

CHAPTER 1

INTRODUCTION

AT SOME POINT in the scientific history, new revolutionary ideas serve as starting points of new topics. Information theory started in 1948 when Claude Shannon's paper "A mathematical theory of information" was published in the *Bell System Technical Journal* [1]. The objectives in the article in many ways were pioneering, but the first thoughts in this direction started more than 20 years earlier.

In 1924 and 1928, Nyquist [2, 3] showed that a signal with a bandwidth W Hz and a durability of T seconds could not contain more than $2WT$ distinguishable pulses. This was later reformulated into what is today known as the sampling theorem. This is an important concept of all systems converting between discrete and analog representations of signals and is widely used in signal processing and communication theory. In 1924, Hartley was first to propose a measure of information in his paper "Transmission of information" [4]. In this study, he realized that the limiting factor in communication is the noise. Without noise there would not be any problem. And then, 20 years later, Claude Shannon further developed the concepts and concluded that a measure of information must be based on probability theory.

The idea is that without choices there is no information. If a variable can have only one outcome, it does not give any information to an observer. If there are multiple outcomes, their presence is determined by their probabilities, and the information measure is therefore derived from the distribution on which the symbol is generated. That means the developed information measures are purely probabilistic functions. At first glance, it might be surprising that the information measure developed, and on which communication systems still rely, does not have anything to do with the content of a sequence. The measure can instead be interpreted as the amount of information required to determine, or reproduce, the sequence. In that perspective, a completely random sequence also contains a lot of information, even though it does not have a meaning to us. The strength in this view is that the theory is valid for all types of information sources and thus all types of communications.

Information theory deals with two closely related terms: information and communication. As stated above, the information measure will depend on the amount of information needed to reproduce the source sequence. Then, if the aim is to transport this sequence or message from a source to a destination, it is only this amount of information that is needed to be transported. In his paper, Shannon expressed that the "fundamental problem of communication is that of reproducing at one point exactly

2 CHAPTER 1 INTRODUCTION

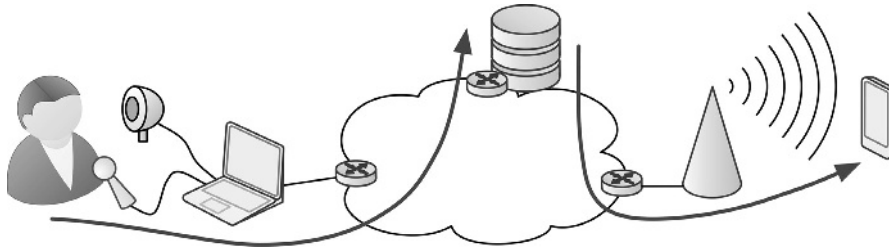


Figure 1.1 A communication system.

or approximately a message selected at another point.” In this rather general statement, a *point* can be referred to as *place* and *time*, meaning that the message selected at on place and one time should be reproduced at another place and another time.

In Figure 1.1, an example of communication is shown. It shows someone recording a video with a computer and then uploading it to a server, e.g., Youtube. Then, at later time someone else downloads the video and displays it on a screen. In this case, the stated *another point* refers to the time and place where the video is displayed. The video that is uploaded is typically compressed using a video compression format, e.g., H.264 or MPEG. Videos are typically compressed quite hard, using a lossy compression. When the video is downloaded and decompressed, it differs quite a lot from the raw video recorded at the source. But for the human eye and perception ability, it should *look* approximately the same. If the message transmitted was a text or a computer program, the same principle can be used, the text is saved on a server and downloaded at a later time. But in case of a program, the reconstruction needs to be an exact copy of the uploaded file. This depicts the difference between the statement *exactly* or *approximately* in the cited case above.

