acknowledgements

We are greatly indebted to one of the reviewers, Prof. Lawrence Ward, for his many helpful suggestions on improving this paper, and to Prof. Shuji Mori for helping us locate relevant background literature. We also wish to acknowledge the assistance of Jordy Koffman in carrying out our experiments. This work has been supported by the Natural Sciences and Engineering Research Council of Canada.

Appendix A. The continuous measure of information

We begin with Eq. (9) and replace the discrete probability function $p(x_j)$ with the probability density function $\tilde{p}(x_j)$, to obtain

$$\begin{aligned} H(X) &= -\sum_j \tilde{p}(x_j)\,\Delta x\,\ln[\tilde{p}(x_j)\,\Delta x] \\ &= -\sum_j \tilde{p}(x_j)\ln[\tilde{p}(x_j)]\,\Delta x \\ &\quad -\sum_j \tilde{p}(x_j)\ln[\Delta x]\Delta x, \qquad \text{(A1)} \end{aligned}$$

where $\Delta x = x_{j+1} - x_j$. Since Δx is constant and $\Sigma_j \tilde{p}(x_j)\,\Delta x = 1$, this equation simplifies to

$$H(X) = -\sum_j \tilde{p}(x_j)\ln[\tilde{p}(x_j)]\,\Delta x - \ln[\Delta x]. \qquad \text{(A2)}$$

If Δx is small (the bin size is small compared with the full range), we can approximate the sum by an integral. Thus,

$$H(X) \simeq -\int_{-\infty}^{\infty} \tilde{p}(x)\ln \tilde{p}(x)\,\mathrm{d}x - \ln[\Delta x], \qquad \text{(A3)}$$

where $\tilde{p}(x)$ is the continuous analog of the discrete probability function $p(x_j)$. A similar expression can be derived for $H(Y|X)$. Regarding the limits of integration, we have defined $\tilde{p}(x) = 0$ for $x < x_1$ and $x > x_m$.

References

Carlton, A.G., 1969. On the bias of information estimates. Psychol. Bull. 71, 108–109.

Garner, W.R., Hake, H.W., 1951. The amount of information in absolute judgments. Psychol. Rev. 58, 446–459.

Garner, W.R., 1953. An informational analysis of absolute judgments of loudness. J. Exp. Psychol. 46, 373–380.

Garner, W.R., 1962. Uncertainty and Structure as Psychological Concepts. Wiley, New York.

Houtsma, A.J.M., 1983. Estimation of mutual information from limited experimental data. J. Acoust. Soc. Am. 74, 1626–1629.

Miller, G.A., 1955. Note on the bias of information estimates. In: Quastler, H. (Ed.), Information Theory in Psychology: Problems and Methods. Free Press, Glencoe, IL, pp. 95–100.

Miller, G.A., 1956. The magical number seven, plus or minus two. Psychol. Rev. 63, 81–97.

Mori, S., 1991. Dynamic Information Model of Identification Performance. Ph.D. thesis, University of British Columbia.

Mori, S., Ward, L.M., 1996, personal communication.

Norwich, K.H., 1993. Information, Sensation and Perception, Academic Press, San Diego CA.

Rogers, M.S., Green, B.F., 1955. The moments of sample information when the alternatives are equally likely. In: Quastler, H. (Ed.), Information Theory in Psychology: Problems and Methods. Free Press, Glencoe, IL, pp. 101–108.

Shannon, C.E., Weaver, W., 1949. The Mathematical Theory of Communication, University of Illinois Press, Urbana, IL.