Universidade de Lisboa

IST

MEEC

DIGITAL TRANSMISSION

1º Practical Work

Information Theory

- Source Coding -

Class: ________**Group:** _______________	**Date:** _____15.12.2024_____
Number: __115593__**Name:** ________Raj Maharjan_________________	
Number: ________**Name:** ____	
Work Evaluation: ____________	**Professor:** ____Prof. Marko Beko___

Year 2024/2025

1. OBJECTIVES

Source coding is a crucial operation in digital transmission systems because it minimizes the transmission rate of the system. This operation is based on the statistical characterization of the information source. The goal of the source encoder is to eliminate redundancy from the symbols generated by the source, thereby reducing the binary rate required to transmit the intended information.

This practical work aims to apply the concepts of information theory to characterize information sources and implement a source encoder/decoder using the **Huffman Coding** method.

Throughout the assignment, variable names are highlighted in **bold**. These variables, and only these, must be grouped into a workspace and sent (**compulsorily**) together with the code file of the work and the properly completed assignment statement.

2. REPORT AND EXECUTION OF WORK IN MATLAB.

2.1) **(Theoretical)** Consider a voice source that continuously generates symbols over time in the form of voltage. Also, assume that the frequency spectrum of the voice signal generated by the source has a maximum frequency of 4 kHz.
Using the Sampling Theorem, determine the minimum sampling frequency (f_s).

8 kHz

2.2) **(Theoretical)** Provide examples of three samples by calculating the instants at which they occur, knowing that the first sample is taken at zero seconds.

T = 1/f = 0.000125. Hence, T(0) = 0 sec, T(1) = 0.000125, T(2) = 0.00025

2.3) **(Theoretical)** Also consider that the voice source generates symbols with discrete amplitude. The total number of symbols (M) generated by the source is 256. How many bits are needed to encode the total number of symbols?

8 bits

2.4) **(Theoretical)** Complete the following table, knowing that the amplitude of the 256 symbols generated by the voice source is uniformly spaced across the normalized interval. $[-1;1[$.

Symbol Index	Symbol Amplitude
1	**1-Δ**
2	**1-(2*Δ)**
.	.
.	.
.	.
254	**1-(254* Δ)**
255	**-1+Δ**
256	**-1**
Δ =	

2.5) Import into MATLAB the sound file corresponding to your student number using the wavread (or audioread and audioinfo) functions.

Example: a student with the number a21401268 **imports the file ST_G6_T2.wav (2+1+4+0+1+2+6+8 = 24 = 2+4 = 6)**

- Name of the variable containing the symbol stream: **`fluxo_simbolos`**
- Name of the variable containing the sampling frequency: **`fs`**
- Name of the variable with the number of bits per symbol: **`n_bits`**

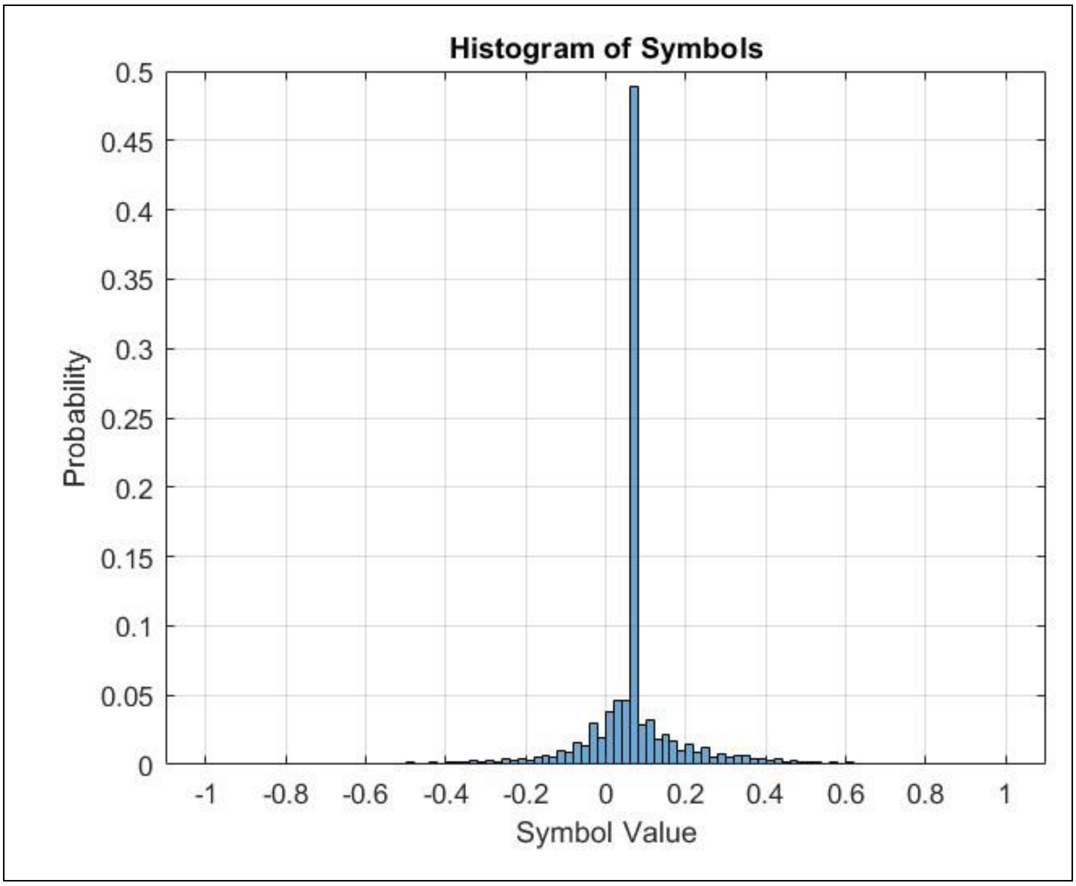
2.6) Determine the number of different (**used**) symbols and their probability of occurrence. How many different (**used**) symbols exist?

