Jackknife and Bootstrap Bias Correction for Single-Subject Information Transfer in Audiological Testing

John Castelloe and George Woodworth
Department of Statistics and Actuarial Science, University of Iowa
February 14, 1996

1. INTRODUCTION

Percent of information transferred (IT) is the average reduction in stimulus entropy given a subject's response, expressed as a percent of total stimulus entropy.

Miller and Nicely (1955) advocated the use of the information transfer IT statistic as a measure of the amount of speech information transmitted from a speaker to a listener. The technique was subsequently widely adopted in speech and hearing research as a tool for investigating the capacity of various hearing augmentation devices to enhance or restore the perception of particular speech features by hearing impaired individuals. *Inter alia*, Von Wallenberg, Hochmair and Hochmair-Desoyer (1990), Dorman, et al. (1990), Van Tasell, et al. (1992), Mckay and McDermott (1993), Summers, et al. (1994), and Cazals, et al. (1994) all report information transfer statistics.

The Iowa consonant confusion test (Tyler, et al., 1983) is a typical example of the type of stimulus-response paradigm used to estimate information transfer. This test involves 12 presentations of each of 10 consonants (p,t,k,c,b,d,v,z,n,m) in completely random order. The results of a single administration can be displayed as a confusion matrix similar to Figure 1.

Response

Stimulus		p	t	k	c	b	d	v	z	n	m
	p	0	2	6	0	2	1	0	0	0	1
	t	0	3	2	0	1	1	1	4	0	0
	k	0	1	3	3	0	0	1	2	1	1
	c	1	1	2	3	2	0	1	0	1	1
	b	1	0	3	0	2	0	1	2	3	0
	d	1	0	5	0	0	2	0	0	0	4
	v	1	2	1	0	0	3	2	1	2	0
	z	1	2	0	0	2	0	3	0	2	2
	n	0	1	0	0	0	0	0	0	4	7
	m	0	0	0	0	0	0	0	0	4	8

Figure 1: A consonant confusion matrix. The consonants p, t, k, c, b, d, v, z, n, and m were presented aurally 12 times each in random order to a hearing impaired subject. Columns indicate the subject's report of what he heard. For example for the 12 presentations of "p" this subject heard "t" and "b" twice, "k" six times, and "d" and "m" once .

Each row of the confusion matrix corresponds to a phoneme stimulus presented aurally to the subject. The columns correspond to the subject's attempts to identify each stimulus. Generally, each stimulus is presented the same number of times (k). In this example, $k = 12$.

Our notation for the contents of the confusion matrix is shown in Figure 2. Here n denotes the number of stimuli ($n = 10$ in Figure 1) and c_{sr} is the number of times the subject reported hearing stimulus r when stimulus s was presented.

$$\text{Stimulus}\ \overset{\text{Response}}{\begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2n} \\ c_{31} & c_{32} & c_{33} & \cdots & c_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & c_{n3} & \cdots & c_{nn} \end{bmatrix}}$$

Figure 2: Confusion matrix notation.

Row, column, and grand totals for the confusion matrix are denoted,

$$\sum_{s=1}^{n} c_{sr} = c_{+r} \ , \quad \sum_{r=1}^{n} c_{sr} = c_{s+} \ , \quad \sum_{s=1}^{n}\sum_{r=1}^{n} c_{sr} = c_{++} \quad .$$

Our theoretical analysis of information transfer rests on some assumptions regarding the statistical properties of a subject's responses. We assume that the rows of the confusion matrix have a product multinomial distribution. That is, the rows of the confusion matrix are statistically independent, and each row vector $(c_{s1},\ldots,c_{sn})$ has a multinomial distribution with parameters $(k, p_{s1},\ldots, p_{sn})$.

The symbol p_{sr} is the probability that the listener will give response r to stimulus s. The assumption that rows are independently distributed multinomial random vectors implies that response probabilities do not change over the duration of a test session and are not affected by context within a session.

The expected number of confusions of r for s is denoted $f_{sr} = E(c_{sr}) = kp_{sr}$. The matrix of expected confusions $\{f_{sr}\}$ is the listener's *true confusion matrix*, while the matrix of observed confusions $\{c_{sr}\}$ is the listener's *observed confusion matrix*, or simply, the *confusion matrix*.

