Information Theory and Dolphin Vocalizations

Review of papers:

B. McCowan, S. F. Hanser and L. R. Doyle. *"Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires"*. Animal Behaviour, 57, 409-419, 1999

R. Suzuki, J. R. Buck and P. L. Tyack. *"The use of Zipf's law in animal communication"*. Animal Behaviour, 69, F9-F17, 2005

Xanadu Halkias @ LabRosa

Is there a structured language?

- Find a metric that could either imply or reject structure within vocalizations.
- Information theory provides tools for the analysis of communication channels.
- Already used to determine possible patterns in animal communication and behavior.

Zipf's law

- Definition:

  The distribution of the logarithm of the signal rank (first, second…most used etc) vs. the logarithm of actual frequency of occurrence for human languages

- Principle of least effort (Zipf 1949): Human language optimizes between unification and diversity.
- This balance between random and uniform is represented by a slope of –1.00

Zipf's law and Shannon

- Zipf's law examines Shannon's entropy for a source.
- First-order entropy provides an insight on repertoire diversity by taking into account some form of internal organization
- Higher order entropies provide a measure of complexity of the system

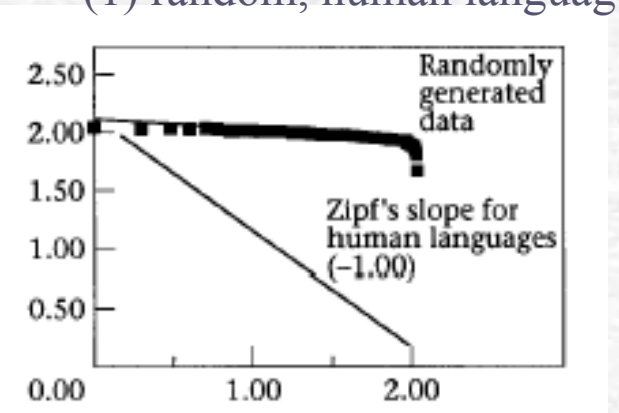
The data

- Use of infant and adult dolphin vocalizations to show potential development in "language".
- Categorize data using a contour similarity technique (K-means on similarity scores of whistles).

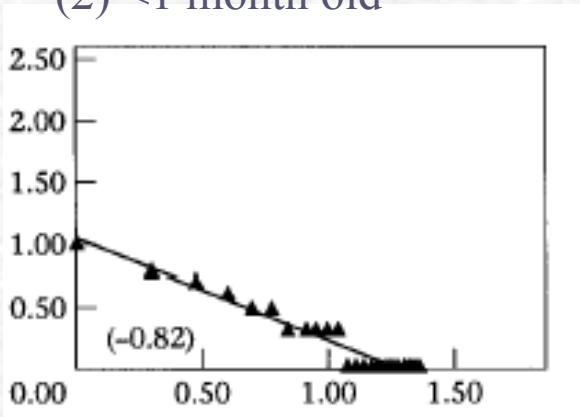
Age	Whistle types	# of whistles
<1 month old	N=23	53
2-8 months	N=73	424
9-12 months	N=60	293
adult	N=102	600

Results

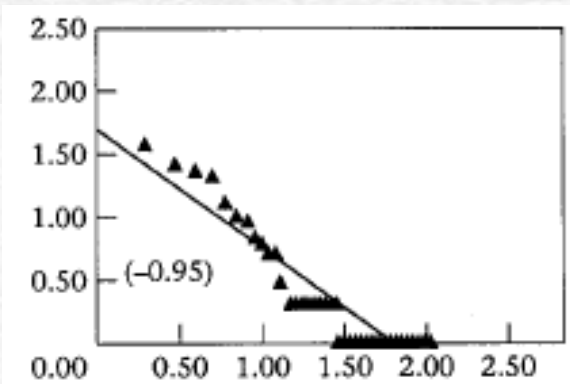
(1) random, human language



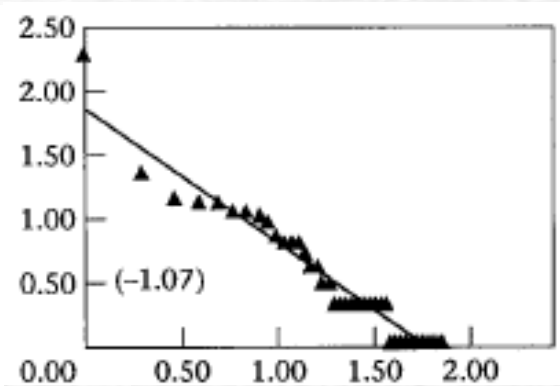
(2) <1 month old



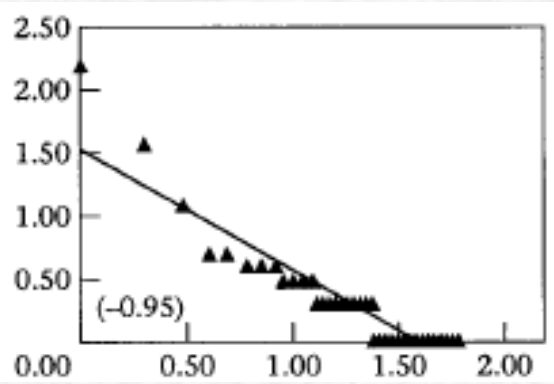
(3) adults



(4) 2-8 months old



(5) 9-12 months old



Comparison

- Lack of data causes artificially low values of higher order entropies or could even hinder their evaluation

