A Mathematical Theory of Communication

Claude Elwood Shannon

Bell System Technical Journal, vol. 27, pp. 379-423, 623-656, July, October, 1948

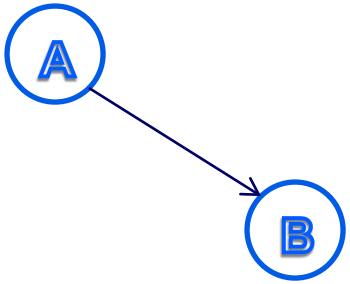
Presented by Andrew Jurik

Theory Lunch

November 6, 2008

“The fundamental problem of communication is…

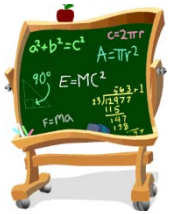
- … that of reproducing at one point either exactly or approximately a message selected at another point”
- Paper seeks to demonstrate a formal way to model communication.
- Insight: a message is selected from a set of possible messages.



The organization of this serial paper is as follows.

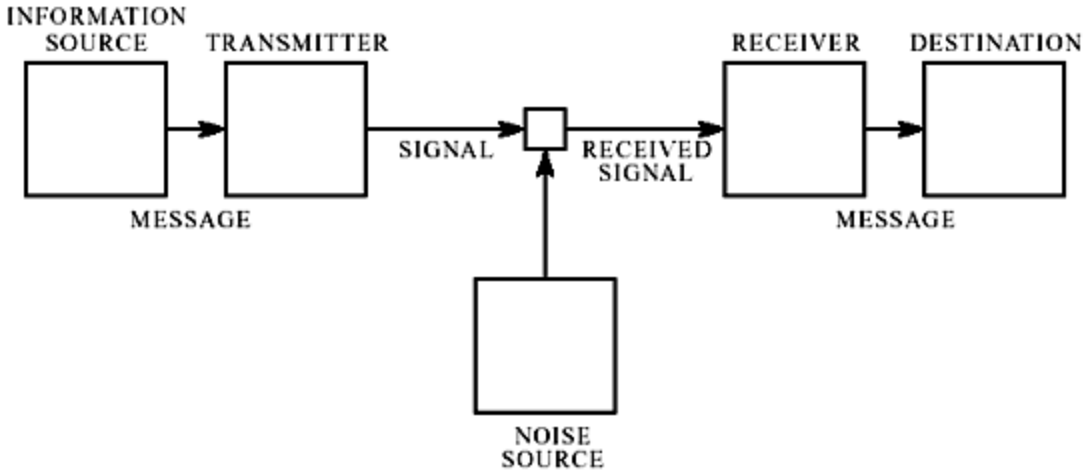
- **PART I – DISCRETE NOISELESS SYSTEMS**
- **PART II – THE DISCRETE CHANNEL WITH NOISE**
- (then appendices)

