Utilizing Information Theory to study the communication system of *Orcinus orca*

Kenna Deyette Smith Lehmann
2111 Quail Ranch Ct
Elverta, CA 95626
(916) 412-9527
kdlehmann@ucdavis.edu

INTRODUCTION

Information Theory

Information Theory, originally designed to study the information encoded in human speech, is now applied to animal communication (McCowan 1999, 2002). Established by Shannon and Weaver (1949), Information Theory is a statistical tool that can be used to determine the structure and organization within communication systems. By analyzing a communication system statistically, Information Theory provides an objective platform on which to compare and contrast any species' communication system.

McCowan et al. (1999) summarized Shannon and Weaver's paper and explained how Information Theory can be used to evaluate the complexity present within call patterns at many different levels. For example, one could evaluate the complexity of letter patterns, or one could evaluate the higher levels of complexity in word or sentence sequences. The first level of complexity is best evaluated using Zipf's statistic (McCowan et al. 1999). Zipf's statistic considers the optimum amount of redundancy and diversity necessary to communicate. A communication system needs a diversity of signals to portray a large amount of information. When a repertoire has too much diversity it becomes redundant, with too many signals coding for the same message. At the same time, noise also affects the communication channel. A channel can be defined by the medium through which it is passed, be it the auditory, visual, tactile or olfactory channel and by the habitat that the species lives in. In the auditory channel, the amount of background noise will vary with the habitat and may interfere with any transmitted signal. In channels with more noise, a signal may need to be repeated to ensure that it is received. This repetition reduces the amount of information that is being transferred. A communication system must possess a balance between repetition and diversity to achieve the optimum transfer of information adapted to the channel of communication (Shannon and Weaver 1949, McCowan et al. 1999).

Zipf's statistic is a graphical representation of a species' optimal balance between redundancy and diversity (McCowan et al. 1999, Zipf 1949). By ranking the signals within a repertoire by frequency and plotting the log of the rank against the log of the frequency, you can compare a species' slope to an optimal information transfer of -1.00 (McCowan et al. 1999, Zipf 1949). The optimal balance between redundancy and diversity will shift for different species depending on their differing ecologies. In a habitat where visual cues are easy to use (i.e. open fields) or in a species that is highly solitary, very little information needs to be passed in the auditory channel and little diversity is necessary. The Zipf Statistic for these species will be steeper than -1.0 (less diversity, more repetition). In species where visual communication is unreliable (i.e. forest or aquatic habitats) or in species that are highly social and require large amounts of information to be passed, the slope will be near -1.0 to optimize the amount of information being transferred. In addition, when comparing age groups within a species, the Zipf statistic will be greater than -1.0 and will shift closer to -1.0 as the individuals mature (McCowen, Hanser et al. 1999). This phenomenon occurs because after birth, young animal's calls are more diverse until they learn the extent of the adult repertoire. It can be hypothesized that the noisier the channel a species is communicating in, the lower its Zipf statistic (McCowan et al. 1999, Zipf 1949). In a noisier channel, a sender may need to repeat its calls to insure that the receiver receives the information (McCowan, Doyle et al. 2002). This would lead to a higher repetition rate in calls and a steeper slope of the regression line (McCowan, Doyle et al. 2002).

Zipf's relation examines the ratio of repetition to diversity within a repertoire (McCowen, Hanser et al. 1999). It does not describe how a communication system is organized. Higher levels of complexity within a communication system are assessed using the Information Theory statistics designed by Shannon and Weaver (1949). The zero-order entropic level (entropy here is defined as the degree of organization) is a measure of the diversity within a repertoire and is calculated with:

$$H_0=\log_2 N$$

where N is number of letters, whistles, or in this case, call types (Shannon and Weaver 1949, McCowan et al. 1999).

First-order entropy is another measure of zipf's statistic and measures the simple complexity within a repertoire (McCowen, Hanser et al. 1999). It is calculated using:

$$H_1(A)= -p(A_1)\log_2 p(A_1) - p(A_2)\log_2 p(A_2) \ldots - p(A_N)\log_2 p(A_N)$$

where $p(A_1)$ is the probability or frequency of occurrence of event (i.e call type) A_1 and so on.

