

## **Information Theory Applied to Animal Communication Systems and Its Possible Application to SETI**

Sean F. Hanser

*Ecology Graduate Group, University of California, One Shields Avenue,  
Davis, CA 95616 USA*

Laurance R. Doyle

*SETI Institute, 2035 Landings Drive, Mountain View, CA 94043 USA*

Brenda McCowan

*School of Veterinary Medicine: Population Health and Reproduction,  
University of California, Davis, 18830 Road 112, Tulare, CA 93274,  
USA*

Jon M. Jenkins

*SETI Institute, NASA Ames Research Center 245-3, Moffet Field, CA  
94035 USA*

**Abstract.** Information theory, as first introduced by Claude Shannon (Shannon & Weaver 1949) quantitatively evaluates the organizational complexity of communication systems. At the same time George Zipf was examining linguistic structure in a way that was mathematically similar to the components of the Shannon first-order entropy (Zipf 1949). Both Shannon's and Zipf's mathematical procedures have been applied to animal communication and recently have been providing insightful results. The Zipf plot is a useful tool for a first estimate of the characterization of a communication system's complexity (which can later be examined for complex structure at deeper levels using Shannon entropic analysis). In this paper we shall discuss some of the applications and pitfalls of using the Zipf distribution as a preliminary evaluator of the communication complexity of a signaling system.

### **1. Introduction to Information Theory and Zipf's Law**

Recognizing and characterizing a communication system of unknown structure and function is a challenge that is difficult to meet even qualitatively. In the animal and plant kingdoms, communication comes in such a variety of forms that it is difficult to specify overlying principles that can characterize intelligent and sophisticated communication. Fortunately the mathematical approach of information theory, as first developed by Shannon (1948), does allow the quantification of a signal system's structure regardless of qualitative considerations such as what the individual signal meanings and usages might be. At the same

time that information theory was being developed, Zipf (1949) developed a mathematical tool for examining human languages that was, essentially, a detailed examination of components of the first-order Shannon entropies. Although Zipf did not specifically establish his work in this area as a branch of information theory, it has been recognized as one of the important tools of communication analysis (see the review in McCowan, Hanser & Doyle 1999). This paper will briefly discuss the application of Zipf's statistic in light of information theory, and its possible application to SETI signals.

The Shannon zero-order entropy,

$$H_0 = \log_2 n \quad (1)$$

where  $n$  is the number of signals in the communication system, characterizes the diversity of the communicative repertoire by indicating how many binary decisions are required to characterize a single choice. The zero order entropy for the English alphabet is  $\log_2 27 = 4.75$  bits/signal (26 letters + space = 27 components), for example.

The first order entropy,

$$H_1 = - \sum_{i=1}^n p(A_i) \log_2 p(A_i) \quad (2)$$

is a measure of the distribution of the probabilities of occurrence,  $p(A_i)$ , of each signal – a measure of repertoire complexity. The component  $\log_2 p(A_i)$  represents the information, and its product with the probability gives the entropy of that signaling unit (which could be phonemes, letters, words, whistles, gestures, color changes, dance directions/steps, facial expressions, chemical signals, etc.; see Hanser, McCowan, & Doyle 2000). The entropy of each signaling unit is summed in Equation 2 to give the overall first order entropy of that communication system (the minus sign being required to ensure that the entropic value is positive since probabilities are always  $\leq 1$ , giving a negative logarithm).

The term information is synonymous with entropy, or uncertainty, in virtually all information theory literature (see the review in McCowan et al. (1999) and Hanser et al. 2000). One must be cautioned, therefore, that the term “information” is not used in its usual sense in information theory. As Weaver puts it, “Information theory is not concerned so much with what one *does* say as much as what one *could* say,” using a given communication system (Shannon & Weaver 1949). Obtaining information, in the sense of knowledge, usually constrains choices (as in making an “informed decision”) whereas in information theory “information” stands for the maximization of a number of choices. Herein we will use the clearer term “entropy” to represent this maximization since it aligns, conceptually, with the term as it is used in thermodynamics (with “choices” replacing “microstates.”)

The second order entropy,

$$H_2 = - \sum_{i=1}^n \sum_{j=1}^n p(A_i B_j) \log_2 p(A_i B_j) \quad (3)$$

introduces conditional probabilities into the concept of entropy (1st-order Markov chains). This has sometimes been used to calculate the transmission of informa-

tion between individuals, but has also sometimes been confused with a transmission of knowledge (see Doyle, McCowan, & Hanser 2003). The terms of the third order entropy,

$$H_3 = - \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n p(A_i B_j C_k) \log_2 p(A_i B_j C_k) \quad (4)$$

take into account the conditional probabilistic dependence of the occurrence of the third signal on the preceding two. This pattern continues on to higher-order entropies that measure the complexity (i.e. nested dependence) of the signals on each other.

A special case of the Shannon first-order entropy, as mentioned, is sometimes known as Zipf's principle of least effort (Zipf 1949). He chose to rank components of human language repertoires by their probabilities of occurrence. Plotting the logarithm, base-10, of the frequencies of occurrence of each signal against the logarithm of their rank (1st, 2nd, 3rd, etc.) he found that most of the dozens of languages he plotted in this way yielded a slope of  $-1$ .