Signal system	Slope	R^2	Entropy			
			Zero order	First order	Second order	Third order
Russian letters	−0.500	0.93	5.00	4.35	3.52	3.01
English letters	−0.566	0.96	4.75	4.03	3.32	3.10
Music	−0.680	0.99	3.00	2.73	2.00	NA
Arabic letters	−0.797	0.96	5.00	4.21	3.77	2.49
Dolphin whistles	−1.334	0.86	4.75	1.92	1.15	0.56
Russian phonemes	−1.519	0.89	5.38	4.77	3.62	0.70

NA: not available.

Example of first-order Markov model

- First-order Markov model obtained from analysis

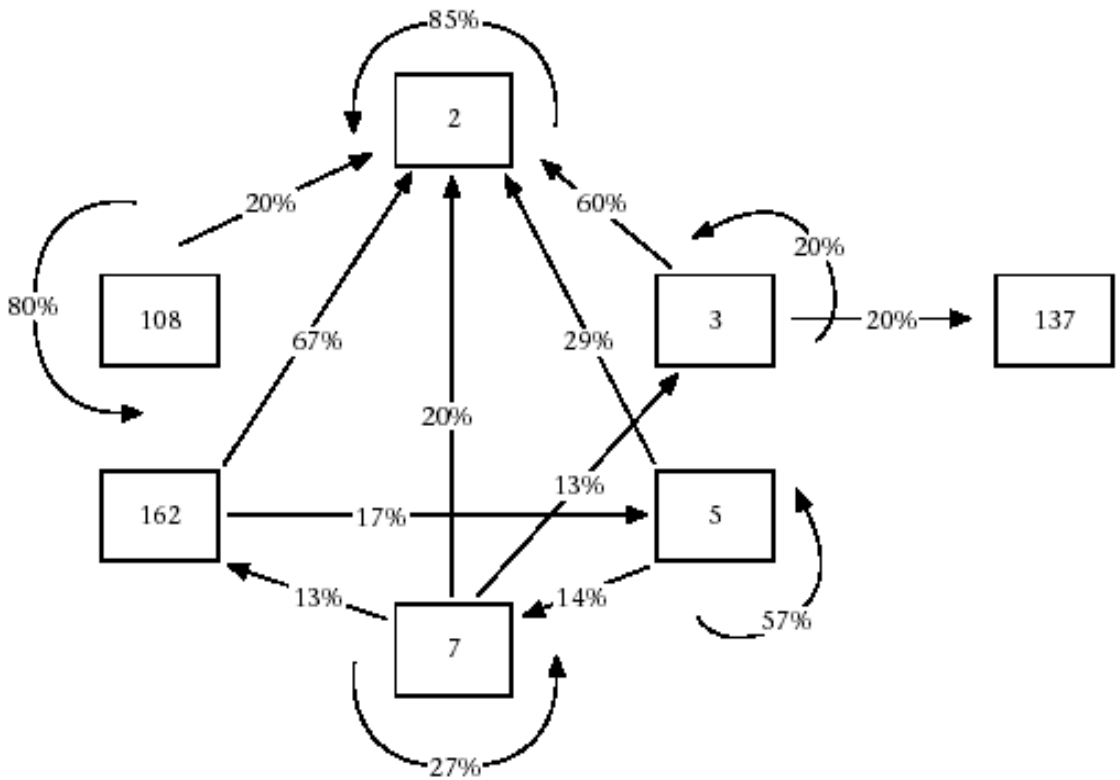


Figure 2. One set of two-whistle sequences shown as a probability tree based on a Markovian first-order (i.e. Shannon second-order entropy) analysis. Numbers in boxes represent whistle types. Percentages and direction of arrows shown represent the probability of one whistle type immediately following a second whistle type. A curved arrow indicates the probability that a whistle of one type immediately follows itself. Probabilities below 0.1 are not shown. The number of whistles for each whistle type (WT) included in the diagram were: WT2=188, WT7=15, WT162=12, WT5=7, WT108=5, WT3=5, WT137=1.

Conclusions

- Good use of tools, but insufficient data causes artificial results that we can't trust.
- More data mining could give insights
- Mutual information could shed more light in structure and interaction of dolphins
- Big problem of getting good data

Reply on first paper by Suzuki et all
The use of Zipf's law in animal communication analysis

- Zipf's law is an unreliable measure of "language" complexity due to high rate of false positives even when adequate data is used.
- Objections on the use of terminology such as nth order entropy etc.
- Reverse Zipf does not hold and it's an empirical law
- Data compression as a counter-example

Main Objection points

- Zipf's law observed in many noncommunicative processes, which means that it is not a good "language" indicator
- Example of die rolling and treating each roll as a letter. Stochastic process with no semantic concept still obeys Zipf's law
- Similar slopes do not imply similar processes
- Lack of data lead to different model that doesn't capture the data