Digital Speech and Audio Coding

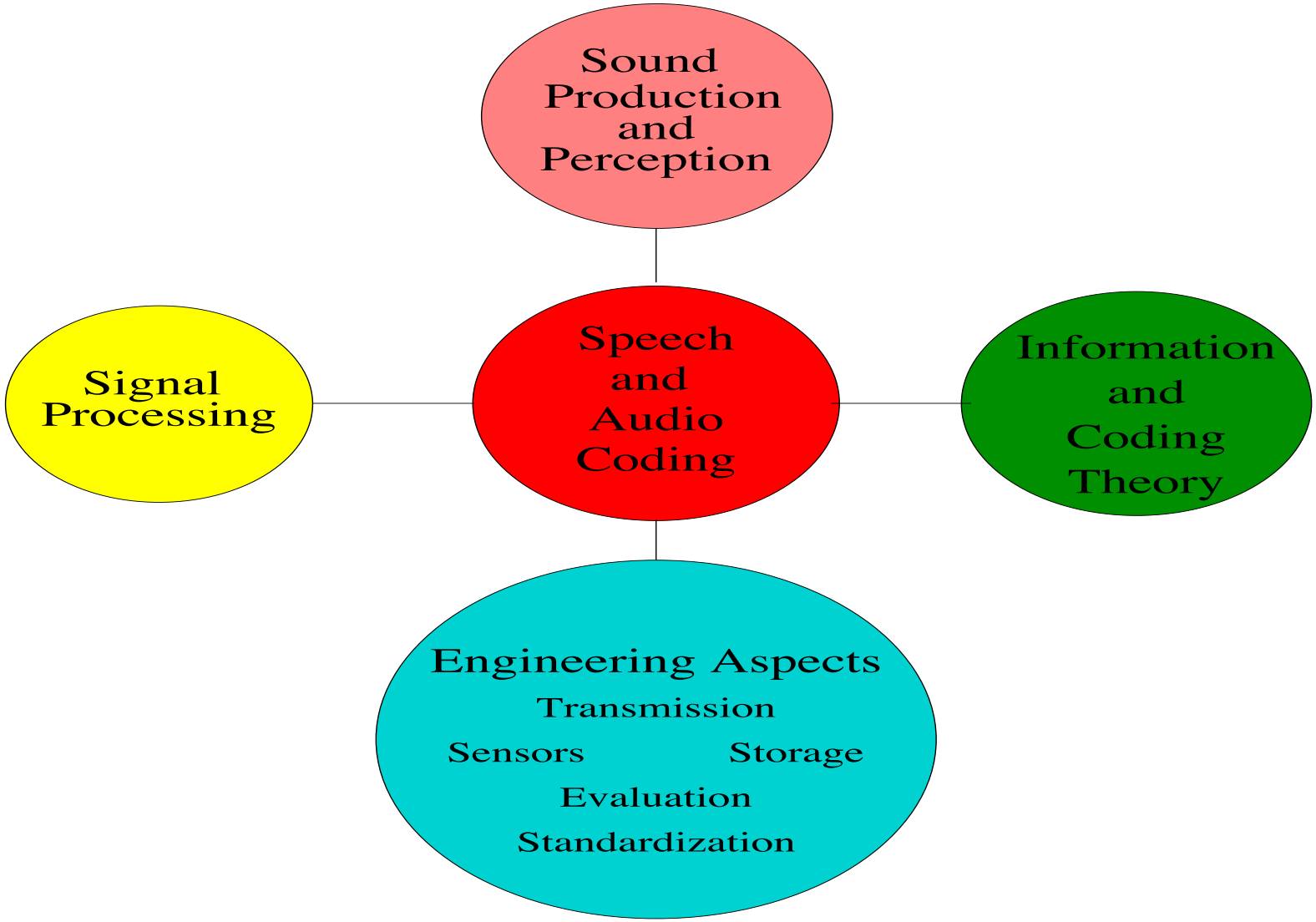
Mathew Magimai Doss and Petr Motlicek

Idiap Research Institute, Martigny

http://www.idiap.ch/

Ecole Polytechnique Fédérale de Lausanne, Switzerland

General Overview



Course Structure

- Introduction (today)
  - Brief introduction to Information Theory and Coding
- Applied Signal Processing
  - Sampling
  - Quantization
  - Signals and Systems
  - Spectral analysis

Course Structure

- Sound Production and Perception
  - Human speech production
  - Human sound perception
  - Speech and music characteristics
- Speech Coding
  - Brief introduction to Information and Coding Theory
  - Waveform coding
  - Transform coding
  - Linear predictive coding

Course Structure

- Audio Coding
  - Perceptual coding
  - MPEG-1 layer 3 (mp3)
  - Advanced audio coding (AAC)
- Other (time permits)
  - Generic coder
  - Bandwidth expansion
- Evaluation of codecs
  - Objective evaluation
  - Subjective evaluation

Lab Exercises

1. Basic signal processing
   - Time domain analysis
   - Frequency domain analysis
   - Psychoacoustics based processing
2. Speech coding
   - Linear prediction
   - Overlap add
   - LP coding
3. Audio coding
   - Quadrature filters
   - Modified discrete Cosine transform
   - mp3

Introduction

Fundamental Questions

- WHAT?
- WHY?
- HOW?