- PART III – MATHEMATICAL PRELIMINARIES

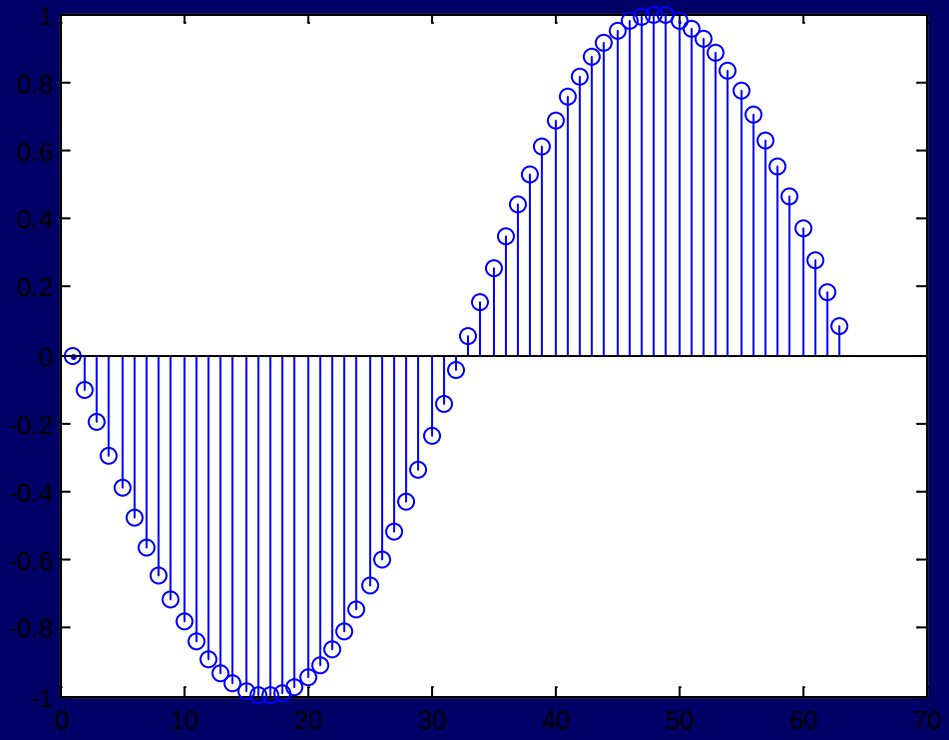


- PART IV – THE CONTINUOUS CHANNEL $\iint$
- PART V – THE RATE FOR A CONTINUOUS SOURCE Δ

Shannon's model of communication has five major components.



DISCRETE NOISELESS SYSTEMS



The capacity of a channel is a tight upper bound on information that can be communicated.

- Channel capacity C (generally in bits per second) is the maximum/gross bit rate or (digital) bandwidth

$$C = \lim_{T \to \infty} \frac{\lg N(T)}{T}$$

- It is not throughput
- It is not goodput (application-level throughput)
- It is not even the maximum theoretical throughput

The discrete source of information can be modeled as a mathematical object.

- Insight: the information source is a stochastic process
- Can be interpreted at many levels of granularity (bits/characters, bytes/words, byte streams/sentences, etc.) and many levels of dependence
- *Properties of an information source:*
- Ergodicity = aperiodic + positive recurrent
- Mixed source – graph made up of a number of pure components (we don't know which ergodic source will be used, but we know one will be chosen)

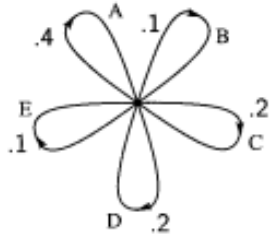


Fig. 3—A graph corresponding to the source in example B.

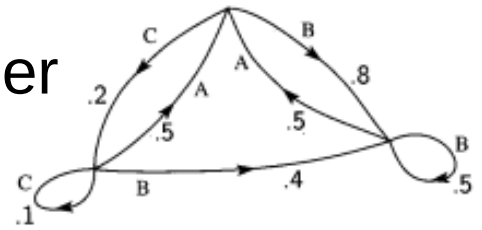


Fig. 4—A graph corresponding to the source in example C.

Tortured English prose?

- THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

… No, just a stochastic process with words independently selected based on their relative frequencies AND a set of transition probabilities from word to word

How can we measure information?

- Entropy is uncertainty. It is also the average information in an ensemble (or event).

$$H(X) := -\sum_{i=1}^{n} p(x_i) \log_b p(x_i)$$

- (Shannon) information content has to do with a particular outcome (−lg p(x)), entropy is based on a set of outcomes
- Several properties:
  - H = 0 iff all p_i are zero, except for one which = 1.
  - For a given number of outcomes, H is a max when all p_i are equal
  - Joint entropy <= sum of individual entropies
  - Averaging increases entropy
  - Joint entropy = entropy of x + entropy of y when x is known
  - Uncertainty of y is never increased by knowledge of x

Let's build some further intuition about entropy.

- Which event has more entropy – a coin flip or the roll of a die?
- Answer: Roll of a die.
- Intuitively, because knowing the result of a die roll discriminates more (produces more information).
- Formally, – [½ * lg(½) + ½ lg (½)] = 1 bit per symbol, whereas – [6 * 1/6 lg(1/6)] ~= 2.58 bits per symbol

What is the entropy of an information source?