2. THE INFORMATION TRANSFER STATISTIC

Miller and Nicely's (1955) information transfer statistic (IT) is the reduction in average stimulus entropy given the subject's responses expressed as a fraction (or percent) of the total stimulus entropy. Miller and Nicely refer to entropy as mean logarithmic probability (MLP).

When expressed in base-2 logarithms, the entropy of a text can be interpreted as the minimum number of bits (binary digits) needed to transmit the text in compressed form. (In our example the text consists of 12 repetitions of 10 consonants in random order.) Given a degraded version of the text (for example a subject's report of what he or she heard) the residual entropy is the number of additional bits needed to correct the degraded text. The information transferred by the degraded text is the total entropy minus the residual entropy.

Mathematically, the stimulus entropy is,

$$\mathrm{H(S)} = -\sum_{s=1}^{n} \frac{f_{s+}}{f_{++}} \log\left(\frac{f_{s+}}{f_{++}}\right),$$

and the residual entropy is,

$$\begin{aligned}\mathrm{H(S|R)} &= \sum_{r=1}^{n} \frac{f_{+r}}{f_{++}} \mathrm{H}(\mathrm{S|R}{=}r) \\ &= -\sum_{r=1}^{n} \frac{f_{+r}}{f_{++}} \sum_{s=1}^{n} \frac{f_{sr}}{f_{+r}} \log\left(\frac{f_{sr}}{f_{+r}}\right) \\ &= -\sum_{r=1}^{n} \sum_{s=1}^{n} \frac{f_{sr}}{f_{++}} \log\left(\frac{f_{sr}}{f_{+r}}\right)\end{aligned}$$

Total information transmitted is $\mathrm{H(S)} - \mathrm{H(S|R)}$; consequently the proportion of information transmitted is,

$$\mathrm{IT} = 1 - \frac{\mathrm{H(S|R)}}{\mathrm{H(S)}} \qquad (1)$$

If the subject responds perfectly, then $f_{ss} = k$ and $f_{sr} = 0,\ \textit{for } s \neq r$. In this case it is easy to see that $\mathrm{H(S|R)} = 0$, and consequently $\mathrm{IT} = 1$. Conversely, if the subject responds randomly, then $f_{sr} = k/n$, in which case it is easy to show that $\mathrm{H(S|R)} = \mathrm{H(S)}$, so that $\mathrm{IT} = 0$.

We will refer to the information transfer statistic computed from the expected confusion matrix as the *true information transfer* (IT). The IT statistic computed from an observed confusion matrix will be called the *sample information transfer* ($\hat{\mathrm{IT}}$). Sample information transfer is intended to estimate true information transfer.

Miller and Nicely (1955) knew that sample information transfer was biased upward and recommended that the ratio of presentations to stimuli (k/n) should exceed 2 for reliable IT estimation. However, we have investigated the magnitude of this bias in confusion matrices typical of cochlear implant wearers for whom information transfer numbers under 20% are not uncommon. In this case we have found that the bias can exceed 100% even when *k/n*=4 (See Figure 3).

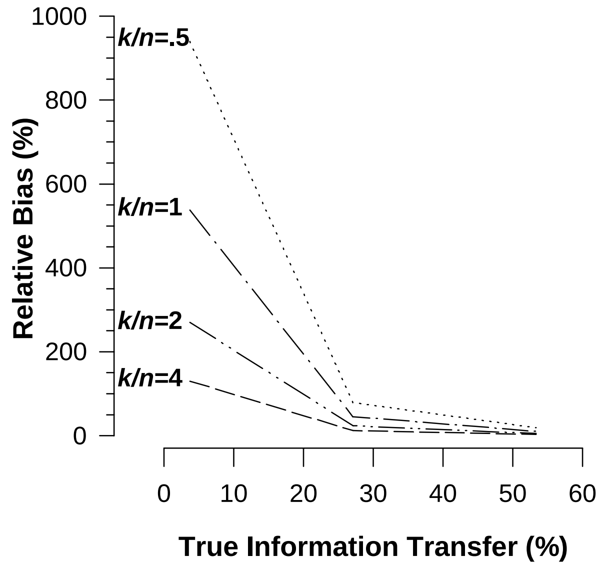

Figure 3. Relative bias of the sample information transfer, ($100 \cdot E((\hat{IT} - IT)/IT)$), based on 1000 simulations of k repetitions of n=10 stimuli for three hypothetical subjects with true information transfer between 3% and 50%.

