Source Efficiency, Sampling Theorem, Source Coding and Data Compression in Brief

Source and Channel Efficiency and Redundancy	**The source efficiency and redundancy are defined by:** Source Efficiency = $\eta = (H(x)/H_{max}(x)).100\%$ Redundancy = R = $(100-\eta)\%$ $H_{max}(x) = Log_2 M$ with all $p_{xi} = \frac{1}{N}$ with both H(x) and H_{max} should have the same units. **Channel Efficiency and Channel Redundancy** Channel Efficiency = $\eta_{ch} = \frac{I}{C}.100\%$ (I = average Mutual Inform. C= channel capacity) Channel Redundancy = $R_{ch} = \frac{C-I}{C}.100\%$
Sampling Theorem	**Sampling Theorem** states that: [Any analog signal can be represented by its sampled version as long as the sampling frequency is greater than $2.F_{max}$ where F_{max} is the maximum signal frequency] or $F_s \geq 2.F_{max}$. (Hz) **Nyquist Rate (F_N)** is the minimum allowed sampling frequency $F_N = 2.F_{max}$. (Hz) **Sampling Interval T_s** = $1/F_s$ (sec) and Nyquist Interval $T_N = 1/F_N$ (sec) **Examples: Find the minimum sampling frequency (i.e. find $F_s = F_N = 2.F_{max}$) of the following signals:** $A)$ $s(t) = 10.\cos(2000\pi t)$ Sol. $F_{max} = 2000\pi/2\pi$ = 1000 Hz , ∴ $F_s = F_N = 2.F_{max}$ = 2x1000 = 2000 Hz or 1 kHz $B)$ $s(t) = 10.\cos(5000\pi t) + 0.8.\cos(2000\pi t)$, Sol. Since we have sum of 2 signals we choose the greatest frequency as F_{max} Fmax = $5000\pi/2\pi$ = 2500 Hz, ∴ $F_s = F_N = 2.F_{max}$ = 2x2500 = 5000 Hz or 5 kHz $C)$ $s(t) = 8.cos^2(2000\pi t)$ Sol. To find F_{max}, one should remove the square and/or multiplications , thus $s(t) = 8.cos^2(2000\pi t)$ Sol (see below) s(t)= $4 + 4\cos(4000\pi t)$ ∴ $F_{max} = 4000\pi/2\pi = 2000$ thus $F_s = F_N = 2.F_{max}$ = 2x2000=4000Hz or 4 kHz. ***D***) $s(t) = 4.cos(2000\pi t).3.cos(3000\pi t)$ *Sol. (see below) ,* *s(t) =0.5x12.cos (3000 πt + 2000 πt)+ 0.5x12.cos (3000 πt- 2000 πt)* *s(t) =6.cos (5000 πt)+ 6.cos (1000 πt). Thus $F_{max} = 5000\pi/2\pi$= 2500 Hz or 2.5 kHz.* *Thus* $F_s = F_N = 2.F_{max}$ = 2x2500=5000Hz or 5 kHz. Useful Relations for above examples. cosB = 0.5 [cos(A-B) +_cos(A+B)] and $\cos^2(A) = 0.5 + 0,5\cos(2A)$
What is Source Coding?	It is the representation of source symbols using new alphabet to match the channel alphabet. Coding to binary alphabet is most common source coding as in ASCII code to match binary channel Binary code alphabet {0,1 }, Ternary code alphabet : {0,1, 2 }, Quaternary code alphabet : {0,1, 2,3 },
Source Coding Definitions, Theorem, and Efficiency	The source with entropy H(x) can be represented efficiently with coded average length L, $L \geq$ H(x) or L_{min}=H(x)), Definition of source coding efficiency : $\eta = (H(x)/L).100\%$ with H(x) and L have the same units. Also, L = $\sum_x P(x_i).l_i$ where l_i is the length of the given code for the source symbol x_i
Fixed Length Source Coding	Fere each symbol is assigned same codeword length L = $\lceil Log_D M \rceil$, where M is the number of source symbols, and $\lceil A \rceil$ is the ceiling of A (i.e. upper integer). So if M=10 for example then L=4 as in BCD representation. Fixed length code is simple to implement and it is a decodable code but it suffers the possible low efficiency specially when there is non-equal probabilities of the source. Ex. For RLE: **Consider the sequence:" AAAAAABBBCCDDDD ". RLE encoding is** 6A3B2C4D

Variable Length Source Coding	Since, L = $\sum_x P(x_i).l_i$ one may suggest assigning small length l_i for source symbol x_i that has large probability $P(x_i)$ to achieve small L. Since the code is variable length code, then it is required to guarantee its decoding correctly (called decodable) this is achieved by selecting the codewords for all source symbols so that: **No codeword is a prefix of any other codeword (Called Prefix- free property)** Example of variable length codes are ; Shannon-Fano Code and Huffman Code (both are decodable code)
Data Compression	Data compression is required to reduce the memory size and/or to reduce the required time of its transmission. Variable length codes that have average length less than fixed length for the same source could be considered as data compression techniques. Data compression performance is measured in terms of **Compression Ratio (CR)** **CR = size before compression / size after compression** **Saving Ratio (SR) is the complement of CR**, or **SR= (size before compression - size after compression) /size after compression** **Example: For file input size of 3Mbytes while the compressed output file is 1Mbytes. CR=3:1 and SR= 2:1** Lossless Compression: No data are lost during decoding or decompression. **(Examples: Huffman, Shannon-Fano, RLE, Fixed Length coding, GIF, PNG, ZIP)** Lossy Compression: Part of data are lost during decoding or decompression. (suitable for audio and video signals) **(Examples: JPEG, MPEG, H2.64, MP3)**