Second-order and higher entropies measure how calls are organized and used in conjunction to transmit information (McCowan et al 1999). To examine information sequencing within the second level of complexity, one must assess the probability that one call follows another:

$$H_2(AB) = -p(A_1B_1)\log_2 p(A_1B_1) - p(A_1B_2)\log_2 p(A_1B_2) \ldots - p(A_1B_N)\log_2 p(A_1B_N)$$
$$-p(A_2B_1)\log_2 p(A_2B_1) - p(A_2B_2)\log_2 p(A_2B_2) \ldots - p(A_2B_N)\log_2 p(A_2B_N)$$
$$\ldots - p(A_NB_N)\log_2 p(A_NB_N)$$

where A_1B_1 is the first event A, which is call type one followed by the second event B, which is also call type one (McCowen et al. 1999). In other words: $p(A_1B_1)$ is how often call type 1 follows call type 1. Similarly, $p(A_1B_2)$ is the probability that call type 2 will follow call type 1. All possible combinations of calls are found and used in the equation and totaled.

When examining the third level of complexity, one must then assess the probability that a certain call will follow two ordered calls and so on giving:

$$H_3(ABC) = H_2(AB) + HAB(C)$$

where HAB(C) is the probability that C will occur given that A and B have already occurred (McCowan et al. 1999).

By evaluating the higher orders and plotting the entropic order, one can determine how the information encoded in a communication system is organized (McCowan et al. 1999). A completely random call set gives a slope of zero, while more organized communication systems will yield a successively steeper negative slope (McCowan et al. 1999).

A study by McCowan et al. (2002) investigated and compared organization and development in bottlenose dolphins, squirrel monkeys and humans. By illustrating the utility of information theory in comparing these species' communication systems, McCowan et al. (2002) has shown that information theory can be a vital tool for studying animal communication.

<u>Communication in *Orcinus orca*</u>

Orcinus orca is a highly social species with many levels of social organization (Bigg et al. 1990). They congregate in matrilines, which consist of a matriarch, her offspring, and often their successive offspring (Bigg et al. 1990, Ford 1991). A pod is a group of matrilines that are often found together. Occasionally, two or more pods are found interacting and are called a superpod. At each level of social organization, there is a certain amount of vocal variation. The more closely related the individuals, the more similar their vocal repertoire (Ford 1991, Miller and Bain 2000).

Two *Orcinus orca* communities reside in the waters off the coast of Vancouver Island: the northern residents and the southern residents. During the summer months, the southern residents inhabit the waters surrounding Vancouver Island, including the Salish Sea, Georgia Strait, Juan de Fuca Strait, and ranging as far south as Puget Sound and as far north as Pender Harbor. The southern residents were listed under the Endangered Species Act in 2005 (NMSF 2006). They are composed of three pods: J, K, and L (Ford 1991). The cohesion of these large family units and the successive levels of organization within them require an extraordinary amount of communication to take place between individuals, matrilines, and pods (May-Collado et al. 2007).

Due to the aquatic habitat of *Orcinus orca*, the majority of communication must be conducted vocally. Cloudy waters make visual communication impossible over longer distances, while olfactory cues do not provide the quick response necessary for intricate social interactions (Bradbury and Vehrencamp 1998). The aquatic habitat and the complex sociality of *Orcinus orca* have led to complex vocal communication. Ford (1989) described three main types of calls: discrete, variable and aberrant. Discrete calls are stereotypic, have been classified into a set of 29 different call-type categories, and given an alphanumeric designation (Ford 1987). Many attempts have been made to show clear correlations between discrete call types and behavior, but none have been found (Ford 1989, 1991). This suggests that orcas do not have a one-to-one, call-to-behavior pattern and that their communication system is much more complex (Ford 1989). Variable calls do not seem to have any obvious structure, but are often used along with aberrant calls in highly social and excited states. Aberrant calls are similar to discrete call types that have been sped up and shortened (Ford 1989).

Orcinus orca lives in a complex social system that requires a large amount of information to be passed to ensure fitness. It has been suggested that boat traffic is a significant factor affecting the health of the southern resident population (NMFS 2006). By adding more noise to the orca's communication channel, *Orcinus orca* may need to increase the repetition of their calls to ensure that the message is received. Increased vessel traffic could have lead to a detectable change in the Zipf Statistic. If so, it would suggest that reduced vessel traffic could increase the amount of information that *Orcinus orca* can share, thereby increasing their fitness.

McCowan et al. (2002) used Zipf's statistic to study vocal learning through development by comparing Zipf's slopes of different age groups. By using information theory, one can study the development of the repertoire and determine how important channel noise is during vocal development. Information theory can also provide a statistic that can be used to compare the communication systems of different species. This statistic provides a unique insight into the amount of information that may be encoded within *Orcinus orca* calls and can determine the level of complexity at which the repertoire functions. By comparing the levels of organization between species, one can gain insight and compare both social and ecological strategies. Using information theory we can test hypotheses that involve many species.

