Information Theory

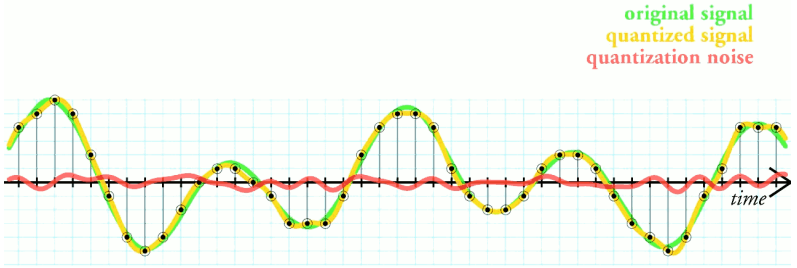
Part 2

Charles Widger 6/17/2025

Analogue to Digital Conversion

Introduction

Example of quantization by Linear Pulse-code Modulation (LPCM). With any PCM, the amplitude is sampled at uniform intervals. With Linear PCM, the quantification levels are linearly uniform. For some telephone speech signals, these levels change, depending on signal amplitude. They are compressed for the signal input, sent, and expanded upon signal output (by companders at each end). This allows more efficient encoding.



- A device or algorithm that does quantizing is called a quantizer. It maps a segment of time, called the *step size*, to a *digital value*.
  - In essence, it is a lossy process. A sample is taken within the step, quantized by rounding to nearest allowed digital value, and output by encoding. The difference between the actual value of the input signal and the quantized value is called *quantification noise*.
- The subject of quantization is divided into two stages, *classification* and *reconstruction*.
- A series of videos by Akash Murthy on engineering considerations for audio applications is available on YouTube. It speaks to choosing the step size, choosing the digital values, demonstrates various kind of noise, and what might be done to make noise less notable, given the hearing range of humans. It illustrates concepts using the music editing software Audacity. This software is cross platform (Windows, MacOS, Linux), public, and can be downloaded at no fee.

Nyquist-Shannon Theorem

Establishes rates for loss-less sampling

- If a function *x* has no frequencies above *B* Hertz, then it can be determined from its ordinates at a sequence of points that are less than *1/(2B)* seconds apart.
  - Claimed to be as quoted in Shannon's 1948 paper.
  - Covers both classification and reconstruction. The accuracy of the quantized results does not matter.
- Proof is based on the Discrete-time Fourier Transform (DTFT).
  - Let x be an integrable function with Fourier Transform $X(f) \triangleq \int_{-\infty}^{\infty} x(t)e^{-i2\pi ft}dt$ (aka the *spectrum* of x) and there are no frequencies above *B* hertz. Define a sampling rate $f_s > 2B$. Let $T \triangleq 1/f_s$; i.e. the *sampling interval*. The proof derives $x(\frac{n}{2B}) = \frac{1}{2\pi}\int_{-2\pi B}^{2\pi B} X(\omega)e^{i\omega\frac{n}{2B}}d\omega$ for all $n \in \mathbf{Z}$ for classification. Further, for construction, let x_n be the n^{th} sample. Then, $x(t) = \sum_{n=-\infty}^{\infty} x_n \frac{\sin(\pi(2Bt-n))}{\pi(2Bt-n))}$.

Open Systems Interconnection (OSI) Model

Computer systems communication channels

- Prior to the development of this model, several networking systems were in use:
  - ARPAnet for the US Department of Defense
  - NPL Network, in the UK, devised by the National Physical Laboratory
  - CYCLADES in France
  - IBM Network Architecture
  - DECnet. The DEC PDP-10 model computer (1966-1983), a 36 bit machine, supported ARPAnet.
- Communication between all was recognized as essential. The OSI was developed to assist achieving that goal, as opposed to dictating particular implementations. For example, older Intel chips were big-endian, Motorola little-endian. Switching between them is done in the Application layer.
- A Frenchman, Hubert Zimmermann, made a first draft of the OSI in February, 1977, while working in Washington, DC. A first draft was released in 1980.

