1 Introduction

Eberhard Hänsler[1] and Gerhard Schmidt[2]

[1] Technische Universität Darmstadt, Germany
[2] Harman/Becker Automotive Systems, Germany

If people would speak digitally speech processing would lack some of its most challenging problems.

One of those tasks is to provide means for a comfortable conversation with a remote partner where one of them or both are in adverse environments. By "adverse environment" we mean noisy offices, railway stations, airports, shop floors, etc. Similar problems have to be solved when a speech recognition system is used. Under "comfortable environment" we understand that a speaker does not have to be "wired", i.e. to carry or to hold a microphone very close to his mouth. He should be able "just to talk" without caring where the microphone(s) is/are located. His partner at a remote location or a speech recognition system should just receive his speech signal. In case of speech recognition background noise should be suppressed as much as possible. For a human listener it may be desirable to communicate some information about the environment his partner stays in.

Systems for speech enhancement have to perform at least three main functions: echo canceling, noise suppression and speech restoration. For all three there is no clear-cut solution. Mathematical approaches have to start with models that are simplified to a high degree and such are a very rough approximation to the reality only. Necessarily, a good deal of heuristics has to enter the solution in order to match it to the real world. Thus, there are no break throughs to final solutions in any one of the subproblems. Advanced technology and cheaper hardware will always stimulate researchers and industrial developers to come up with more sophisticated methods that promise better results. The appearance of powerful simulation tools and high-capacity personal computers over the last decades have speeded up this process. The simulation and the real-time verification of algorithms do no longer require costly dedicated soft- and hardware.

This book provides an overview of recent developments and new results reported from the key researchers in speech and audio processing.

1.1 Overview about the Book

The succeeding seventeen chapters are organized in five parts. This introduction is followed by chapters on *Speech Enhancement* in Part one. *H. Löllmann* and *P. Vary* review the design of uniform and non-uniform analysis-synthesis filter-banks employed sub-band processing of speech and audio signals. Their main aim is to achieve low signal delay. They introduce and analyse the concept of the filter-bank equalizer. In certain applications of noise reduction information originally generated for other purposes can be used to support the solution of a problem. In case of noise reduction in a passenger car the exact speed of the engine is provided by a bus-system in the car. *H. Puder* discusses a pre-filter to reduce harmonics that are proportional to the speed of the engine from car noise. *M. Krini* and *G. Schmidt* improve noisy speech signals by partially reconstructing the spectrum. By overlaying the conventionally noise reduced signal and the reconstructed one it is possible to avoid the robot-like sound of pure synthesized speech signals. Telephone signals are transmitted with a reduced bandwidth. *B. Iser* and *G. Schmidt* explain how the bandwidth of a transmitted speech signal can be extended at the receiver, such, that a listener feels a more natural sound. Reverberations often degrade the fidelity and intelligibility of speech signals and decrease the performance of automatic speech recognition devices. *E. A. P. Habets*, *S. Gannot*, and *I. Cohen* develop a post-processor for the joint suppression of the residual echo, the background noise and the reverberation. *A. Sugiyama* assumes the existence of a reference microphone and extends this classical noise reduction technique for the case where the crosstalk from the primary signal source to the auxiliary source can not be neglected. He describes the application of this technique together with a speech recognition system in a human-robot communication scenario.

Part two of the book deals with nonlinear *Echo Cancellation*. *O. Hoshuyama* and *A. Sugiyama* address the nonlinearity of the echo path in hands-free cellphones. Since common approaches like Voltera filters are too demanding for such devices they propose a nonlinear echo canceller based on the correlation between spectral amplitudes of the residual echo and the echo replica.

In Part three, *Signal and System Quality Evaluation*, two chapters are concerned with diagnosing the quality of speech signals. The ultimate criterion here is the judgment of human listeners. Tests of this kind, however, are time consuming and costly and not free of problems. *U. Heute* explains these tests for various scenarios and also explains possibilities for automatic quality measurement not involving human listeners. The chapter by *F. Kettler* and *H.-W. Gierlich* primarily focuss on hands-free terminals installed in passenger cars. They describe subjective tests and the necessity of objective laboratory tests. They show how the the scores for different aspects of a hands-free system can be documented by a “quality pie”.

PART FOUR deals with topics on *Multi-Channel Processing.* The use of microphone and loudspeaker arrays opens a new dimension in speech and audio signal processing. *J. Scheuring* and *B. Yang* estimate the time difference of arrival in a multi-source reverberant environment. They resolve the ambiguity by using the facts that the cross-correlation maxima from a direct path and the echo show the same distance than the extrema in the autocorrelation of the microphone signals and that the cyclic sum of the time differences of arrival of all microphone pairs is zero. The performance of microphone arrays may degrade severely due to mismatched microphones. *M.* Buck, T. Haulick and *H.-J. Pfleiderer* introduce a new model that allows to study the effect of differences of microphones in an array mathematically and by simulation. They discus methods for fixed and for adaptive calibration of the microphones where in case of mass production adaptive calibration is clearly preferable. Convolutive blind source separation for noisy mixtures is the concern of the chapter by *R. Aichner, H. Buchner* and *W. Kellermann.* In contrast to conventional procedures no a-priori knowledge about source and sensor positions are required. Their method combines pre- and postprocessing algorithms such, that residual cross-talk and background noise can be handled. The results are confirmed by experiments. In their chapter on binaural speech segregation *N. Roman* and *D. L. Wang* describe the principles of binaural processing. Their special interest is an automatic sound separation in a realistic environment. Solutions of this problem are of interest in many applications of speech and speaker recognition. Substantial improvements are achieved by utilizing only reliable target dominant features and by a target reconstruction method for unreliable features. In the final chapter of this part of the book *S. Spors*, H. Buchner and *R. Rabenstein* gives a unified description of spatio-temporal adaptive methods of sound reproduction and trace them back to the problem of inverse filtering. They introduce eigenspace adaptive filters to decouple the multichannel problem. An exact solution, however, would need data-dependant transformations. Therefore, wave-domain adaptive filtering serves as an approximate solution.

Selected Applications are described in PART FIVE. *K. Wiklund* and *S. Haykin* report on a system that allows to test algorithms designed for hearing aids. It allows to simulate the real acoustic environment and the impairments of patients. Thus, hearing aids can be tuned to the needs of patients and the amount of time consuming and costly real life tests can be reduced. Automobiles provide very undesirable acoustic environments. On the other hand, passengers request pleasant acoustic conditions for listening to audio programs, to carry hands-free telephone calls, or talking to other passengers. *M. Christoph* shows how – under a practical point of view – a vehicle dependent tuning can be accomplished. Since this is done when the engine is off and the car is not moving control procedures are required to maintain the acoustic quality during the time the car is operated. Inputs for this algorithms can be signals that are available from the car-electronics or additional microphones in the passenger compartment. *A. Sehr* and *W. Kellermann* address the problem of automatic speech recognition in reverberant environments. Reverberation disperses a signal and thus, it cannot be modelled by an additive or a multiplicative term. Conventional methods for dereverberation are described. A new concept called reverberation modelling that combines the advantages of the former methods is introduced.