- Probability of a message with N symbols…

$$p = p_1^{p_1 N} p_2^{p_2 N} \cdots p_n^{p_n N}$$

- H is approximately the log of the reciprocal probability of a typical long sequence divided by the number of symbols in the sequence

$$H \doteq \frac{\lg 1/p}{N}$$

Fundamental Result #1: Source Coding Theorem

- The number of bits needed to represent an uncertain event is given by its entropy

- It is possible to encode information such that it is possible to transmit at the maximum rate the channel allows

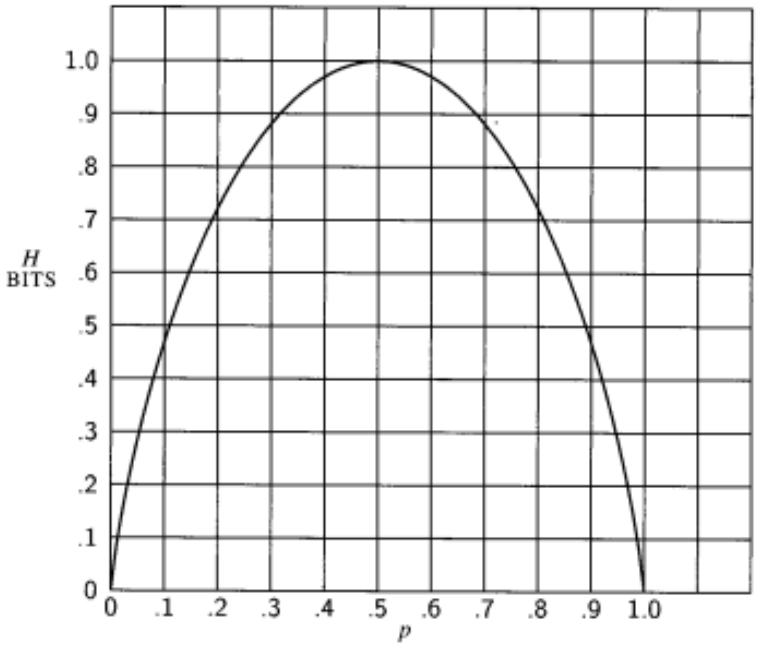


Fig. 7—Entropy in the case of two possibilities with probabilities p and $(1 - p)$.

Example

- Symbols: A (1/2), B (1/4), C (1/8), D (1/8)

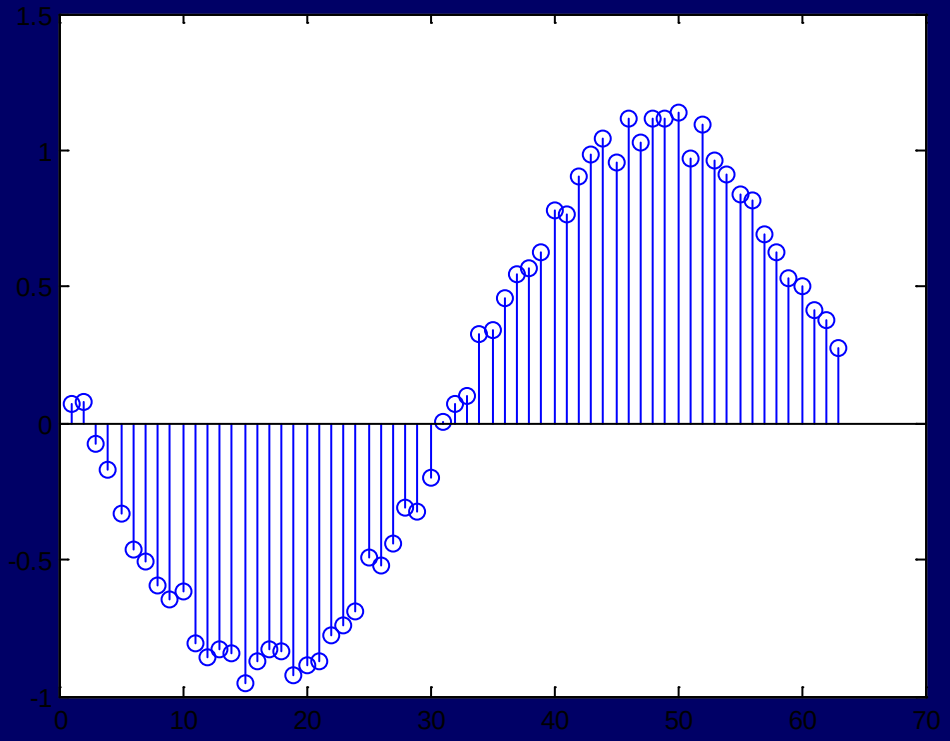
$$H = -(1/2 \lg 1/2 + 1/4 \lg 1/4 + 2/8 \lg 1/8)$$