Precursor (to Speech and Audio Coding):

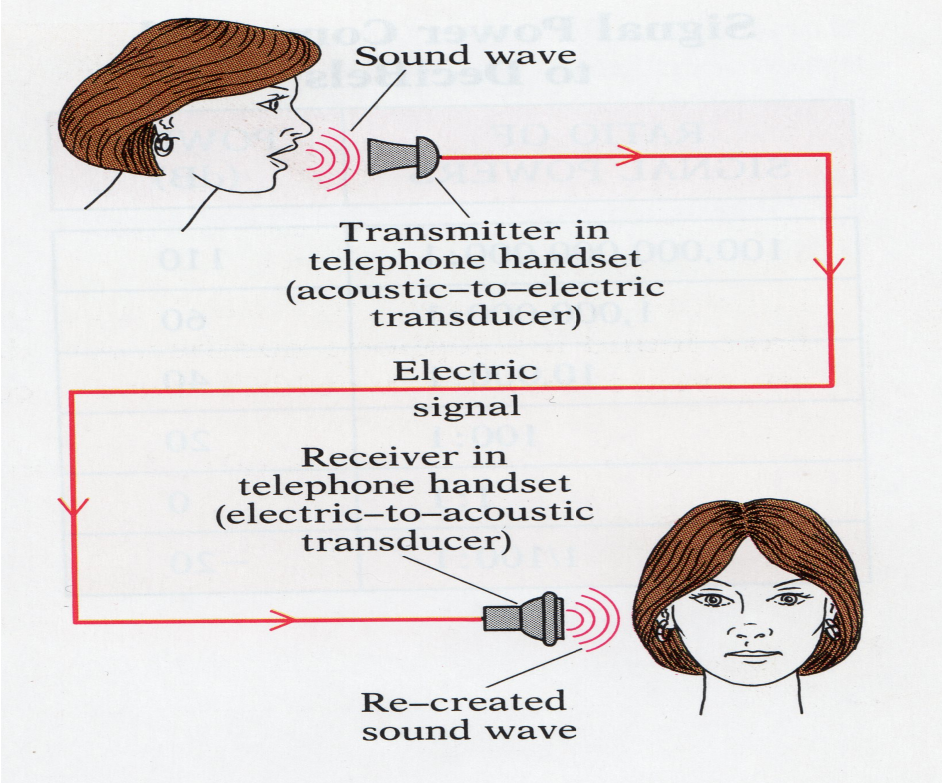
- efficient communication
- storage and re-creation

Modes of Communication (Historical View)

- Speech
  - Good for short distance communication.
  - Long distance communication through messengers is not good.
- Sign
- Drawing and Painting
- Written
  - Good for long distance communication.
  - Postal system
  - Telegraphy
  - How many characters to be sent?

Telephone

"If I could make a current of electricity vary in intensity precisely as the air varies in density during the production of a speech sound, I should be able to transmit speech telegraphically." Alexander Graham Bell



Courtesy- Signals: The Science of Telecommunications by John R. Pierce and A. Michael Noll

A few milestones

Milestone	Year
Commercial Telegraph	1844
Transatlantic Telegraph	1858
Telephone	1876
Phonograph	1877
Flat Disc	1887
Radio	Early 20th Century
Radio Program	1906
Transcontinental Telephone	1915
Transatlantic Radiotelephone	1927
Pulse Code Modulation (PCM)	1938
Speech and Hearing Research	1939