204

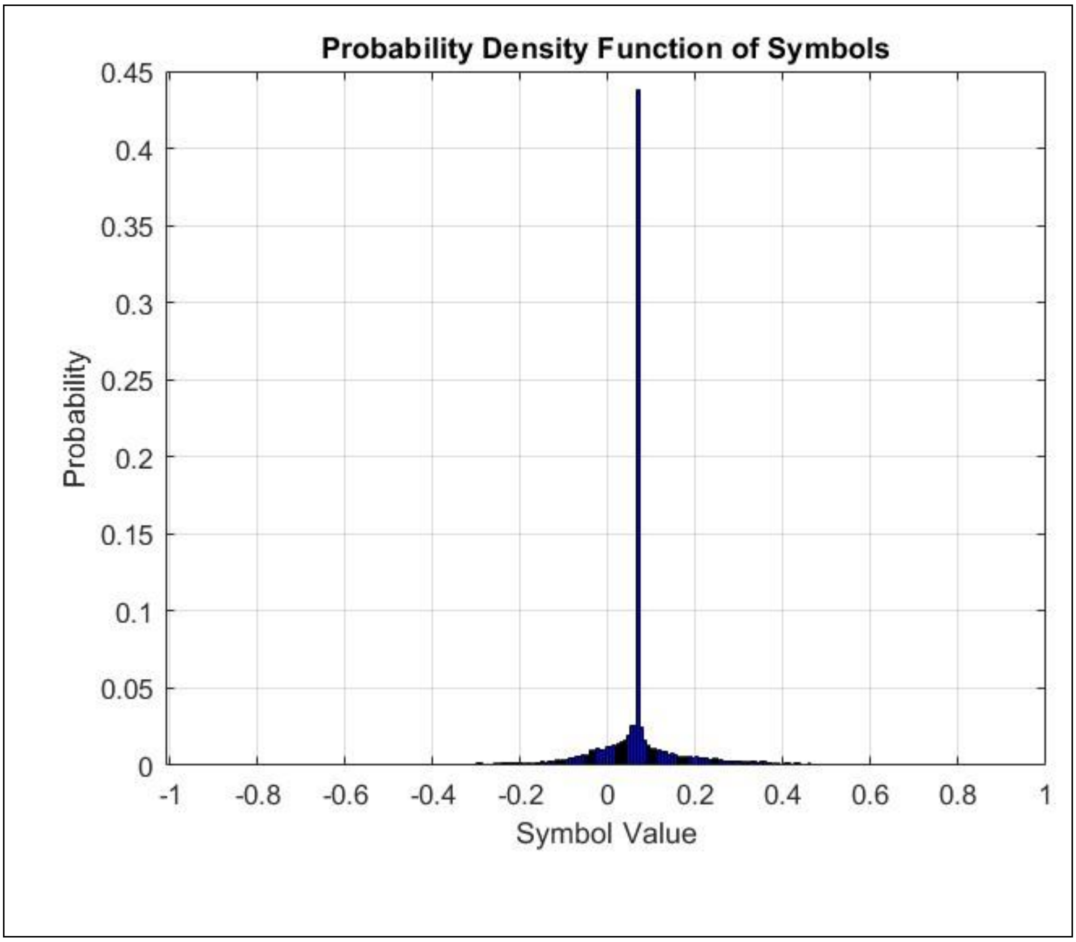
- Name of the variable containing the different (used) symbols: **`simb`**
- Name of the variable containing the symbol probabilities: **`prob_simb`**