To achieve $k/n = 4$ in our example, a subject would be required to respond to 40 presentations of 10 consonants, a total of 400 presentations -- requiring about 25 minutes to complete at the Iowa Cochlear Implant Project. In clinical practice it would not be practical to ask a subject to respond to that many stimulus presentations. Therefore, we have investigated the use of two statistical methods of bias reduction to improve the validity of the sample information transfer for clinically practical values of k/n: Quenouille's jackknife and Efron's bootstrap. This paper reports the results of simulation studies of the performance of the bootstrap and jackknife.

3. THE JACKKNIFE AND BOOTSTRAP

The bias of a statistic is the difference between its expected sample value and its true value as computed from the population. The key idea of the bootstrap is to approximate the bias by resampling from a sample. In other words, the obtained sample is used as a proxy for the population, and samples from the sample are used as proxies for repeated sampling from the population. The difference between the observed statistic and its mean over all possible resamples provides an estimate of the bias of the statistic compared to the population. This bias estimate is then subtracted from the statistic to reduce its bias.

Suppose that S is the statistic based on the observed sample and that S^* is the same statistic computed from a sample of the same size drawn with replacement from the observed sample. The bootstrap estimate is,

$$S_{boot} = S - \hat{b}ias_{boot} \ ,$$

where the bootstrap bias correction is,

$$\hat{b}ias_{boot} = E_{resamples}(S^*) - S \ .$$

Bootstrap bias correction is computer-intensive, often requiring hundreds of resamples to obtain a stable estimate. Simulating the performance of the bootstrap is doubly computer intensive as it is necessary to simulate hundreds of hypothetical samples from which hundreds of resamples must be drawn to compute each replication of the bias-corrected statistic.

The less computer-intensive jackknife bias correction for a statistic computed from *N* independent observations involves leaving out observations one at a time. Suppose that S is the statistic based on the complete sample, and $S_{(i)}$ is the statistic computed from the same sample after deleting the i^{th} observation. The bias correction is,

$$\hat{b}ias_{jack} = \frac{(N-1)}{N} \sum_{i=1}^{N} \left(S_{(i)} - S \right) \qquad (2)$$

We made a small simulation study to provide a rough assessment of the comparative performance of the bootstrap and jackknife bias corrections for the information transfer statistic. We used three hypothetical subjects with consonant information transfer numbers of 3.7%, 27.1% and 53.3%, respectively. Their true confusion matrices were synthesized by pooling confusion matrices from several cochlear implant wearers participating in the Iowa Cochlear Implant project. True information transfer values were computed using (1). The three true (expected) confusion matrices, along with their corresponding IT values, are shown in Figure 4.

In our simulation studies, sample confusion matrices were randomly generated for $k = 5$, 10, 20, and 40 repetitions of each stimulus using the multinomial assumptions discussed above (Figure 5). Multinomial probabilities are computed as,

$$p_{sr} = f_{sr} \ / \ f_{s+}$$

For each of the three hypothetical subjects, and for each of four different values of *k*, a sample of $M = 15$ confusion matrices was generated. The information transfer for each of these 180 matrices was estimated using three different methods: raw, uncorrected $\hat{IT}$, jackknife-corrected $\hat{IT}_{jack}$, and bootstrap-corrected $\hat{IT}_{boot}$. The relatively small number of replications (15) was dictated by the extensive computing time required for generating 300 bootstrap re-resamples for each of 180 samples; however, it was sufficient to indicate that the jackknife gives superior bias correction (Figure 6).