In the world of information theory and communication theory, Figure 1.1, and most other communication situations, is better represented by Figure 1.2. There the source is the recording of the (raw) video that gives the sequence X . To save disk space as well as uploading and downloading time, the source sequence X should be compressed through a source encoder. The aim of the source encoder is to extract the pure information, and in the case of lossy compression a representation with less

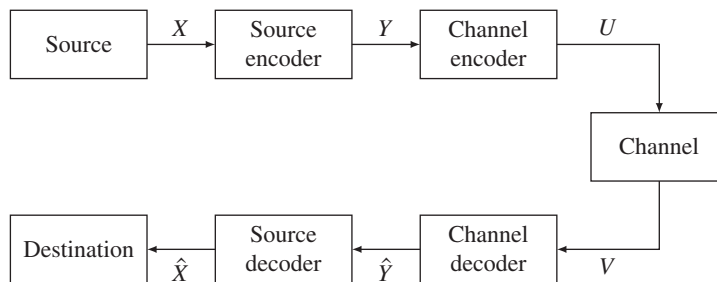


Figure 1.2 Shannon's model for a communication system.

information gives approximately the same experience as the raw video. Since pure information is very sensitive to errors in transmission and storage, it is protected by a channel code. The purpose of this code is to introduce redundancy such that the channel decoder can correct errors. In the case of the video uploading to a data center, there are several occasions of encoding and decoding. First when the file is uploaded, there is encoding and decoding in the transmission protocols. At the data center, there are often error-correcting codes, protecting from hard drive failure. At downloading, there is again the transmission protocol using error-correcting codes. Finally at the destination, the compressed sequence is decompressed and sent to the screen.

The aim of this text is to give an introduction to the information theory and communication theory. It should give an understanding of the information measures and the bounds given by them, e.g., bounds on source coding and channel coding. It also describes source coding and channel coding on the basis of some well-known algorithms. Chapter 2 gives a short review of probability theory, where most of the basic theory needed in the text is presented. The intention is to present a quick and handy lookup of the theory regarding probabilities. Even though there are attempts to explain intuitively the theory, most of the mathematical proofs are omitted. For a more thorough treatment of the subject, the reader should refer to the literature, e.g., [5, 6]. The chapter is complemented by an appendix where some of the most common distributions are shown, both discrete and continuous.

In Chapter 3, the information measures used in the theory are defined. It also presents their relations, and how they can be interpreted. In Chapter 4, the measures are used to set up bounds on source coding. For the case of vectors with identical and independently distributed (i.i.d.) variables, a simple, yet optimal, code construction due to Huffman is given. Even though Huffman's code construction gives optimal codes for the i.i.d. case, in reality sources for compression are seldom independent variables and the distribution is seldom known. In Chapter 5, adaptive compression methods are described. These include the widely used the Lempel–Ziv algorithm, that is used in, for example, zip, png, and other well-known standards.

In Chapter 6, the concept of asymptotic equipartition property is presented, which is a consequence of the law of large numbers. It will give the possibility to show two of the most important theorems on information theory, i.e., the source-coding theorem and the channel-coding theorem. In this chapter, the channel capacity will also be introduced as the possible amount of information carried in a transmission over a noisy channel. In Chapter 7, the concepts of error-correcting codes are discussed, introducing both block codes and convolutional codes.

In Chapter 8, the information measures defined in Chapter 3 for discrete variables are extended to the continuous case. The relation between the interpretations for the measures in the discrete and continuous cases is discussed. Depending on the logical level of the channel model, it can be discrete or continuous. The discrete channels were treated in Chapter 6, whereas in Chapter 9 the continuous counterpart is discussed, with a special focus on the case for Gaussian noise. To reach the channel capacity for this case, it turns out that the transmitted variable should be Gaussian distributed, whereas in reality it is often both discrete and uniformly distributed. This case is treated specially in Chapter 10, where a closer relationship is presented with

4 CHAPTER 1 INTRODUCTION

practically used modulation schemes like pulse amplitude modulation and quadrature amplitude modulation.

In Chapter 11, the concept of distortion is introduced. In reality, it is often acceptable with a certain amount of distortion, e.g., in image and video coding the human eye will tolerate distortions and in a communication scenario there might be tolerance for few errors in the transmission. This is incorporated in the previous theory by using the rate distortion functions. The chapter is concluded with a description of lossless compression and an introduction to transform decoding, used in, e.g., jpeg.

Over the years there has been several books written in the subject of information theory, e.g., [11, 13, 67, 71, 80, 81, 84, 86, 92]. The aim of this book is to present the theory for both discrete and continuous variables. It also aims at applying the theory towards communication theory, which is especially seen in Chapter 10 where discrete input Gaussian channels are considered.