1

Introduction

Language is the most essential means of human communication. It is used in two modes: as spoken language (*speech communication*) and as written language (*textual communication*). In our modern information society both modes are greatly enhanced by technical systems and devices. E-mail, short-messaging, and the worldwide web have revolutionized textual communication while

- digital mobile radio systems,
- acoustic human–machine communication, and
- digital hearing aids

have significantly expanded the possibilities and convenience of speech communication.

Digital processing of speech signals for the purpose of transmission (or storage) is a branch of information technology and an engineering science which draws on various other disciplines such as physiology, phonetics, linguistics, acoustics, and psychoacoustics. It is this multidisciplinary aspect which makes digital speech processing a challenging as well as rewarding task.

The goal of this book is a comprehensive discussion of fundamental issues, standards, and recent trends in speech communication technology. Speech communi-

Digital Speech Transmission: Enhancement, Coding and Error Concealment
Peter Vary and Rainer Martin

cation technology helps to mitigate a number of physical constraints and technological limitations, most notably

- bandwidth limitations of the telephone channel,
- shortage of radio frequencies,
- acoustic background noise,
- interfering acoustic echo signals from loudspeaker(s), and
- (residual) transmission errors caused by the radio channel.

The enormous advances in signal processing technology have contributed to the success of speech signal processing. At present, integrated digital signal processors allow economic real-time implementations of complex algorithms which require several thousand operations per speech sample. For this reason advanced speech signal processing functions can be implemented in cellular phones as illustrated in Fig. 1.1.

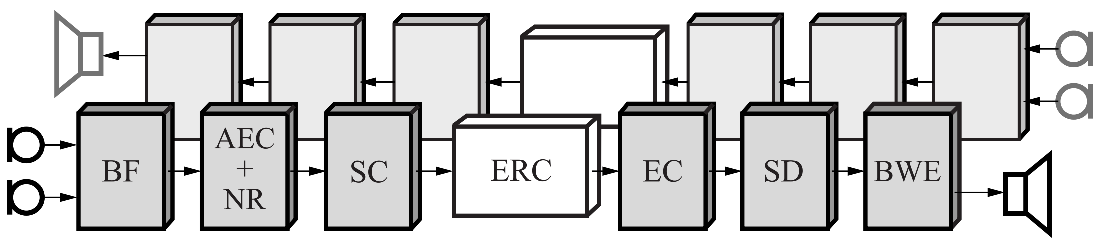

Figure 1.1: Speech signal processing in a handsfree mobile terminal
BF : Beamforming
AEC : Acoustic Echo Cancellation
NR : Noise Reduction
SC : Speech Coding
ERC : Equivalent Radio Channel
EC : Error Concealment
SD : Speech Decoding
BWE: Bandwidth Extension

The handsfree terminal in Fig. 1.1 facilitates communication via a microphone and a loudspeaker. Handsfree telephone facilities are installed in motor vehicles in order to enhance road safety and to increase convenience in general.

At the near end of the transmission system, three different pre-processing steps are taken to improve communication in the presence of ambient noise and loudspeaker signals. In the first step, two or more microphones are used to enhance the near-end speech signal by **beamforming** (BF). Specific characteristics of the interference, such as the spatial distribution of the sound sources and the statistics of the spatial sound field, are exploited.

Acoustic echoes occur when the far-end signal leaks from the loudspeaker of the handsfree set into the microphone(s) via the acoustic path. As a consequence, the far-end speaker will hear his or her own voice delayed by twice the signal propagation time of the telephone network. Therefore, in a second step, the acoustic echo must be compensated by an adaptive digital filter, the **acoustic echo canceller** (AEC).

The third module of the pre-processing chain is **noise reduction** (NR). Single channel noise reduction systems are most effective for short-term stationary noise. They are based on optimal filters and estimation techniques.

Speech coding (SC), **error concealment** (EC), and **speech decoding** (SD) facilitate the efficient use of the mobile radio channel. Speech coding algorithms for mobile communications with typical bit rates between 4 and 13 bit/s are explicitly based upon a model of speech production and exploit properties of the hearing mechanism.

At the receiving side of the transmission system, the speech quality is ensured by means of error correction (channel decoding), which is not within the scope of this book. In Fig. 1.1 the (inner) channel coding/decoding as well as modulation/demodulation and transmission over the physical radio channel are modeled as an **equivalent radio channel** (ERC). In spite of channel coding, quite frequently residual errors remain. The negative auditive effects of these errors can be reduced by **error concealment** (EC) techniques. In many cases, these effects can be reduced by exploiting both residual source redundancy and information about the instantaneous quality of the transmission channel.

Finally, the decoded signal might be subjected to artificial **bandwidth extension** (BWE) which expands narrowband (0.3 – 3.4 kHz) to wideband (0.05 – 7.0 kHz) speech. With the introduction of true wideband speech coding into telephone networks this step will be of significant importance as, for a long transition period, narrowband and wideband speech terminals will coexist.

Some of these processing functions find applications in multimedia terminals and digital hearing aids.

The book is organized as follows. The first part (Chapters 2–5) deals with the *fundamentals* of speech processing: models of speech production and hearing, spectral transformations, filter banks, and stochastic processes.

The second part (Chapters 6–8) covers the issue of *speech coding*. Quantization, differential waveform encoding, linear prediction, and especially the concepts of code excited linear prediction (CELP) are discussed. Finally, some of the most relevant speech codec standards are presented. Recent developments such as the *Adaptive Multi-Rate* (AMR, narrowband and wideband) codec for GSM and UMTS or variable rate coding for Internet telephony are addressed.

The third part of the book (Chapters 9–13) is concerned with the measures of speech enhancement of Fig. 1.1: error concealment, single channel noise reduction, acoustic echo cancellation, multi-channel noise reduction and beamforming, and, finally, artificial bandwidth extension.