	Subject 1 (IT= 3.7%) Response										Subject 2 (IT= 27.1%) Response										Subject 3 (IT=53.3%) Response									
Stimulus	p	t	k	c	b	d	v	z	n	m	p	t	k	c	b	d	v	z	n	m	p	t	k	c	b	d	v	z	n	m
p	6	6	7	8	8	12	10	3	4	2	8	6	1	1	2	3	0	2	0	1	10	6	2	0	0	0	0	0	0	0
t	9	8	4	11	6	10	5	6	5	2	2	9	1	4	1	3	0	3	0	1	3	14	0	0	0	1	0	0	0	0
k	3	11	9	9	6	8	6	6	2	6	5	4	5	4	0	4	0	1	1	0	6	6	4	0	0	0	1	1	0	0
c	4	7	4	15	3	8	1	12	8	4	0	1	1	18	0	2	1	1	0	0	0	0	0	16	0	0	0	2	0	0
b	6	10	5	8	8	7	5	5	3	9	6	0	1	2	8	3	2	0	0	2	1	0	0	0	6	1	7	2	0	1
d	6	8	6	4	6	7	6	8	8	7	0	2	1	3	4	10	1	1	0	2	0	0	0	0	1	9	6	2	0	0
v	5	3	7	9	8	5	13	4	6	6	5	0	2	5	3	4	2	2	0	1	0	0	0	0	0	4	9	5	0	0
z	7	1	7	6	3	11	7	8	13	3	3	1	2	5	0	2	1	3	5	2	0	0	0	1	0	3	3	11	0	0
n	6	8	9	2	4	6	8	5	9	9	2	0	1	0	1	0	0	0	4	16	0	0	0	0	0	0	0	2	10	6
m	4	7	6	3	5	7	5	6	14	9	1	0	0	0	0	0	0	0	7	16	0	0	0	0	0	1	0	1	5	11

Figure 4. True expected confusion matrices for three hypothetical subjects used in the simulation study. Matrices were developed by aggregating data from several cochlear implant wearers participating in the Iowa Cochlear Implant Project.