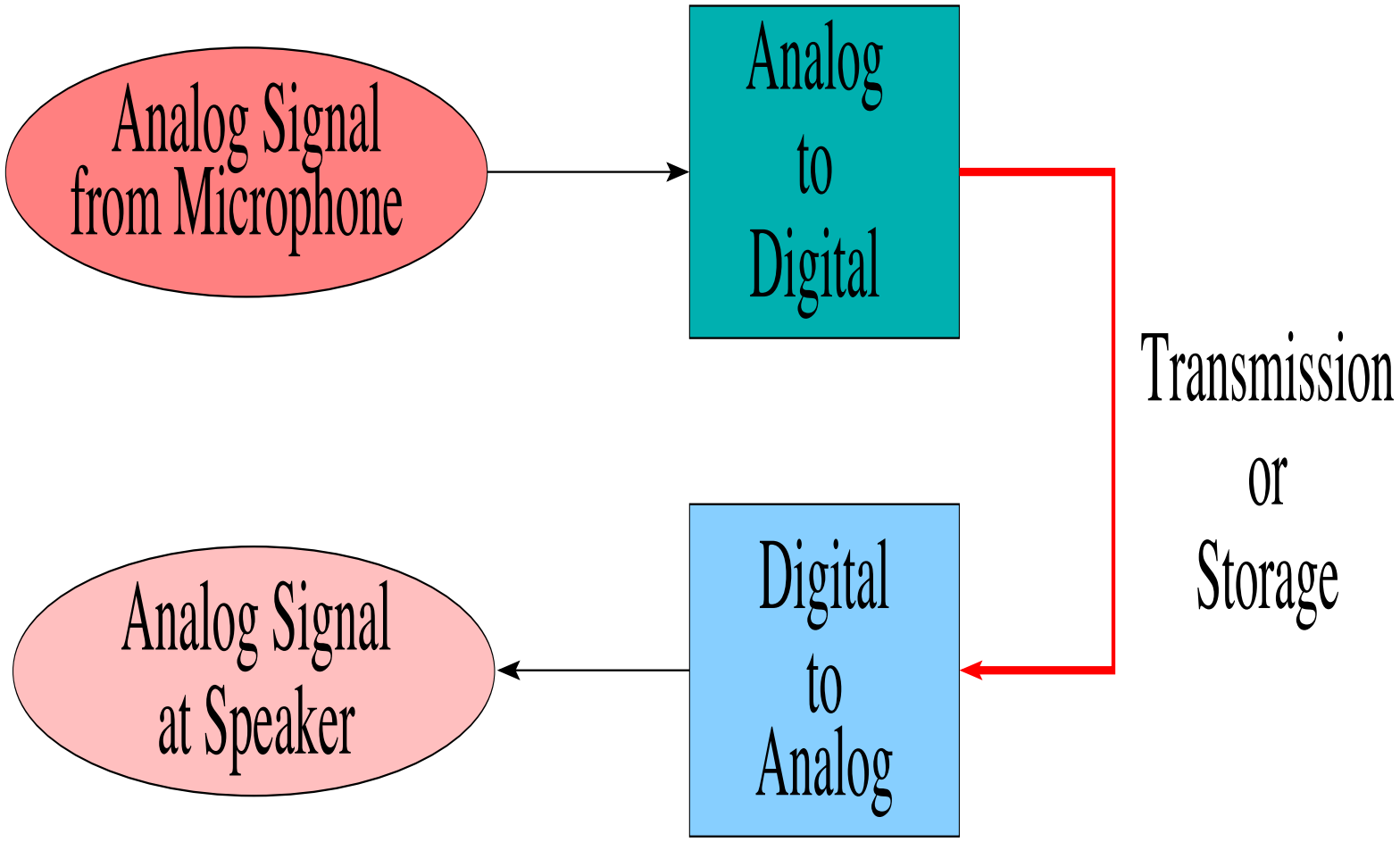
A few milestones

Milestone	Year
Transistor	1947
Information Theory	1948
$33\frac{1}{3}$ LP	1950
Transatlantic Telephonecable	1956
Stereo LP	1958
Commercial Digital Computers	1958
Communication Satellites	1962
Digital Telephone Transmission (using PCM)	1962
Integrated chips	1971
Singlechip Digital Signal Processing	1981

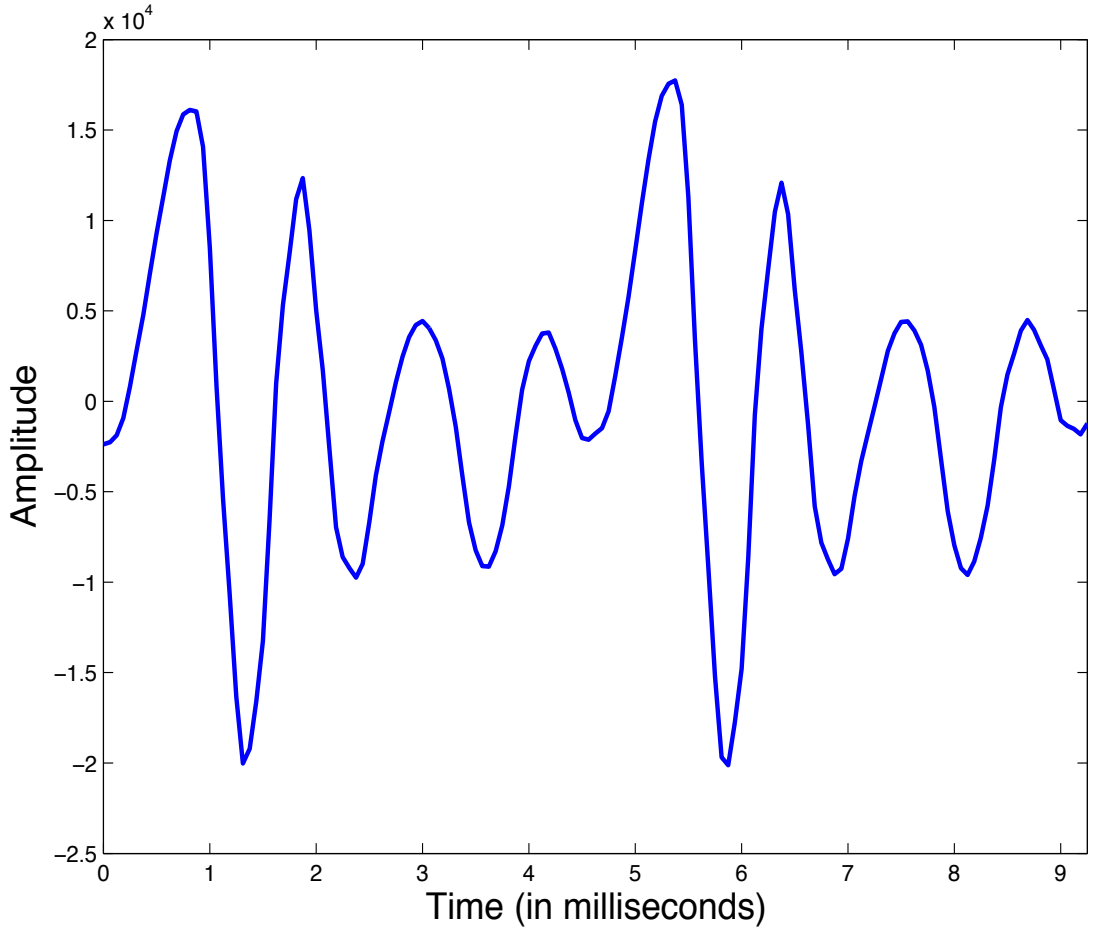
A few milestones

Milestone	Year
Compact Disc	1982
Cellular telephone	1984
Groupe Spécial Mobile (GSM)	1989
Digital Versatile Disk (DVD)	1995-1996
Voice over Internet Protocol (VoIP)	around 1996
3rd Generation (3G)	2001
In US, Western Union discontinued Telegraph	2006