- = 7/4 bits per symbol

- A: 0
- B: 10
- C: 110
- D: 111

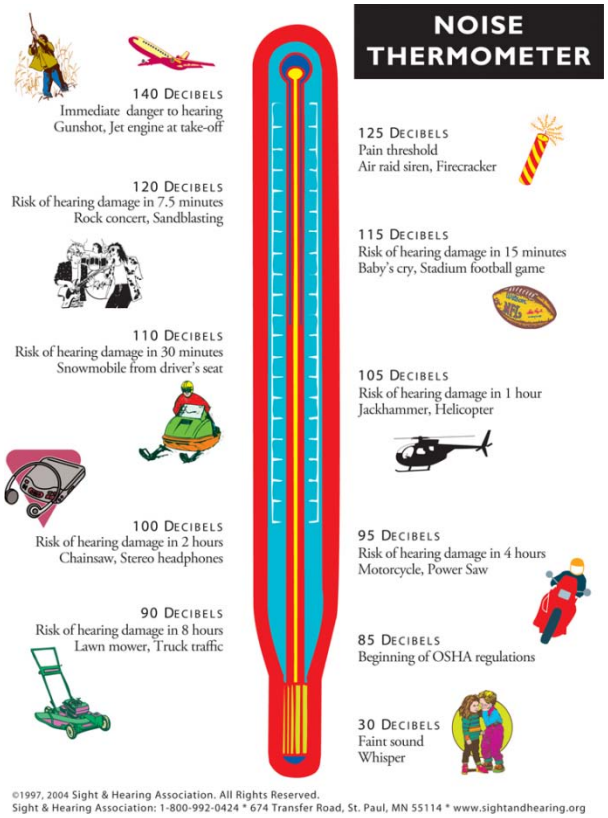
- Average number of bits for N symbols
  = N(1/2*(1 bit) +1/4*(2 bits)+2/8*(3 bits)) = 7/4*N

THE DISCRETE CHANNEL WITH NOISE



Noise can be modeled as a function.

- Distortion = received signal is a definite function of the transmitted signal
- Insight: model received signal as a function of 2 variables – the transmitted signal and noise
- Equivocation, $H_y(x)$ = the information missing in the received signal, equivalent to the conditional entropy of the message (knowing the received signal)
- Capacity $C = Max(H(x) - H_y(x))$



Noisy channel example

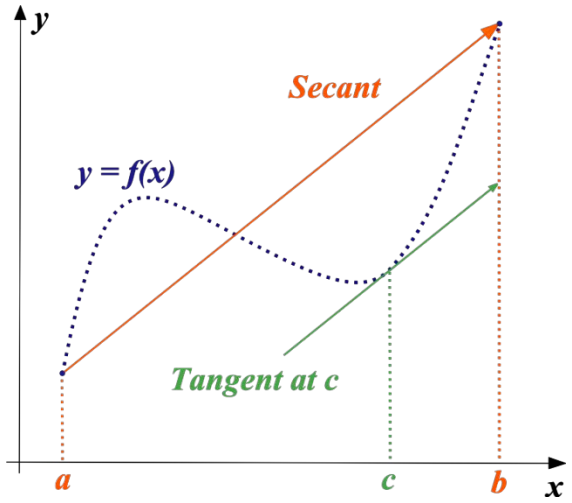
- Say we transmit an A or B with equal probability at 1000 symbols per second. But, 1 in 100 is flipped in the channel. What is the rate of transmission of information?