METHODS

Recordings of *Orcinus orca* will be obtained from different locations in the Salish Sea, off the coast of Washington state between 2005 and 2007 (Table 1). Recordings from Lime Kiln Point were taken in 2006 on a stationary permanent hydrophone as part of The Whale Museum of Friday Harbor's Sea Sound project. Lime Kiln recordings were recorded to computer using a signal recognition program, and then stored to CD. The Lime Kiln hydrophone is located 15 meters from shore at a depth of 8 meters (S. Veirs personal communication). Archived sound files were recorded in 2005 and 2006 from the *Gato Verde* (a 42 foot sailing catamaran) by students participating in the Beam Reach Marine Science and Sustainability School. In 2007, additional recordings will be taken from the *Gato Verde,* using towed hydrophones lowered into the water when *Orcinus orca* are located. The recordings were taken at an FFT rate of 44100 and 16 bit rate. The *Gato Verde* recordings were taken in and around Haro Straight and the Straight of Juan de Fuca.

Southern resident pods interact, communicate, and share a repertoire (Ford et al. 2000). Therefore, I used recordings of all resident pods and all combinations of resident pods. The necessary sample size of 3,654 was calculated as described in McCowan et al. (1999), using the formula

$$N(r)=n!/r!(n-r)!$$

where r is the entropic order(3) and n is the number of call types(29). Recordings will be viewed using Xbat (Harold Figueroa) in Matlab 7.4.0 (Mathworks). As sound files are viewed, I will record call types that can be heard and/or seen on the spectrogram. Call types will be identified using Ford's (1987) catalogue of calls produced by *Orcinus orca*. I will record call types and the order in which they occur as a PDF. I will then use the find function in Adobe Reader 8.1 (Adobe) to count the number of occurrences of any one call and any sequence of calls from 2-3 ordered call types. Probability will be calculated and then entered into Excel 2000 (Microsoft) using the number of occurrences of a call or ordered sequence being tallied, divided by the total number of calls or sequences.

Measures of the different entropic levels and Zipf's Statistic will be calculated in Excel using the equations outlined in the introduction. The different levels of entropy will then be graphed and the regression statistic calculated using Excel. The slope of the levels of entropy will then be compared to other species that have been previously calculated in the scientific literature.

Table 1. Sound file specifications

Location	Year	Source	Microphone	Recorder
Gato Verde	2007	Beam Reach 2007	LAB core Hydrophone Peak Sensitivity ~ 5000 Hz Down 30 dB at ~200 Hz-10, 500 Hz	Flat from 10 Hz to 40 kHz (+ 0.1, -0.5 dB)
Gato Verde Archived	2006	Beam Reach 2006	International Transducer Corporation (ITC) hydrophone	Marantz PMD660
Gato Verde Archived	2005	Beam Reach 2005	ITC hydrophone	Marantz PMD660
Lime Kiln Point	2006	Whale Museum	Cetacean Research Technology Hydrophone	Desktop computer with built in sound

LITERATURE CITED

Bigg, M. A., P. F. Olesiuk, and G. M. Ellis. (1990). "Social organization and genealogy of resident kill whales (*Orcinus orca*) in the coastal waters of British Columbia and Washington state. Reports of the International Whaling Commission Special Issue No. 12: 383-405.

Bradbury, J. W. and S. L. Vehrencamp (1998). <u>Principles of animal communication</u>. Sunderland, Mass, Sinauer Associates

Ford, J. K. B. (1989). "Acoustic behavior of resident killer whales (*Orcinus orca*) off Vancouver Island, British Columbia." <u>Canadian Journal of Zoology-Revue Canadienne De Zoologie</u> **67**(3): 727-745.

Ford, J. K. B., G. M. Ellis, and K. C. Balcomb. (2000). <u>Killer Whales: The natural history and genealogy of Orcinus Orca in British Columbia and Washington</u>. Vancouver, Toronto
Seattle, WA, UBC Press
University of Washington Press.

May-Collado, L. J., I. Agnarsson, and D. Wartzok. (2007). "Phylogenic review of tonal sound production in whales in relation to sociality". BMC Evolutionary Biology **7**: 136.

McCowan, B., L. R. Doyle, and S. F. Hanser. (2002). "Using Information Theory to assess the diversity, complexity, and development of communicative repertoires." <u>Journal of Comparative Psychology</u> **116**(2): 166-172.

McCowen, B., S. F. Hanser, and L. R. Doyle. (1999). "Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires." <u>Animal Behavior</u> **57**: 409-419.

Northern Marine Fisheries Service (2006). Proposed Conservation Plan for Southern Resident Killer Whales (*Orcinus orca*). Seattle, Washington, National Marine Fisheries Service, Northwest Region.

Zipf, G. K. (1949). <u>Human Behavior and the Principle of Least Effort</u>. Cambridge: Addison-Wesley Press.