1 • Introduction

Telecommunication systems make intensive use of speech coders. In wireless systems, where bandwidth is limited, speech coders provide one of the enabling technologies to reach more users and furnish better services. In wireline systems, where bandwidth can be less of an issue, speech is also digitized and compressed to a certain extent depending on the system.

In general, a speech encoder takes sampled speech at its input and produces a compressed bitstream for transmission or storage. At the receiver (or at playback from a storage device), a speech decoder takes the bitstream and produces a synthesized speech signal that is desired to sound as close to the original speech as possible. This is illustrated in Figure 1. The compression ratio of the speech encoder and the quality of the synthesized speech depend on the system requirements and application needs. Very often, the application requirement is that synthesized speech should have quality as high as possible with a bit rate as low as possible. Any voice application with limited channel capacity and/or memory space for storage can greatly benefit from existing low-rate speech coders.

The main objective of this chapter is to give the reader an appreciation of the ways signal processing is used inside and around modern speech coders to meet specific application needs and to solve acoustical problems in the communication link. Examples will focus mostly on standards since these are well documented and thoroughly tested. Only the main concepts of speech coders
will be presented here. For in-depth reviews of speech coders, the reader is referred to the several excellent tutorials written on the subject. For example, see [1,2].

The outline of this chapter is as follows. In Section 2, principles of modern speech coders are briefly reviewed. Section 3 presents methods to address issues related to input conditions. These include the presence of acoustical noise, the effect of audio bandwidth and the variability of the input signal characteristics. In Section 4, problems related to channel conditions, such as packet loss and variable transmission delay, are addressed. Finally, Section 5 discusses current developments and future challenges.

2 • Speech Coding Concepts

Advances in speech coding span several decades. It would be impractical to present here all the advances that have been made in this field. However, looking at most voice communication systems using speech coders today, we observe that the speech coders used in most systems are based on variants of linear predictive (LP) coding [3]. The basic principle in LP coding is to model the speech signal using a time-varying linear filter with a proper excitation signal at the input of the filter. The time-varying linear filter is often called the synthesis filter in the sense that it produces the synthetic speech, provided it has an appropriate excitation at its input. Moreover, the excitation signal is very often modeled as the weighted sum of two components: one component selected from an adaptive codebook and the other component from the innovative codebook. The adaptive codebook is nothing more than the past excitation signal, which is also available at the decoder – or at least approximately if there are channel errors. The adaptive codebook has thus a content which varies in time, as the past excitation is shifted after encoding new speech samples. The innovative codebook is also called the fixed codebook. In the simplest form, its content does not change.