This plot has been extended to several non-human signaling systems including porpoises, chickadees, orcas, squirrel monkeys, dolphins, ground squirrels, and even the chemical signals from cotton plants (see reviews in McCowan et al. 1999; Hanser et al. 2000; McCowan, Doyle, & Hanser 2002; Doyle et al. 2003). For human languages, the Zipf relationship was found to be robust for phonemes, letters, and words, but it also fits many other phenomena that rely upon optimization. Because it cannot take into account the complexity of a communication system (i.e. the higher entropic orders) it has been generally discarded as a useful linguistics tool. However, it can be applied both for obtaining a first estimate of the potential complexity of a communication system as well as for measuring the evolution of this complexity. To date, this has been done for bottlenose dolphins and squirrel monkeys (McCowan et al. 2002). Adult bottlenose dolphins show a Zipf slope of  $-0.95$  indicative of additional complexity in their signaling systems (which is being explored at present) while adult squirrel monkey chucks produce a Zipf slope of no greater than about  $-0.6$ , indicative of a less complex signaling system. McCowan et al. (1999, 2002) have also demonstrated that one can use Zipf's technique to study the evolution of complexity of communication with age; the progression of infant babbling to adolescent redundancy to adult complexity (slopes of  $-0.8$  to  $-1.1$ , to  $-1.0$ ) in bottlenose dolphins follows a developmental path identical to that of human speech.

## 2. Possible Applications to ETI Signals

We have stated above that a  $-1$  slope indicates that the communication system may be approaching optimization (i.e. a kind of Huffman coding optimization; Press et al. 1992). It is true that many phenomena can produce such a relationship, but this, too, is often the result of optimization. (While a completely random process should produce a slope of 0, even the natural distribution of numbers produces a minimum slope of about  $-0.2$ , sometimes referred to as "Benford's Law"; Hill 1995).

Slopes approaching  $-1$  may also be generated by random processes ordered in an optimized way. For example, if one combines the orders of entropy of a random process, one can generate a Zipf slope that indicates more complexity than is actually present. (A Zipf plot of single die throws combined with two, three, and four dice throws will approach a  $-1$  slope, for example; see also Hailman & Ficken 1986; Hailman, Ficken, & Ficken 1986). Finally, if the signaling units are undersampled (such as plotting only the dots and dashes of Morse Code instead of combinations of 3 or 4 at a time) one will get an indication of less communication complexity in the signaling system than is actually present (i.e. a slope of about  $-0.3$ ). In this sense, a Zipf plot of parts-of-signals, whole signals, and sequential combinations of signals may be helpful in distinguishing the primary signaling units, under the assumption that the communication system is fairly well optimized. Undersampling (i.e. not enough signals to be statistically stable or ergodic) will conversely tend to indicate more complexity than is present.

A slope of  $-1$ , however, need not reflect that the system has used up all, or even most, of the possibilities for combinations of repertoire components. Consider written English: the average word size is about five letters, allowing (via the binomial equation) for 101 583 combinations or choices (i.e. microstates) for 27 components (26 letters and a space). Those combinations are all possible when we are not following the constraints of English spelling, but overlying rules are applicable, such as 'u' always follows 'q,' never three or more of the same letter in a row, etc. The result is that we have an optimized communication system despite the fact that we use a fraction of the possible five letter microstates. The number of words (some, of course, greater than 5-letters) used by the King James Bible compared to the theoretical pool of possible five-letter words is about 8%, and Shakespeare's word choice represents roughly 30% of the total theoretical pool.

Present radio searches for ETI signals look for a narrow-band (i.e. 1 Hz bandpass) carrier wave signal (see many articles in this volume addressing the SETI.) The modulated signal, with the actual information in it, may require an increase in sensitivity of about  $4\times$  to detect. Once the modulated signals are picked up and correctly classified, a log-log plot might immediately reveal whether the signal is random or worth investigating for higher level of complexity (i.e. higher-order entropic values). Early human language symbols were close to whole words, but as the need to translate and then to transmit emerged, the more efficient way to communicate was to nest information in higher orders of complexity. There is a hierarchy of signaling system complexity, quantified by the highest degree of Shannon entropy, emerging in studies of non-human animal communication systems compared with human communication systems. Many animals (and plants, for that matter) will likely show a first-order entropic value only. Bottlenose dolphins (and many others, possibly) show structure to at least 2nd and likely 3rd-order entropic values. Humans, depending on the signaling system used, show a great degree of "nested" complexity-8th or 9th order for words (i.e. 30 - 40th order letter entropy) (Yaglom & Yaglom 1983). Although meaning might be illusive for some time, one might nevertheless immediately test any ETI signaling system for the highest-order of entropic complexity. It would be sobering to discover a species communicating at an entropic order of 50 or more. We would at least be able to calibrate where our species stands on the evolutionary line of communication complexity.

**References**

- Doyle, L. R., McCowan, B., & Hanser, S. F. 2003, *J. Theoretical Biology*, to be submitted
- Hailman, J. P., & Ficken, M. S. 1986, *Anim. Behav.*, 34, 1899
- Hailman, J. P., Ficken, M. S., & Ficken, R. W. 1986, *Semiotica*, 56, 191
- Hanser, S. F., McCowan, B., & Doyle, L. R. 2000, in *ASP Conf. Ser. Vol. 213, A New Era in Bioastronomy*, ed. G. Lemarchand & K. Meech, 613
- Hill, T. P. 1995, *Statistical Science*, 10, 354
- McCowan, B., Doyle, L. R., & Hanser, S. F. 2002, *J. of Comp. Psychol.*, 116, 166
- McCowan, B., Hanser, S. F., & Doyle, L. R. 1999, *Anim. Behav.*, 100, 409
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes* (Cambridge: Cambridge University Press)
- Shannon, C. E. 1948, *Bell System Tech. J.*, 27, 379, 623
- Shannon, C. E., & Weaver, W. 1949, *The Mathematical Theory of Communication* (Urbana, Illinois: University of Illinois Press)
- Yaglom, A. M., & Yaglom, I. M. 1983, *Probability and Information* (Boston: Reidel)
- Zipf, G. K. 1949, *Human Behavior and the Principle of Least Effort* (Cambridge: Addison-Wesley Press)