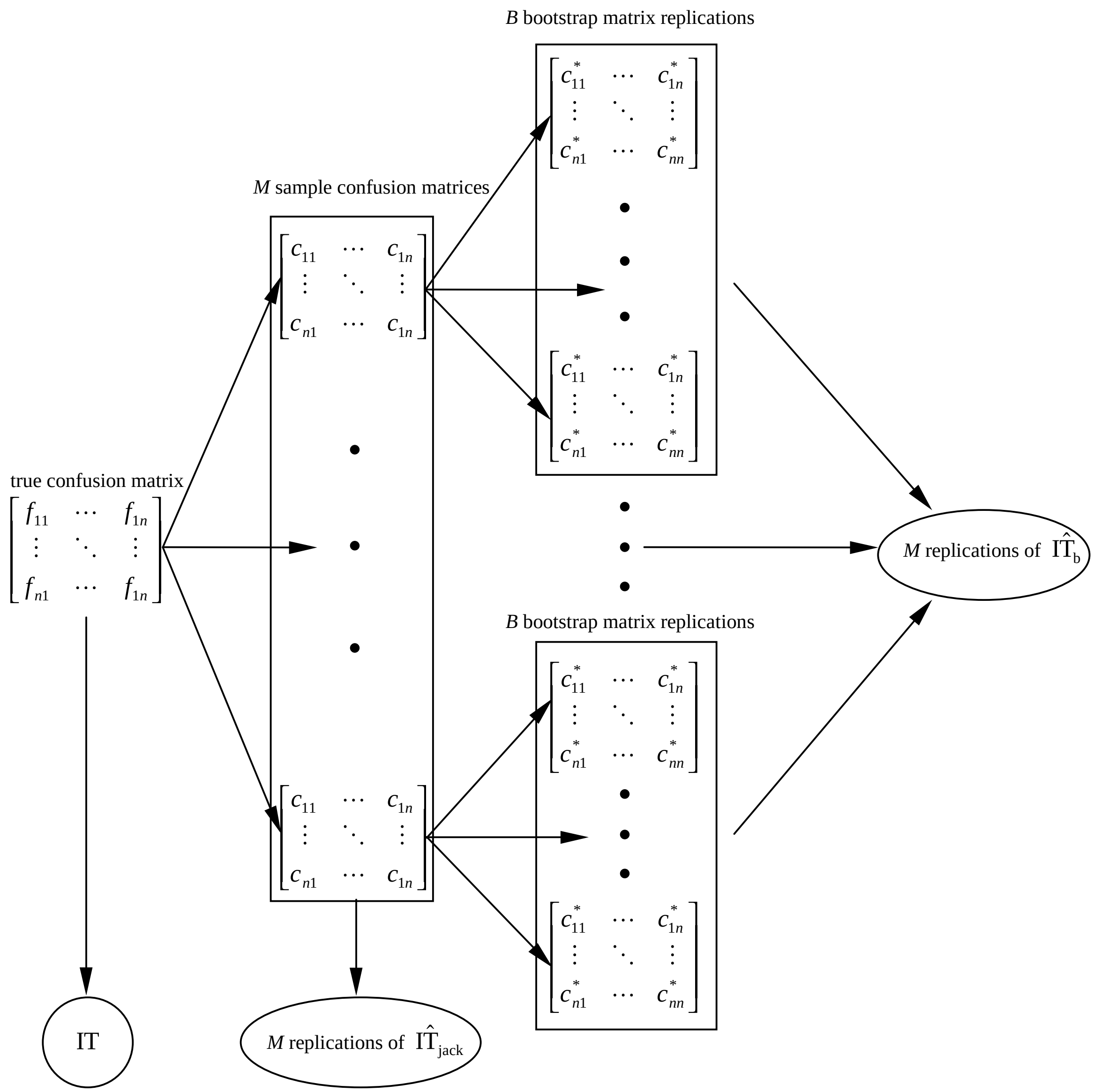

Figure 5. Design of the simulation study. M sample confusion matrices were generated by multinomial sampling from each of 3 true confusion matrices (Figure 4). M=15, supplemented to 1000 for the jack-knife. For each sample confusion matrix B=300 bootstrap re-samples were drawn to compute the boot-strap bias correction.

We supplemented the small simulation study with 1000 additional replications for the uncorrected and jackknife-corrected information transfer statistic. The results are shown in Figure 6. We derived the estimate of the bias of $\hat{IT}$ for a given sample confusion matrix **C** by treating **C** as a function of $N = nk$ consonant presentations. That is, letting $\mathbf{C_{sr}}$ denote the observed confusion matrix **C** with entry c_{sr} replaced by $c_{sr} - 1$ for $c_{sr} > 0$. Then from (2),

$$\hat{\text{bias}}_{\text{jack}} = (N-1)\left(\frac{1}{N}\sum_{c_{sr}>0} c_{sr}\hat{IT}(\mathbf{C_{sr}}) - \hat{IT}(\mathbf{C})\right). \quad (3)$$

The jackknife bias-corrected estimate of the information transfer estimate is given by,

$$\hat{IT}_{\text{jack}} = \hat{IT} - \hat{\text{bias}}_{\text{jack}} \quad (4)$$

(Efron and Tibshirani, 1993, p 138).

A sample confusion matrix can be interpreted as $N = nk$ observations of a subject's response to a stimulus. A jackknife sample is determined from the original collection of observations by removing one of these responses. A new confusion matrix is formed with the remaining $N-1$ observations. The information transfer of this new matrix is calculated to give one jackknife replication of the information transfer statistic. The process is repeated N times, once for each of the nk observations. The difference between the average of these N replications and the information transfer of the original matrix **C** is then multiplied by $N-1$ to give the jackknife estimate of bias. Note that the jackknife replications corresponding to any two observations in the same cell of the matrix will be equal. Thus, it is only necessary to calculate one jackknife replication for each non-zero cell in the original confusion matrix. **C.** The following example should make these points clear.