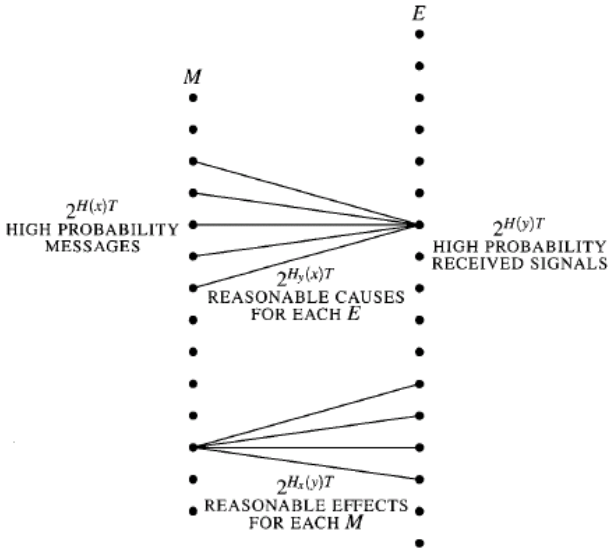
- First, what's the equivocation?
- Because of the .01 error rate, if an A was received, the probability that an A was transmitted is 0.99.
- $H_y(x) = -[0.99 \lg (0.99) + 0.01 \lg (0.01)]$ = 0.081 bits/symbol (81 bits/second)
- System transmits information at 1000 – 81 = 919 bits per second

Fundamental Result #2: Noisy-channel coding theorem

- No matter how contaminated a communication channel may be, it is possible to communicate information nearly error-free up to a given maximum rate through the channel

- Proof idea: based on average frequency of errors of a group of codes less than ε, there must exist at least one code with error less than ε





Puzzles whose solutions can be justified with information theory

- You've been given 12 objects (one of which weighs more or less than the others) and a balance, find the odd one with least number of weighings

- Smallest number of yes/no questions to identify an integer between 0 and 31?

- Strategy: Make the outcome of each trial equiprobable (then the information content will be highest)

Information theory in programming languages: Information Flow

- Topic in PL research
- Label variables with annotations (e.g., group into “high” and “low” variables)
- Enforce some confidentiality policy to distinguish unacceptable flows of information
- Example of information theory in “Quantitative Information Flow as Network Flow Capacity” by McCamant & Ernst

