Chapter 1

Introduction

Rainer Martin, Ulrich Heute, Christiane Antweiler

In the era of mobile communication networks, of Voice over IP (VoiP), and of hands-free voice interfaces, new opportunities as well as new challenges arise. While traditional voice services are confined to the rather narrow frequency range below 4 kHz, new technologies enable the transition to higher bandwidth and thus better quality speech transmission systems. At the same time, speech communication systems are increasingly used in *adverse acoustic conditions*, i.e., noisy and reverberant environments. It turns out that a convincing end-to-end quality improvement is obtained only when all components of the transmission chain – analog and digital – are optimized to a comparable level of quality. This entails improved methods for acoustic front-end processing as well as for speech coding and transmission. As a consequence, new wideband speech coding systems, new speech enhancement and error concealment algorithms, and new quality assessment methods have emerged.

Another important application in digital speech transmission is hearing instruments that comprise increasingly powerful digital processors. Advanced speech processing algorithms are currently integrated into hearing aids and will result in significant improvements for hearing-impaired people.

Likewise, the recognition performance of automatic speech recognizers (ASR) can also be improved dramatically if models of noise and reverberation are integrated. Furthermore, a satisfying user experience requires these systems to become aware of *who* is actually using them and to adapt to the needs of specific users. Many challenging tasks in the conceptual design and in the signal processing modules of modern speech transmission systems need to be solved before a uniformly pleasant user experience is accomplished.

Advances in Digital Speech Transmission Edited by R. Martin, U. Heute and C. Antweiler

The general theme of this book is the presentation and the analysis of solutions for improved-quality design of speech transmissions under adverse acoustic and heterogeneous network conditions. The book is organized into six parts:

I. **Speech Quality Assessment**
II. **Adaptive Algorithms in Acoustic Signal Processing**
III. **Speech Coding for Heterogeneous Networks**
IV. **Joint Source-Channel Coding**
V. **Speech Processing in Hearing Instruments**
VI. **Speech Processing for Human–Machine Interfaces,**

which will be briefly introduced below. Each chapter comes with an extensive list of references that may serve as a resource for further study.

Intrinsically related to the *Advances in Digital Speech Transmission* is the question of how to measure the quality of voice communication systems. This topic is treated in **Part I, Assessment of Speech Quality**. In his chapter on ***Speech-Transmission Quality: Aspects and Assessment for Wideband vs. Narrowband Signals*** Ulrich Heute discusses the impact of increasing the signal bandwidth first on speech processing algorithms in general and secondly on the methods for quality assessment in particular. He argues convincingly that the traditional total quality assessment approach, for instance, in the form of total quality listening tests and *mean opinion scores* (MOS) should be succeeded by quality assessment procedures with diagnostic abilities. Therefore, the task is to find more or less orthogonal descriptors for wideband speech quality that provide a basis for the computation of total quality scores, and to develop algorithms for the computational assessment of speech signals. In his chapter on ***Parametric Quality Assessment of Narrowband Speech in Mobile Communication Systems***, Marc Werner provides an overview of and insights into automated *non-intrusive* quality monitoring for mobile voice communication channels. His approach is based on mapping measurable system parameters at the receiving end to speech quality measures. Such parameters are, for instance, the signal power or the frame-error and the bit-error rate (FER / BER). The quality measure is then computed via a linear function of these parameters whose coefficients are optimized in the minimum mean-square error sense. Speech transmissions in the GSM and the UMTS systems serve as examples.