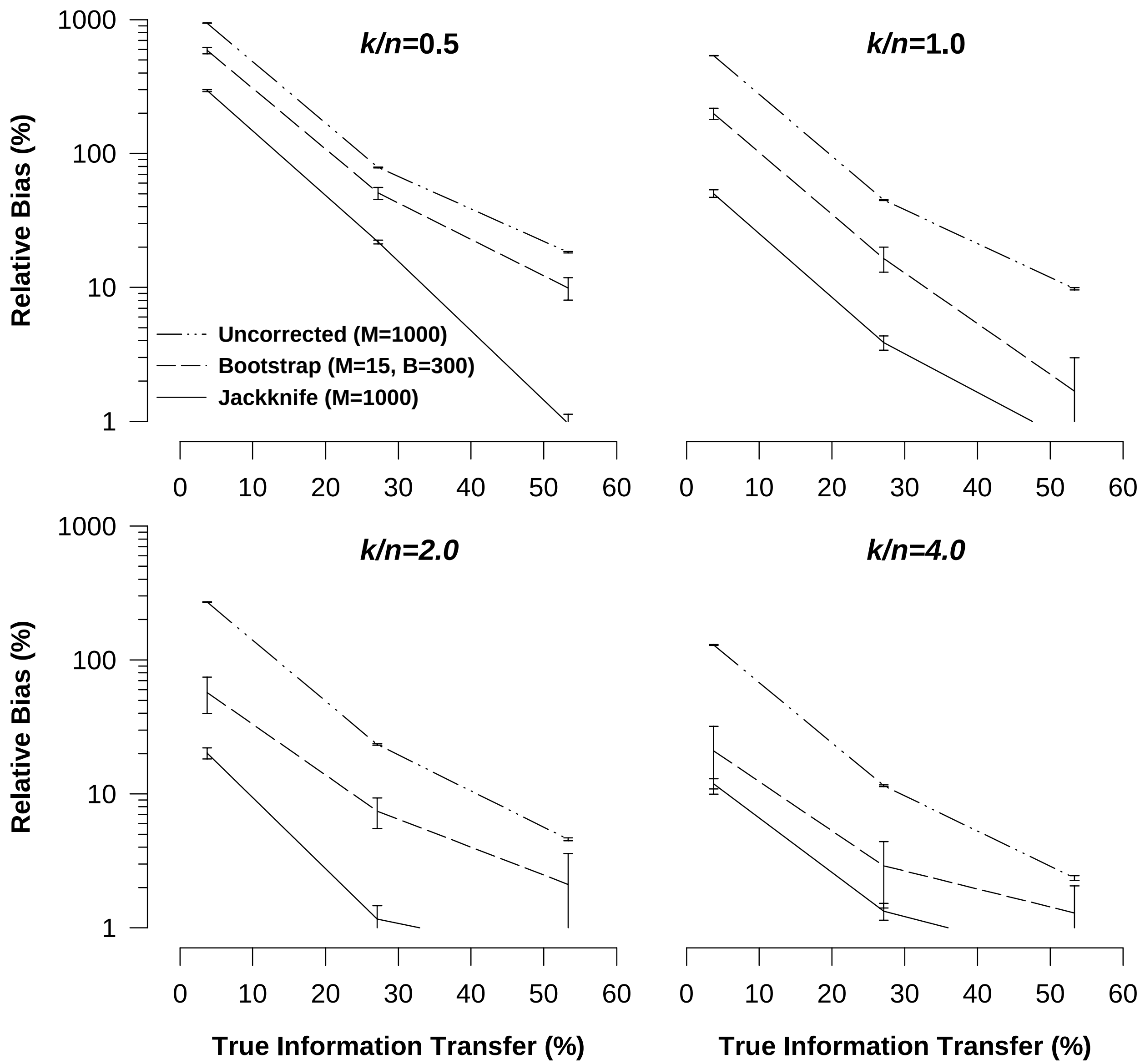

Figure 6. Relative absolute bias ($100 \cdot |IT - E(\hat{IT})| / IT$*) of uncorrected, jackknife-corrected, and bootstrap-corrected Information Transfer. Based on 1000 (uncorrected, jackknife) or 15 (bootstrap, B=300)) simulated confusion matrices for three hypothetical hearing-impaired subjects. Error bars are plus or minus one standard error.*

Let $\mathbf{C} = \begin{bmatrix} 3 & 1 \\ 0 & 4 \end{bmatrix}$ be a sample from a true confusion matrix, and let IT be the true IT. Note that $n = 2$ and $k = 4$. Using (1), we estimate IT as $\hat{\mathrm{IT}}(\mathbf{C}) = 0.5488$. To compute $\hat{\mathrm{IT}}_{\mathrm{jack}}$ we remove one observation at a time. When the first observation is removed, the resulting confusion matrix is $\hat{\mathrm{IT}}_{\mathrm{jack}}$. Removing any one of the first three observations will yield this same confusion matrix. Removing the fourth observation yields $\mathbf{C_{12}} = \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}$. Finally, removing any one of the last four observations yields $\mathbf{C_{22}} = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$. Thus, the average of the 8 jackknife replications of information transfer is computed as

$$\frac{1}{8}\left(3\hat{\mathrm{IT}}(\mathbf{C_{11}}) + \hat{\mathrm{IT}}(\mathbf{C_{12}}) + 4\hat{\mathrm{IT}}(\mathbf{C_{22}})\right)$$
$$= \frac{1}{8}\left(3(0.4766) + 1.000 + 4(0.5295)\right) = 0.5685$$