Note: The i-th position of the variable **`prob_simb`** should contain the probability of the symbol in the i-th position of the variable **`simb`**.

2.7) Plot the histogram of the determined symbols.



2.8) Plot the probability density function of the symbols.



2.9) Which symbol is the most probable, and what is its probability? What is the least likely symbol **used** and what is its probability?

Symbol = 0.0703, P(0.0703) = 0.4380

- Name of the variable with the minimum probability: **`prob_simb_min`**
- Name of the variable with the symbol having the minimum probability: **`simb_prob_min`**
- Name of the variable with the maximum probability: **`prob_simb_max`**
- Name of the variable with the symbol having the maximum probability: **`simb_prob_max`**

2.10) **(Theoretical)** Which of these symbols represents the most information? Why?

The information content of a symbol is inversely proportional to its probability of occurence. Hence, the symbol with minimum probability will have the maximum information.

2.11)

Based on the previous variables, create a new pair of variables (**used** symbol, probability) containing only the symbols **used** by the source (symbols with non-zero probability) **ordered** in ascending order of probability.

- Name of the variable containing all used and ordered symbols: **`simb_no_zero`**
- Name of the variable containing the probabilities of the used and ordered symbols: **`prob_simb_no_zero`**

Note: The **i-th** position of the variable **`prob_simb_no_zero`** should contain the probability of the symbol in the **i-th** position of the variable **`simb_no_zero`** and should have a higher probability than position $\boldsymbol{i-1}$.

2.12) Calculate the average information per symbol or Entropy (H), the Decision Content (D), and the Redundancy (R) of the symbols. Comment on the results.

Note: Each symbol emitted by a source represents a Decision Content (*D*), *D* = $\log_2(n^o)$ (*bits/symbol*), where "n°" represents the number of different (**used**) symbols. The difference between the Decision Content and the Entropy is the source Redundancy.

3.0795

- Name of the variable containing the Decision Content: **`Conteudo__Decisao`**

- Name of the variable containing the Entropy: **`Entropia`**
- Name of the variable containing the Redundancy: **`Redundancia`**

2.13) (**Theoretical**) Comment on the possibility of designing a more efficient code.

As Redundancy > 0, symbols are not efficiently encoded. For efficient encoding, we can use techniques like Huffmann Coding; to generate optimal code length.

2.14) **(Theoretical)** What would be the average information per symbol or Entropy (H), the Decision Content (D) and the Redundancy (R) if all the symbols from the considered voice source were equiprobable?

In case where all the symbols have equal probability of occurrence, designing a efficient code is not possible, hence H = D and R=0.

2.15) Calculate the Information Rate (R), the Decision Rate (Rd), and the Redundant Transmission Rate. Comment on the results.

Note: If a source emits r_s symbols/s, The information Rate *bits/s* can be defined as, R = $r_s H$.

R = 36.45 kHz; Rd = 61.09 kHz; Db_Redund = 24.63 kHz

- Name of the variable containing the Information Rate: **`R`**
- Name of the variable containing the Decision Rate: **`Rd`**
- Name of the variable containing the Redundant Transmission Rate: **`Db_Redund`**

2.16) (**Theoretical**) In an optimal code, what is the Redundancy of the code? Why?

For an optimal code, Redundancy of the code should be minimal, because that means minimal number of bits will be used to encode the message.

2.17) To implement Huffman coding, you first need to construct the **coding table** based on the symbol probabilities. Build the coding table for the symbols of your voice source.

- Name of the variable containing the table: ***`tabela_final`***

The variable **`tabela_final`** should have the following structure:

It contains as many rows as the number of symbols.

It has 4 columns:

1st Column: Symbol

2nd column: Probability

3rd column: Lenth of the Binary code

4th column: Binary code

The rows are sorted by probability, with the first row corresponding to the symbol with the highest probability and the last row corresponding to the symbol with the lowest probability.