Layer			Protocol data unit (PDU)	Function[26]
Host layers	7	Application	Data	High-level protocols such as for resource sharing or remote file access, e.g. HTTP.
	6	Presentation		Translation of data between a networking service and an application; including character encoding, data compression and encryption/decryption
	5	Session		Managing communication sessions, i.e., continuous exchange of information in the form of multiple back-and-forth transmissions between two nodes
	4	Transport	Segment	Reliable transmission of data segments between points on a network, including segmentation, acknowledgement and multiplexing
Media layers	3	Network	Packet, Datagram[27]	Structuring and managing a multi-node network, including addressing, routing and traffic control
	2	Data link	Frame	Transmission of data frames between two nodes connected by a physical layer
	1	Physical	Bit, Symbol	Transmission and reception of raw bit streams over a physical medium

Spread-Spectrum Techniques

Multiple applications

- Applies to cable, electromagnetic radiation, and some acoustic channels, such as sonobuoy and geophone fields.
- Used for bandwidth multiplexing, anti-detection, anti-jamming, avoiding multi-path and fading, resistance to eavesdropping. Implemented for WiFi and Bluetooth.
- Technologies implemented are frequency-hopping, direct-sequencing, and chirping. Direct-sequencing makes the signal bandwidth greater than the information bandwidth. Invented by Gustav Guanella, patented in 1942. A chirp either increases or deceases frequency during its duration and covers the whole communication band. It can be helpful with moving stations.
- Some history:
  - Frequency hopping used by Guglielmo Marconi in 1899 as an attempt to reduce interference. He used spark transmitters, with no modulation. His application was telegraphy, so bandwidth was not an issue. Spark transmitters became illegal in the 1920s.
  - Frequency hopping developed by Telefunken before 1908. Limited use during World War I.
  - Hedwig Eva Maria Kiesler, better known as Hedy Lamarr, obtained a US patent for a radio signal guidance system for torpedos to counter jamming by frequency hopping. The synchronization technique was player piano rolls. She collaborated with a musician.
  - The US and its allies in World War II used SIGSALY (1943-1946), developed by Bell Laboratories with help from Alan Turing. It had a impressive list of first used technologies. Subsequent systems include the Secure Telephone Unit (last model was STU-III), Secure Terminal Equipment (STE), and the current Voice over IP (VoIP) via the Secret Internet Protocol Router Network (SIPRNet).

Cryptography

Covert Channels

- Cryptography applies to instances in which the meaning of the communication needs to be hidden. One way to do that is to hide the meaning in the message. Cryptography is the means to do this. Typically, both a means and some kind of *key* need to be shared. The original message is called *plaintext*, the encrypted result, *ciphertext*. Eavesdroppers are intent of guessing both the means and the key by applying Cryptanalysis.
- One of the uses of spread spectrum is to hide that any messages are being sent. A more general category is *steganography*. This word comes from Greek, steganós (στεγανός) meaning covered or hidden, plus -graphia (γραφή), meaning written. Use of this term dates from 1499.
  - A use is noted by Herodotus around 440 BCE. A nobleman had the head of a servant shaved, the message inscribed on the skin of the head. They waited for the hair to grow back, and then sent the servant to the recipient, who was told by the servant how to read the message.
  - A more typical example is the use of invisible ink. An example is a message by Benedict Arnold, who added information to a letter written by his wife. Several inks and viewing methods are available, even for toys, as this method is easily compromised. It might do for marking a catalytic converter with a security code for tracing stolen items.
  - A more modern variant is embedding information in a JPEG or MPEG file. Another is embedding information in the TCP/IP packets of the Transport Level of the OSI Model.
- The Common Criteria for Information Technology Security Evaluation collects these under the name, “covert channels.” These can be created by misusing systems. For example, a back-up warning signal might be used to send a message by Morse Code.

Cryptography

Some Encryption Schemes

- Julius Ceasar lived (100 BCE - 44 BCE). In his day, the Latin alphabet had 21 letters, or 23, with the addition of two more, Upsilon and Zeta, from Greek. The means was to write the alphabet in normal order, shift it three spaces left, and substitute the letter in plaintext with the letter as shifted.
  A B C D E F G H I K L M N O P Q R S T V X Y Z
  D E F G H I K L M N O P Q R S T V X Y Z A B C
  - So, the message SRXYLRFHVKRSVRFYVLRF renders from *post hoc, ergo proctor hoc*. This would be easier to see if the ciphertext also inserted spaces at the end of words. This is a type of *substitution* cipher, some letter gets substituted by another.
  - An eavesdropper might know that this kind of shift is used, but might not know by how much or in which direction. The sequence MSG is repeated. Whiz Kids at *Wheel of Fortune* would be entertained. The Dutch cryptographer Auguste Kerchkoffs (1835-1903) suggested such an eavesdropper should be assumed to know what kind of cipher is used just by examining the message. The security of the system rests in the key.
- The entropy of the Ceasarian cipher is $\log_2(23) = 4.523562$, so a bit more than 4 and half bits. If we changed the alphabet to English, so 26 letters, and instead of a shift, made the substitution letter pseudorandom, the entropy changes to $\log_2(26!) \approx 88$. This might take a while to crack. Still, there are other things to apply. The most common letter in English is e. A start would apply that by finding the most common letter in the ciphertext and seeing what happens when that is e. The technique is called frequency analysis. It might get the message pretty quick if the message is “the bee’s knees.”
- The French diplomat, translator, and cryptographer Blaise de Vigenère (1523-1596) created a modification of the Ceasarian cipher. Instead of a single shift, a code word or phrase is chosen as a key. The shift changed at every letter of the key. Suppose the key is the word, VICTORY. For the first three shifts, we have
  A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
  V W X Y Z A B C D E F G H I J K L M N O P Q R S T U
  I J K L M N O P Q R S T U V W X Y Z A B C D E F G H
  C D E F G H I J K L M N O P Q R S T U V W X Y Z. A B
  - This type of cipher was used by the Confederate States during the Civil War. It was completely broken in 1863 by Prussian officer Friedrich Kasiski (1805-1881). More available here.