Digital Signal Transmission/Storage

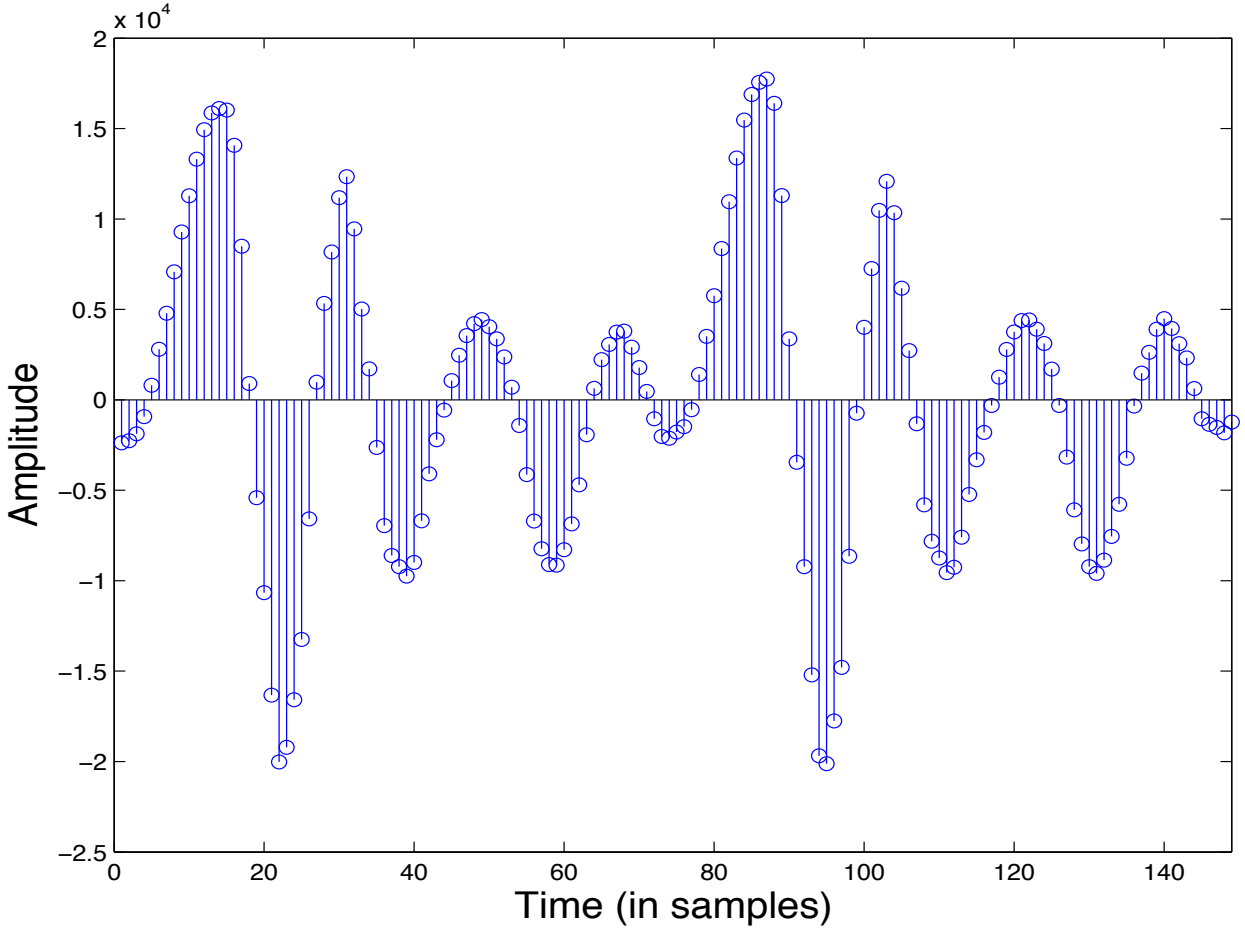


Analog Speech Signal



Both amplitude and time are continuous.

Digital Speech Signal



Both amplitude and time are discrete.

Digital Speech Signal

- Sampling: Discretization of time axis.
- Quantization: Discretization of amplitude (turn continuous value into binary 0/1 representation).

  example: $3 \rightarrow 011$

  Each symbol 0 or 1 is referred to as one bit.
- Bit rate (bits/second):

$$sf \times n$$

  sf: rate of sampling (sampling frequency)

  n: number of bits to quantize

Why Digital?