From (3) the jackknife estimate of bias is then given by,

$$\hat{\mathrm{bias}}_{\mathrm{jack}} = (2(4)-1)(0.5685 - 0.5488) = 0.1379 .$$

The jackknife bias-corrected estimate of information transfer from (4) is,

$$\hat{\mathrm{IT}}_{\mathrm{jack}} = 0.5488 - 0.1379 = 0.4109 ,$$

which is an improved estimate of IT.

4. DISCUSSION

Uncorrected estimates of small IT values are grossly biased. For IT = 3%, relative biases of 300% are possible even when $k/n = 2$. The most surprising finding is the poor performance of the bootstrap. This procedure generally outperforms the jackknife (Efron and Tibshirani 1993, p 145-6), and its behavior for the IT statistic needs further investigation.

For $k/n \le 1$ and small IT, it is not possible to reduce the bias to acceptable levels. For $k/n \ge 2$, the jackknife reduces the relative bias to 20% or less for IT ≥ 3% and to less than 2% for IT ≥ 30%. The bootstrap performs consistently worse than the jackknife for all values of k/n and IT. Since the bootstrap is considerably more computationally intensive, it is therefore not a useful option.

Jackknife-corrected information transfer numbers have less than 20% relative bias for IT as small as 3% when $k/n \ge 2$, as recommended by Miller and Nicely. Since increasing k/n to 4 does not materially reduce the bias, there is only a slight benefit in using $k/n > 2$. Thus the Miller-Nicely rule, when combined with jackknife bias correction, appears to be a reliable rule of thumb.

Uncorrected IT estimates are extremely biased for small IT values and should not be used without jackknife correction. Since the jackknife is easily implemented, we recommend that audiological researchers routinely correct IT numbers for finite-sample bias.

REFERENCES

Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (1994), "Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants," *J. Acoust Soc America*, 96 (4), 2048-2054.

Dorman, M.F., Soli, S., Dankowski, K., Smith, L.M., McCandless, G., and Parkin, J. (1990), "Acoustic cues for consonant identification by patients who use the Ineraid cochlear implant," *J. Acoust. Soc. Am.*, 88 (5), 2074-2079.

Efron, B. and Tibshirani, R (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

McKay, C.M., and McDermott, H.J. (1993), "Perceptual Performance of Subjects with Cochlear Implants Using the Spectral Maxima Sound Processor (SMSP) and the Mini Speech Processor (MSP)," *Ear and Hearing*, 14 (5), 350-367.

Miller, G.A., and Nicely, P.E. (1955). "An Analysis of Perceptual Confusions Among Some English Consonants," *J. Acoust. Soc. Am.*, 27, 338-352.

Summers, I.R., Dixon, P.R., Cooper, P.G., Gratton, D.A., Brown, B.H., and Stevens, J.C. (1994), "Vibrotactile and electrotactile perception of time-varying pulse trains," *J. Acoust Soc. Am*, 95 (3), 1548-1558.

Tyler, R., Preece, J., and Lowder, M.(1983) *The Iowa Cochlear Implant Tests*. Iowa City, IA: The University of Iowa Department of Otolaryngology-Head and Neck Surgery.

Van Tassell, D.J., Greenfiled, D.G., Logemann, J.J., and Nelson, D.A. (1992), "Temporal cues for consonant recognition: Training, talker generalization, and use in evaluation of cochlear implants," *J. Acoust. Soc. Am.*, 92 (3), 1247-1257.

Von Wallenberg, E.L., Hochmair, E.S. and Hochmair-Desoyer, I.J. (1990), "Initial Results with Simultaneous Analog and Pulsatile Stimulation of the Cochlea," *Acta Otolaryngol (Stockh)*, Suppl. 469, 140-149.