2.18) Calculate the average code length ($\overline{N}$) and its efficiency (η)

N’ = 4.5906; efficiency: 99.34%

- Name of the variable containing the average code length: **`L_med`**
- Name of the variable containing the efficiency: **`Eficiencia`**

2.19) Based on the coding table, convert the symbol stream into a binary stream.

- Name of the variable containing the binary stream: **`fluxo_bin`**

2.20) Calculate the total number of bits in the binary stream resulting from the encoding.

125469 bits

- Name of the variable containing the number of bits in the binary stream: **`n_fluxo_bin`**

2.21) Calculate the average encoded binary rate. Comment on its significance by comparing it with the value calculated in 2.15).

4.5859 b/s

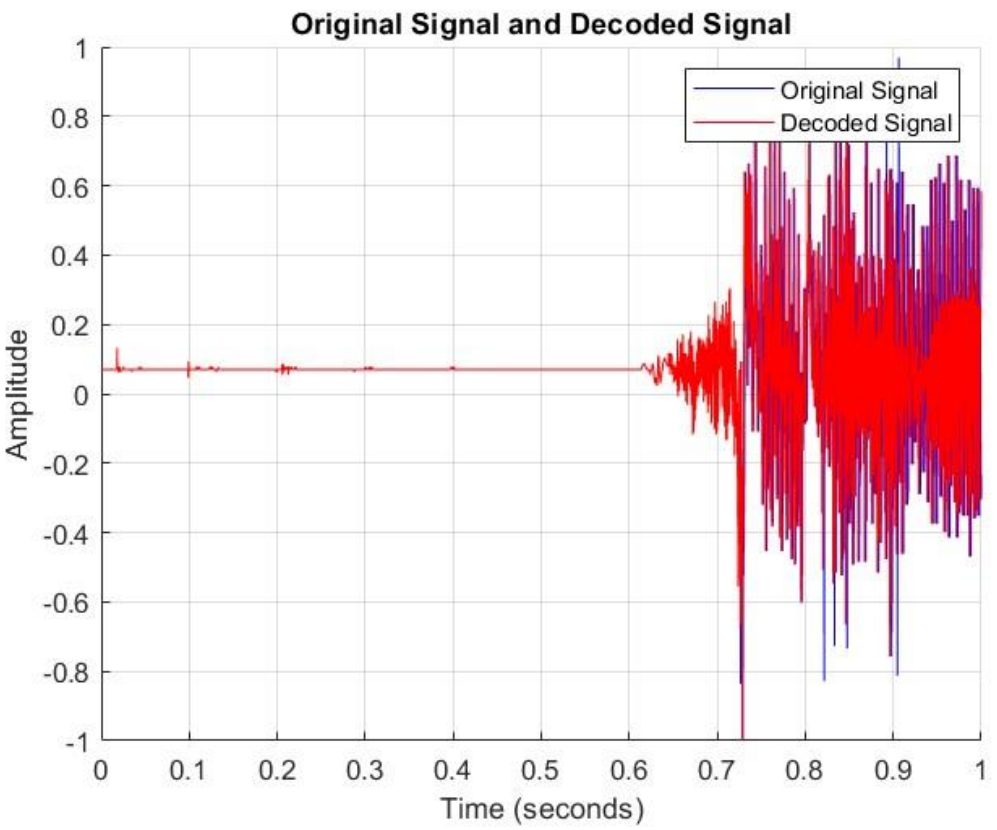
- Name of the variable containing the rate: **`R_cod_med`**

2.22) Decode the binary stream encoded in 2.19).

- Name of the variable containing the decoded symbol stream:

`fluxo_simbolo_descod`

2.23) Plot the amplitude of the resulting signal and the original signal in the same figure up to the first second.



2.24) **(Theoretical)** Under what conditions would it be possible to use the coding table you constructed to encode another voice signal? Justify.

Another voice signal should have similar features like its statistical properties, sampling rate. But in real world this is not probable as we have signals that are dynamic in nature.

2.25) **(Theoretical)** Assume it is possible to encode another voice signal based on the coding table you constructed. Would the resulting average code length be different? Why?

Yes, because this coding table values will not be optimal values for another voice signal as it will have varying probability distribution.

2.26) Save only the variables highlighted in **bold** into a workspace named:

y_T1

2.27) Save the MATLAB code file with the following name:

y_T1

Note: The notation is as follows:

- y corresponds to your student number.