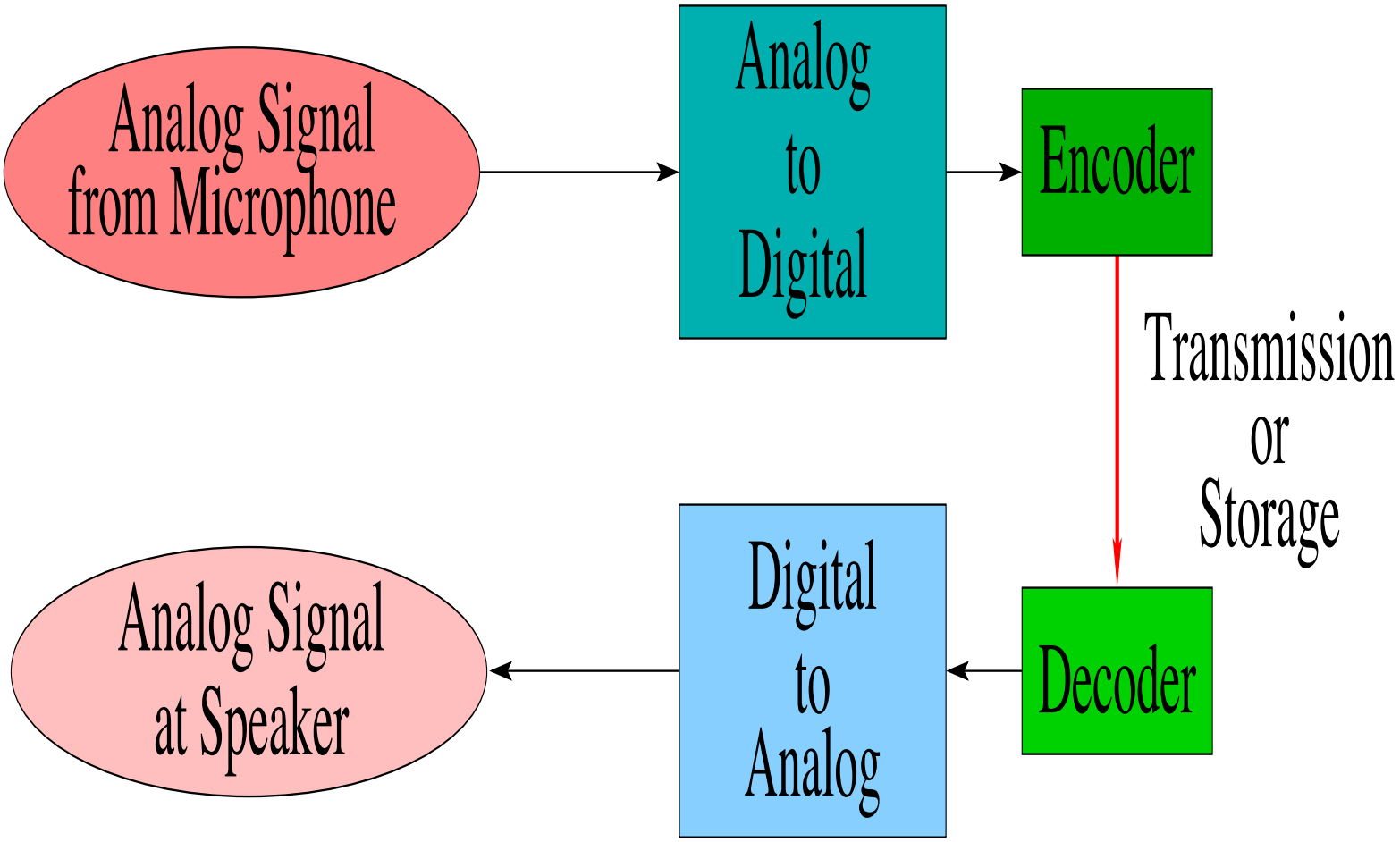
- Analog transmission and storage is fragile and thus can be easily corrupted.
- Error correction.

Coding

Goal: Reduce the bit rate!!!

Cost of transmission or storage $\propto$ Bit rate

Digital Signal Coding and Transmission/Storage



Device capable of encoding (compress) and decoding (decompress) is called as codec.

Attributes of Codecs

- Bit rate: measured as bits per seconds or bits per sample.

Application	Year	Bit Rate
Network Telephony	64 kbps	1972
Network Telephony	32 kbps	1984
Undersea cable	24,40 kbps	1988
Undersea cable	16,24,32,40 kbps	1990
GSM	13.2 kbps	1988
GSM	5.6 kbps	1994
Audio storage	128-384 kbps	1992

kbps: kilo bits per second

Attributes of Codecs

- Signal Quality: Based on human judgement. Measured on five-point scale.

  Bad Poor Fair Good Excellent

  Use of voice activity detector to add comfort noise.

- Processing delay: sum of delays occurring while encoding and decoding.

  delay $>$ 400 milliseconds leads to conversation breakdown in telephone conversation.

  delay has to be less for video or audio conferencing.

Attributes of Codecs

- Complexity: number of machine instructions plus memory requirement (for DSP chips or PC).

  In case of application specific integrated circuit, in addition count number of transistors and gates.

  Power consumption depends on it.

- Robustness towards error especially when streaming audio over packet-switched and wireless networks.

- Quality of Service (QoS)????

Speech and Audio Signal

- Signal: sequence of numbers
- Human can perceive sounds with in the frequency range of 20 Hz - 20000 Hz (not a hard threshold).
  - Audio signal: covers the entire range of 20 Hz - 20000 Hz.
  - Speech signal: most of the relevant information about speech sounds (phonemes/phones) lies with in 8000 Hz.