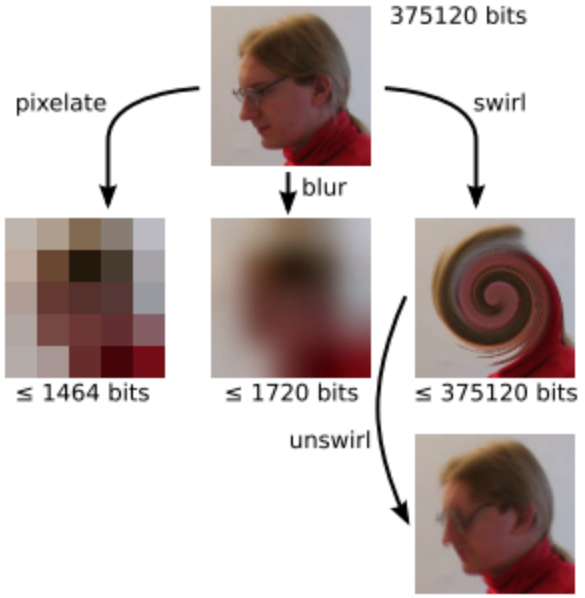
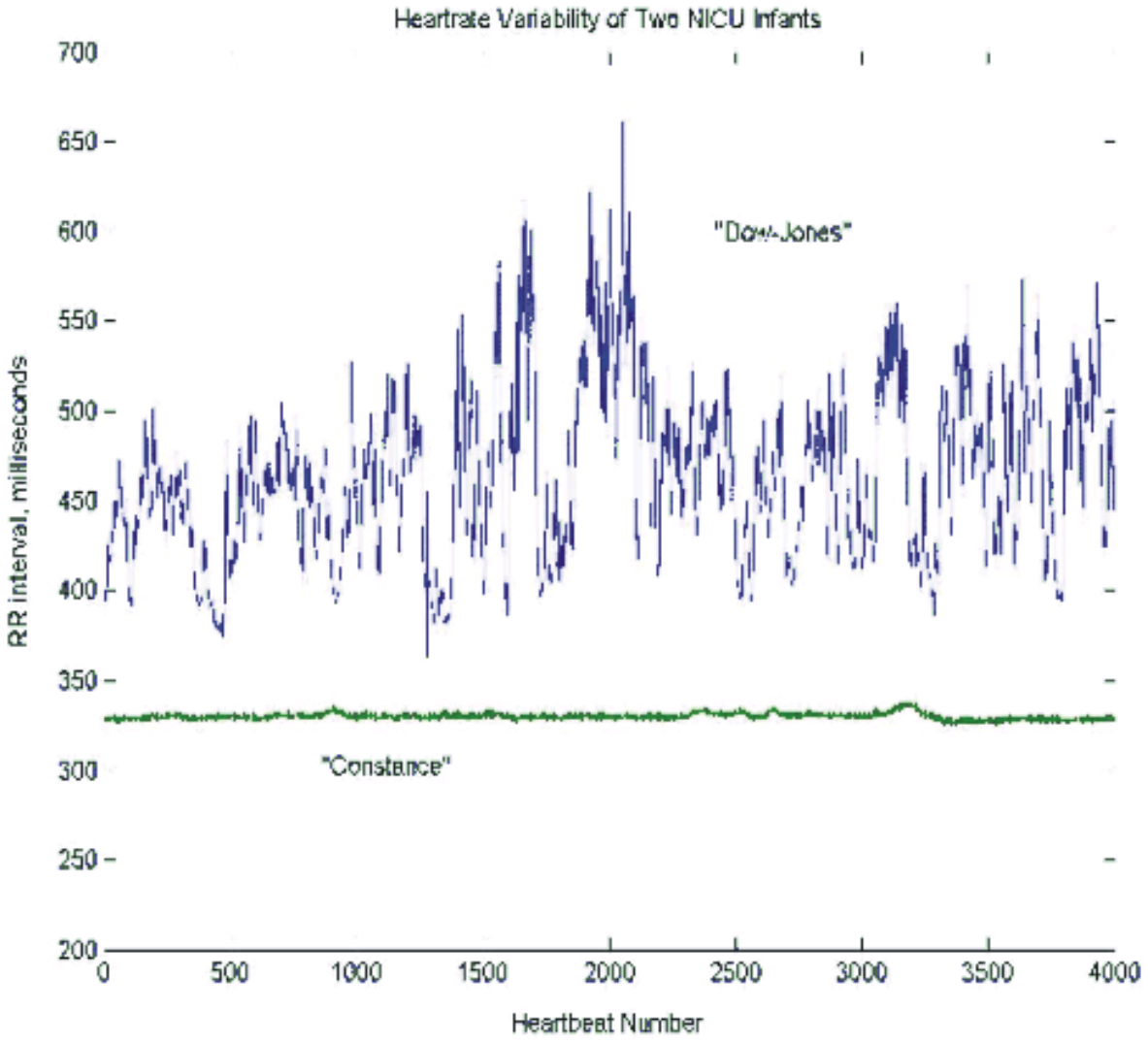


Figure 5. Image transformations vary in how much information they preserve. Our tool verifies that pixelating (left) or blurring (middle) the original image (top, 375120 bits), reveals only 1464 or 1720 bits respectively. By contrast, the bound our tool finds for the information revealed by a twisting transformation (right) is 375120 bits, no less than the input size. Applying the same transformation with the opposite direction to the twisted image gives back an image fairly close to the original (lower right).

Entropy in heart rate variability: good or bad?



Conclusion

- Decoupling the notion of information (the “message”) from the physical propagation of electrons through circuits, light through fiber, etc., (the “signal”) was an important breakthrough

- Communication is fundamental to networks, so having a mathematical model for interpreting messages between nodes (whatever nodes may be) is useful

- (Digital) **Signal** Processing is particularly concerned with (and applies) principles of information theory

References

- David J.C. MacKay's "Information Theory, Inference, and Learning Algorithms"

- Available: http://www.inference.phy.cam.ac.uk/mackay/itila/