Cryptanalysis

Code Breaking Natural Languages

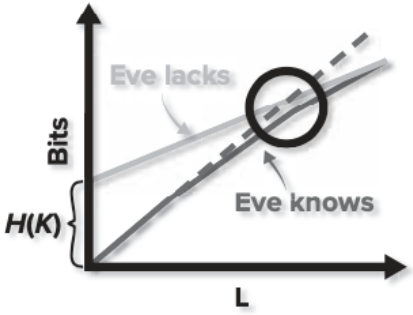
- Not every string of letters is a word in English, although author (and mathematician) Lewis Carroll (1832-1898), AKA Charles Lutwidge Dodgson, would get inventive from time to time.
  Twas bryllyg, and þᵉ slythy toves
  Did gyre and gymble in þᵉ wabe
  - This significantly reduces the amount of possible plaintext of a coded message. Instead of 26! ($\approx 4.0 \times 10^{28}$) strings, there are a mere 250,000 or so, most of them pretty short, to boot.
  - Artificial Intelligence now examines text to predict what the next word(s) will be. This can be used for cryptanalysis, too.
- Ciphertext may not give away the size of words, but if it does, the cryptanalysis becomes very much easier by frequency analysis of letters and words. Al-Kindi (ca. 801 - ca. 873) used this technique.
- Knowing the source or the purpose of the message can help, too. Maybe the plaintext is a weather prognosis or the height of a tide.
- For decrypting the World War II German Enigma messages, it eventually became apparent that many messages started with “Heil Hitler.” This kind of hint is called a “crib.”
  - It also helped that the users got lazy about applying the machine as specified.

Labelled roadside marker

Cryptanalysis

Shannon Viewpoint per Schumacher

- Alice encrypts plaintext P as cipher text C of length L using key K and sends it to Bob. Eve intercepts C and attempts to decode. Initially, Eve has no information. When she gains C, she now knows H(C) and L. Define $H(C) = h_C$ and $H(P) = h_P$. Eve knows Lh_C. She lacks $H(K) + Lh_P$. Although $h_C = 4.7$ bits per letter, h_P is less that 2 bits per letter, due to redundancy in English. Shannon's estimate is 1.5 bits. So the slope of the line that Eve learns from receipt of C is greater than the slope of the line that Eve lacks.
- Shannon defines $D = h_C - h_P$ as the *unicity distance*. If the length L is big enough, there is enough information to decipher the code. For a substitution code, $H(K) = 88$ bits, $D = 3.2$ bits per letter. Converting bits to letters, $D = 88/3.2 = 27.5$ letters. The longest undecipherable cipher text is 27 letters.
- A significant challenge with keys is their management. Both sender and receiver need to have the same one, without the beans being spilled to spies like Eve. One method of dispersion is by special couriers.



Unbreakable Codes

One-time Pads

- We have seen from the Shannon viewpoint that the trick to an undecipherable message is to change the key often enough. One imagines a set of keys printed on pages of a pad and distributed to the expected audience. When the message gets to a point of a change, the next printed code is used. When the last code of the page runs out, the page is torn off, destroyed, and the first code on the next page of the pad is started.
  - SIGSALY was used in World War II for voice telephone communication between the USA and the UK. Stations were installed in London at 10 Downing Street, and the War Cabinet rooms underground, and the White House in Washington DC. The one-time pads were phonograph records. Two turntables were queued up at all stations for smooth transitions during conversations.
- Modern garage door openers do not use a one-time pad. They use a *rolling code*. Both the operating remote and the door opener have simple computers to compute the next code in some sequence. They just need to be synched up to work together.