    Wideband speech: covers upto 7000 Hz

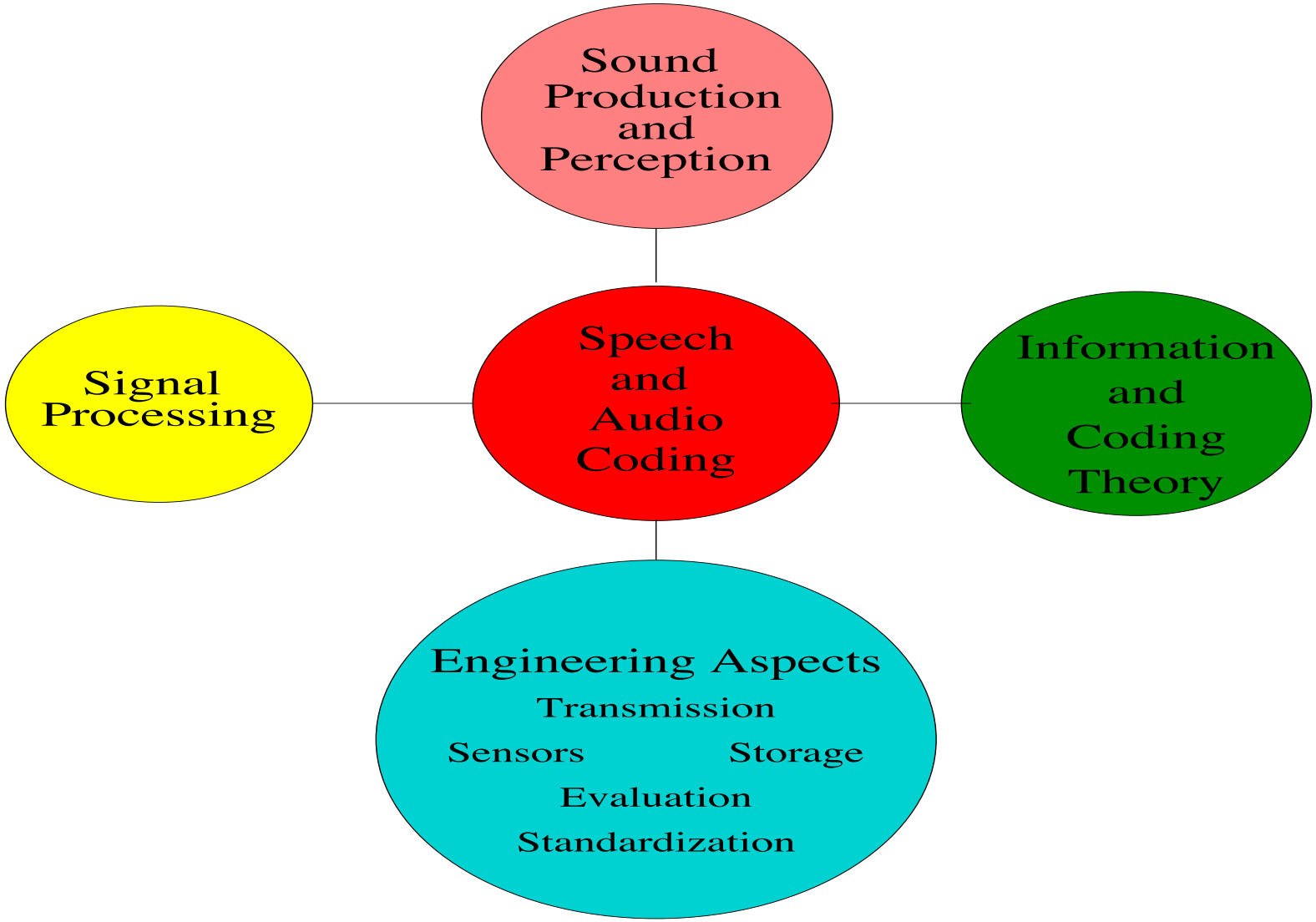
    Narrowband (telephone) speech: 320 Hz - 3200 Hz

Suggested Readings

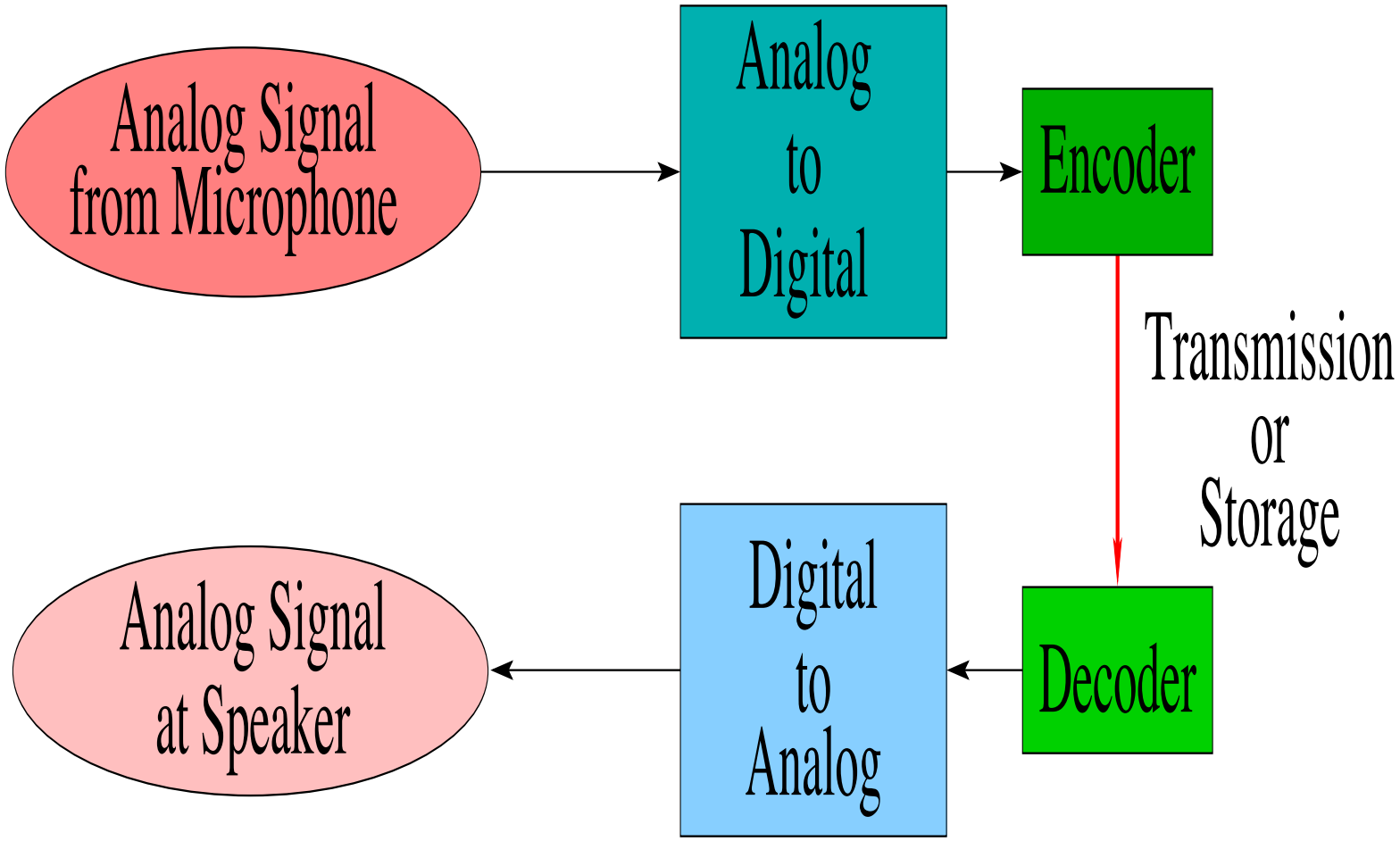
1. Signal Compression, Edited by Nikhil Jayant.
2. Audio Signal Processing and Coding by Andreas Spanias, Ted Painter and Venkatraman Atti.
3. Digital Coding of Waveforms, Nikhil Jayant and P. Noll.
4. Papers put on the course website.