Part II is dedicated to **Adaptive Algorithms in Acoustic Signal Processing**. It begins with Gerald Enzner's chapter on ***Kalman Filtering in Acoustic Echo Control: A Smooth Ride on a Rocky Road***, which takes a fresh look at the acoustic echo control problem and develops a model-based approach. By dropping the assumption of a deterministic echo-path model and by replacing it with a statistical model, he arrives at a unifying solution to the acoustic echo-control problem. His solution comprises an echo canceler and a post-filter and turns out to be very robust to echo-path variations, while at the same time it is conceptually elegant and simple to implement. The chapter ***Noise Reduction – Statistical Analysis and Control of Musical Noise*** by Colin Breithaupt and Rainer Martin investigates the statistical fluctuations in the spectral parameters of state-of-the-art noise reduction systems. A careful analytic analysis of the most widely used spectral enhancement approaches explains the emergence of musical noise in these algorithms. It is shown that the histogram of log-spectral amplitudes provides a good indication of audible spectral fluctuations. Furthermore, a solution is presented for controlling these fluctuations without impairing the perceived speech quality. In their chapter ***Acoustic Source Localization with Microphone Arrays***, Nilesh Madhu and Rainer Martin provide an overview on time delay of arrival (TDOA) and source localization techniques. This includes the popular generalized cross-correlation (GCC) method as well as multi-microphone techniques such as steered response power (SRP), multiple signal classification (MUSIC), and maximum-likelihood (ML) approaches. The chapter elaborates on the links between these methods and provides simulation examples highlighting their respective performance. The chapter ***Multi-Channel System Identification with Perfect Sequences – Theory and Applications*** by Christiane Antweiler explains how sequences with perfect correlation properties can be used to identify multiple input – single output (MISO) systems. Perfect sequences are a most elegant tool to be used in conjunction with the normalized least mean-square (NLMS) algorithm. This chapter extends this method to the real-time identification of multiple channels and provides an example in medical technology: the online assessment of the Eustachian tube.

Part III, Speech Coding for Heterogeneous Networks, is opened by an overview on ***Embedded Speech Coding: From G.711 to G.729.1*** by Bernd Geiser, Stéphane Ragot, and Hervé Taddei. The authors review the general theory and the methods for successive refinement coding of speech. Tree-structured vector quantization (TSVQ), multi-stage vector quantization (MSVQ), and transform domain vector quantization with progressive decoding and the application in state-of-the-art speech coders are discussed in detail. Furthermore, this chapter summarizes the latest developments in embedded wideband coding for VoiP speech transmission systems and the network aspects of such schemes. Bandwidth extension (BWE) of speech signals constitutes another important research area in speech signal processing that currently enjoys renewed and significant interest. In Chap. 9, ***Backwards Compatible Wideband Telephony***, Peter Jax discusses the implications of this technology for the migration from narrowband to wideband speech transmission systems and outlines his approach to this task. He explains solutions that do not need side-information (stand-alone BWE) and those that make use of information about the wideband speech envelope. If the side-information is embedded into the speech signal by means of watermarking technology, the result is a speech signal that can be reproduced either on a narrowband or on a wideband terminal.

Part IV, Joint Source-Channel Coding, presents a series of four chapters that deal with the exploitation of redundancies in speech and channel coding schemes for improving the quality of the received signal. Tim Fingscheidt's chapter on ***Parameter Models and Estimators in Soft Decision Source Decoding*** first presents an overview of soft decision source decoding (SDSD) and of modeling techniques for source parameters. Secondly, it discusses estimators for the recovery of degraded, received speech parameters by means of extrapolation and interpolation and presents simulation results for various transmission scenarios. In Chap. 11, this work is extended towards a general discussion of ***Optimal MMSE Estimation for Vector Sources with Spatially and Temporally Correlated Elements*** by Stefan Heinen and Marc Adrat. While the general solution of this task would lead to an overwhelming computational complexity, the authors tackle this problem by imposing a Markov property on the speech parameters in the temporal *and* the vector dimensions. Thus, they are capable of deriving MMSE and near-optimal MMSE estimators with significantly reduced computational requirements. Furthermore, they provide simulation results for digital audio broadcast (DAB) and GSM systems. In their chapter on ***Source Optimized Channel Codes & Source Controlled Channel Decoding*** Stefan Heinen and Thomas Hindelang establish the link between source and channel (de-)coding in two competing approaches. In the source-optimized channel coding (SOCC) system the channel codes are tailored to the source statistics. It thus achieves an efficient utilization of the available bit rate and a low reconstruction error at the receiver. The second approach of source-controlled channel decoding (SCCD) improves the decoding of convolutional codes by exploiting residual redundancies in source parameters. The authors compare and discuss the merits of both approaches on the basis of a single, general source and transmission channel model and are thus able to draw interesting conclusions. Part IV concludes with the chapter on ***Iterative Source-Channel Decoding & Turbo DeCodulation*** by Marc Adrat, Thorsten Clevorn, and Laurent Schmalen, which builds on the two approaches of the previous chapter. Here, the authors extend soft source and channel decoding techniques towards iterative methods, also known as *turbo*-decoding methods. Besides a review of turbo-techniques, two novel approaches, iterative source-channel decoding (ISCD) and Turbo DeCodulation (TdeC), are developed and analyzed by means of extrinsic information transfer (EXIT) charts. It is shown that ISCD outperforms non-iterative transmission schemes in terms of the signal-to-noise ratio of the reconstructed speech signal.