Thermodynamics

Review

- My beginning physics course in college was taught from the textbook by Halliday and Resnick (the combined edition of the two volumes cost $10.50 in 1961). It covers only the 1st and 2nd laws of thermodynamics. These chapters were not covered in the course I took. Elsewhere are discussions of a third law, and even a zeroth law, the last to be added to the enumerations of such laws.
  - Wikipedia has a lengthy time line of pertaining research starting with a thermometer invented by Galileo Galilei in 1593 and ending in 1977 with the Nobel Prize for Chemistry award to Ilya Prigogine (1917-2003) regarding dissipating systems, those well beyond near thermodynamic equilibrium. Examples include tornadoes and typhoons.
  - Rudolph Claudius (1822-1888) coined the word "entropy" in a 1865 paper. He also established what is now the first and second laws of thermodynamics: a) the energy of the universe is constant, and b) the entropy of the universe tends to a maximum. Nobel prize awardee in physics Leon Cooper (1930-2024) is noted to have said entropy was a " ... word that meant the same thing to everybody: nothing." He defined entropy by $S = k_B \ln(\Omega)$, where k_B is the Boltzmann constant, and Ω is the number of particles. This view provides the link with Communications Theory.
  - Later work by Ludwig Boltzmann (1844-1906) explained the macroscopically observable entropy from a microscopic view of mixed uniform particles. This was later extended by Josiah Willard Gibbs (1839-1903) to allow for non-uniform particles.
  - Mathematician Constantine Carathéodory (1873-1950) published an axiomatic version of the second law in 1909. Einstein liked it, others quibbled. The paper is available, translated into English.
- Sir Ralph Howard Fowler (1889-1944) established the zeroth law of thermodynamics. It states that if two systems are in thermal equilibrium with a third, then they are in thermal equilibrium with each other. In other words, thermometers work just fine. He had been the thesis advisor of several noted scientists, including Paul Dirac, Garrett Birkhoff, and Subrahmanyan Chandrasekhar.
- Walter Nernst (1864-1941) established the basis for the 3rd law of thermodynamics, for which he was awarded the Nobel Prize for Chemistry in 1920. The 3rd law says the entropy of a closed system at thermodynamic equilibrium approaches a constant value when its temperature approaches absolute zero.

Entropy

Thermodynamics *vis-á-vis* Information Theory

- A *microstate* of a system is one of the possible arrangements of its particles in terms of its positions and momenta, sometimes referred to as its *phase space*. For practical systems, this can be a whopping big number of data. Recall Avagrado's number, $6.02214076 \times 10^{23}$ per mol.
- Gibbs expressed entropy as $S = k_2 \sum_m p(m) \log_2 \left(\frac{1}{p(m)} \right)$, where m is a microstate with probability $p(m)$ of occurrence and k_2 is the Boltzmann constant that depends on the choice of logarithm. Recall the communication entropy is $H = \log_2 L$, where L is the length of of the message M. Each bit position in M is a microstate and $\forall m, p(m) = p(1) = 1/L$. So,
  $S = k_2 \sum_m L^{-1} \log_2 L^{-1} = (k_2 \times (L) \times (L^{-1}) \times (-\log_2 L) = k_2 \log_2 L.$
- A conclusion from this is that information does not disappear from the universe, because entropy increases.

Maxwell's Demon

A Thought Experiment

- James Clerk Maxwell (1831-1879) proposed a thought experiment in 1867 of how the 2nd law of thermodynamics might be violated. Suppose one had Jar A filled with oxygen, and Jar B filled with nitrogen. Connect these by a tube with a valve. Heat the jars to the same temperature and open the valve. The gasses will flow between the jars and become mixed in both, as the second law states and experiment confirms. Now suppose the valve is controlled by a clever individual, later called a daemon by Lord Kelvin (1824-1907). Can this control re-sort the gasses so Jar A contains only oxygen and Jar B only nitrogen? For example, oxygen has higher density than nitrogen (see). At given temperature, oxygen molecules travel a bit slower on average than nitrogen. So open the gate from A to B only when the molecule is going fast, and from B to A only when it is going slow. After a while, the first process is reversed.
- In 1929, Leó Szilárd (1898-1964) observed the demon would need energy to measure the speed of the molecules. He estimated this to be greater than the entropy lost.
- In 1960, Rolf Landauer (1927-1999) applied information theory, instead. The demon could measure the speed, but then would either have to store the result, or erase it. Erasing would raise entropy, and but storing would not.
- In 1982, Charles Bennett (1943-) (my age) argued that the demon would eventually run out of memory for storage and would have to make more room by erasure.
- Other researchers in 2012 performed experiments to measure the energy lost by storage deletion.
- Further information may be found in Wikipedia.

Information Theory

End Part 2

Charles Widger 6/17/2025