Information Theory and Coding: brief introduction

General Overview



Digital Signal Coding and Transmission/Storage



Device capable of encoding (compress) and decoding (decompress) is called as codec.

Information Theory

- Transmit symbols
- How to quantify the information content?

Information Theory

- $X = \{x_1, x_2, \cdots x_k, \cdots x_K\}$ denote a set of K symbols.
- $P = \{p_1, p_2, \cdots p_k, \cdots p_K\}$ denote probability of respective symbols
- Three things about a function $I(p_k)$ that measures information
  - $I(p_k) \geq 0$
  - $I(p_j, p_k) = I(p_j) + I(p_k)$
  - $I(p_k)$ is a continuous function of p_k

Information Theory

- Measure of surprise

$$I(p_k) = \log(\frac{1}{p_k})$$

- Uncertainty, surprise and information gain are related
- Average information or Entropy

$$H = \sum_{k=1}^{K} p_k \cdot \log(\frac{1}{p_k})$$

Information Theory

- Base of log function
  - base 2 log: bits
  - base e log (natural log): nat
  - base 10 log: Hartley
- Definition of information is objective, i.e., does not cares about the importance of a message

  Will you come out for dinner?

  Will you marry me?

  Answer to both questions is simple yes or no! (1 bit of information)

Information Theory

Analogy with signals

- Single tone (lowest entropy, H_{tone})
- Wideband noise (highest entropy, H_{wb})
- Entropy of speech and music signals lie between H_{tone} and H_{wb}.

Try Interpreting it in frequency domain

Coding

- Binary encoding: encoding each symbol by 0 or 1.
- Example: ASCII code
- Binary state devices are reliable than multistate devices.
- Length of code (B): $2^B \geq K$
- Uniform code length for symbols, e.g., 7 bit ASCII code (in practice 8 bits)
- Bits can be reduced by variable length codes
  - Example: decide code length based on probability of symbol (Huffman Coding)

Entropy Coding

- Uniform code length is *inefficient* if symbols have unequal frequency
- Variable code length (based on symbol probability)
  - Huffman coding
  - Rice coding
  - Golomb coding
  - Arithmetic coding

Entropy Coding

- Encode a message with an ensemble of codes such that the code is
  - Uniquely decodable
  - Prefix free
  - Optimum (i.e. provides minimum redundancy)

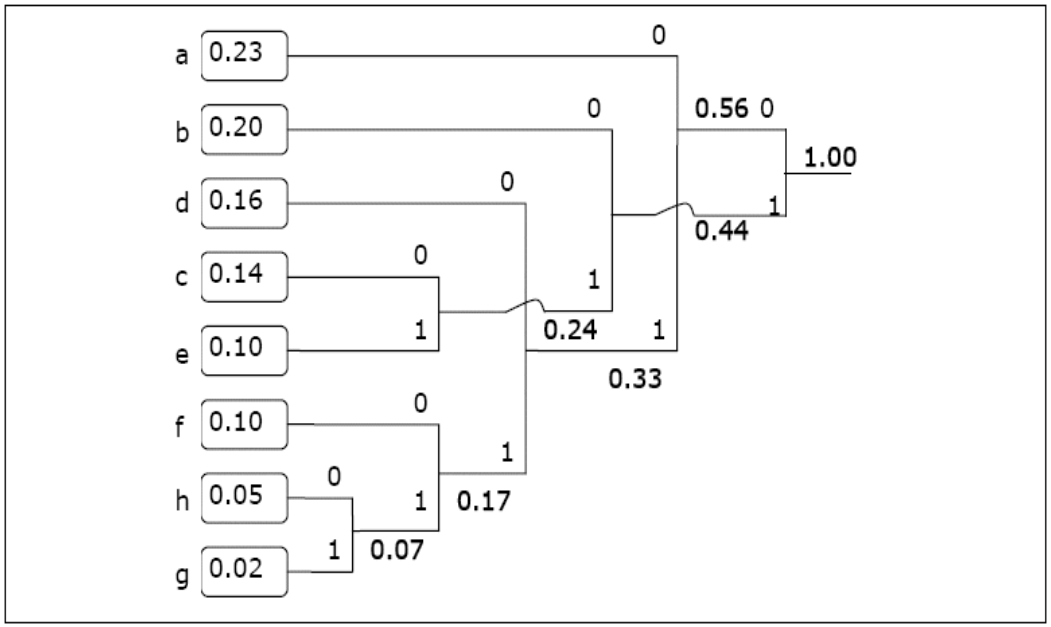
$$R = \sum_{k=1}^{K} p_k \cdot L_k$$

where, R is the compression rate, L_k is length of symbol x_k

Huffman Coding

- Assign fewest bits to most probable symbol and most bits to least probable symbol
- Most effective if the *estimated* probability of the symbols match with the input symbols
- Audio signals:
  - shorter frame length better modelled by Gaussian pdf
  - longer frame length better modelled by Laplacian pdf
- Example application: MPEG-1 Layer 3 (mp3), DVD

Huffman Coding



Huffman Coding Tree

L(K)	C(K)	K	P(K)
2	00	a	0.23
2	10	b	0.20
3	010	d	0.16
3	110	c	0.14
3	111	e	0.10
4	0110	f	0.10
5	01110	h	0.05
5	01111	g	0.02