An interesting area for research in digital speech transmission is the application of **Speech Processing in Hearing Instruments**. **Part V** is dedicated to this topic. The first chapter in this part is authored by Volkmar Hamacher, Ulrich Kornagel, Thomas Lotter, and Henning Puder and discusses ***Binaural Signal Processing in Hearing Aids: Technologies and Algorithms***. The realization of a wireless data link between the left ear and the right ear hearing device is a challenge in itself but also enables the employment of binaural processing schemes. The authors first describe the design of the wireless link as implemented in a commercial hearing aid. Then, they discuss the potential that lies in the possibility to exchange data at various bit rates between the left and the right side. They show that considerable potential resides in these processing schemes, especially in terms of user comfort. Advanced signal processing schemes such as binaural beamformers, however, require higher data rates. Thus, many interesting research questions are still to be answered before such methods can be applied. In Chap. 15, ***Auditory-profile-based Physical Evaluation of Multi-microphone Noise Reduction Techniques in Hearing Instruments***, by Koen Eneman, Arne Leijon, Simon Doclo, Ann Spriet, Marc Moonen, and Jan Wouters discusses multi-microphone beamforming approaches for hearing devices and introduces an assessment method for noise reduction algorithms that takes the speech perception capabilities of hearing impaired listeners into account. This evaluation is based on the *auditory profile* of the listener, i.e., the characterization of the hearing loss in terms of various measures such as the audiogram. The newly developed instrumental measures provide a prediction of the usefulness of signal enhancement algorithms for people with hearing deficiencies. It can be expected that such measures will reduce the effort that comes with listening tests and thus will facilitate the development of new and more effective algorithms.

Part VI deals with **Speech Processing for Human–Machine Interfaces**. Voice driven human–machine interfaces are typically used in a hands-free mode. Thus, the automatic speech recognizer (ASR) has to cope with significant levels of noise and reverberation. In Chap. 16, Hans-Günter Hirsch discusses ***Automatic Speech Recognition in Adverse Acoustic Conditions***, which entail ambient noise and reverberation. He gives an overview on state-of-the-art approaches and presents a new database that was developed to allow the evaluation and comparison of recognition experiments in noisy and reverberant conditions. Furthermore, he presents a new model adaptation approach that dramatically improves the recognition performance in such environments. The topic of the final chapter is ***Speaker Classification for Next-Generation Voice-Dialog Systems***. In this chapter, Felix Burkhardt, Florian Metze, and Joachim Stegmann discuss how voice services may be personalized when the speaker can be classified as belonging to a certain target group. Useful classification criteria are, for instance, the age and the gender of the caller, or they may be related to the emotions of the caller when accessing the dialog system. The authors provide an overview on classification methods and present corresponding algorithms. They show how these classification criteria can be used in dialog systems and present evaluation results for their specific approach.

The editors believe that each of these chapters by itself will serve as a valuable resource and reference for students and researchers and that cross links between these topics will trigger new ideas and thus contribute to the progress of the field.