Rice Coding

- Signal exhibits Laplacian distribution
- Can be considered as Huffman coding with Laplacian PDF
- Code integer into four parts
  - one sign bit
  - m LSBs
  - number corresponding to remaining MSBs as 0
  - a stop bit '1'
- if x denotes the signal

$$m = \log_2(\log_e(2)E(|x|))$$

- Example: integer $I = 65$ and $m = 4$, $binary(I) = 100001$
  [0 0001 0000 1]
- Application: SHORTEN

Golomb Coding

- Exponentially decaying probability distribution of positive integers.
- Given unique parameter m, code integer I into three parts
  - Binary(m mod I)
  - Unary($\frac{I}{m}$)
  - a stop bit '0'
- Prefix codes
- Example: $I = 65$ and $m = 16$

  Binary(16 mod 65) = Binary(1) = 1

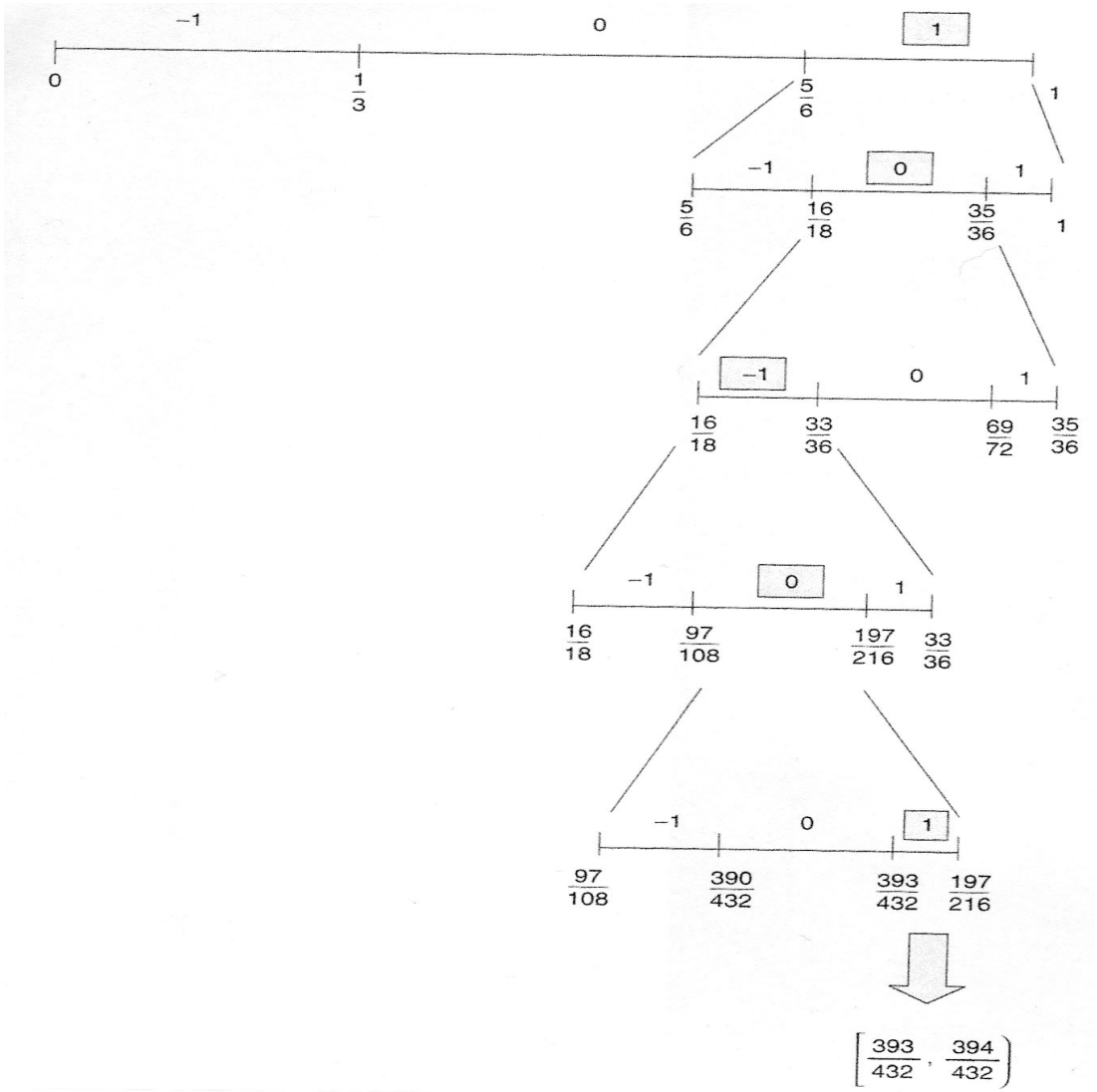
  Unary($\frac{65}{16}$) = Unary(4) = 1110

  [1 1110 0]
- Application: Audio-Pak

Arithmetic Coding

- Codes a sequence of numbers
- Given the symbols $\{x_1, \cdots x_k, \cdots x_K\}$ and their respective probabilities $\{p_1, \cdots p_k, \cdots p_K\}$, encode a sequence $y = [y_1 \cdots y_l \cdots y_L]$ as a rational number in the half open interval [0 1).
- Efficient than Huffman if the probability mass is concentrated on only a few symbols
- Example: $x = \{-1, 0, 1\}$, $p = \{\frac{1}{3}, \frac{1}{2}, \frac{1}{6}\}$, $y = [1\ 0\ -1\ 0\ 1]$

Arithmetic Coding



Next Course

Applied Signal Processing

- Sampling
- Quantization
- Signals and Systems
- Spectral analysis