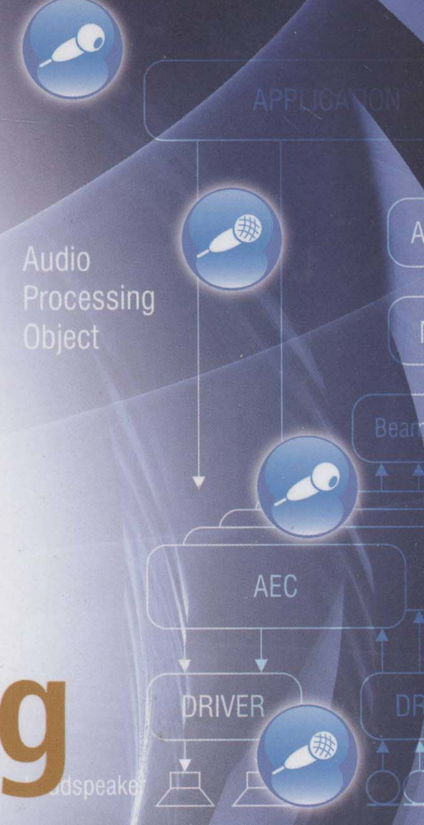
IVAN TASHEV

Sound Capture and Processing

PRACTICAL APPROACHES



Companion Website

WILEY

Sound Capture and Processing
Practical Approaches

Ivan J. Tashev
Microsoft Research, USA

WILEY
A John Wiley and Sons, Ltd., Publication

This edition first published 2009

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Library of Congress Cataloging-in-Publication Data

Tashev, Ivan J. (Ivan Jelev)
Sound capture and processing : practical approaches / Ivan J. Tashev.
p. cm.
Includes index.
ISBN 978-0-470-31983-3 (cloth)
1. Speech processing systems. 2. Sound–Recording and reproducing–Digital techniques. 3. Signal processing–Digital techniques. I. Title.
TK7882.S65T37 2009
621.382'8–dc22

2009011987

A catalogue record for this book is available from the British Library.

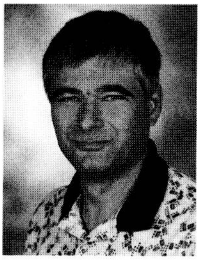
ISBN 978-0-470-31983-3 (H/B)

Typeset in 11/13pt Times by Thomson Digital, Noida, India.
Printed and bound in Great Britain by CPI Antony Rowe, Chippenham, Wiltshire.

Sound Capture and Processing

To my family: the time to write this book was taken from them

About the Author



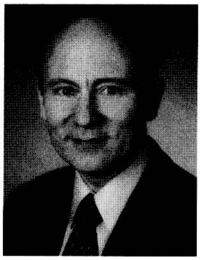
Dr Ivan Tashev took both his Engineering Diploma in Electronics and PhD in Computer Science degrees at the Technical University of Sofia, Bulgaria, in 1984 and 1990 respectively. After his graduation he worked as R&D engineer and researcher in the R&D Department of the same university. Dr Tashev became assistant professor in 1989. He created and taught two courses, "Data and signal processing" and "Programming of real-time systems" to the students of fourth and fifth year in the Department of Electronics.

Dr Tashev joined Microsoft in 1998 and held positions in various product teams until 2001 when he moved to Microsoft Research. Here he was involved in projects such as RingCam (now a Microsoft product – Round Table Device), microphone array (currently part of Windows Vista), and many others related to sound capturing devices and audio signal processing. Currently he is a member of the Speech Technology Group in Microsoft Research lab at the Microsoft headquarters in Redmond, Washington.

Dr Ivan Tashev is senior member of IEEE and IEEE Signal Processing Society, member of Audio Engineering Society and its Pacific Northwest Committee. He is reviewer for most of the audio and signal processing journals and conferences. Dr Tashev has published three books, more than fifty scientific papers and is listed as inventor of five granted U.S. patents and seventeen U.S. patent applications.

The research interests of Dr Tashev include sound capturing devices, signal processing for arrays of transducers, speech enhancement algorithms, and signal processing of audio, speech and biological signals.

Foreword



Just a couple of decades ago we would think of "sound capture and processing" as the problems of designing microphones for converting sounds from the real world into electrical signals, as well as amplifying, editing, recording, and transmitting such signals, mostly using analog hardware technologies. That's because our intended applications were mostly analog telephony, broadcasting, and voice and music recording. We have come a long way: small digital audio players have replaced bulky portable cassette tape players, and people make voice calls mostly via digital mobile phones and voice communication software in their computers. Thanks to the evolution of digital signal processing technologies, we now focus mostly on processing sounds not as analog electrical signals, but rather as digital files or data streams in a computer or digital device. We can do a lot more with digital sound processing, such as transcribe speech into text, identify persons speaking, recognize music from humming, remove noises much more efficiently, add special effects, and so much more. Thus, today we think of sound capture as the problem of digitally processing the signals captured by microphones so as to improve their quality for best performance in digital communications, broadcasting, recording, recognition, classification, and other applications.

This book by Ivan Tashev provides a comprehensive yet concise overview of the fundamental problems and core signal processing algorithms for digital sound capture, including ambient noise reduction, acoustic echo cancellation, and reduction of reverberation. After introducing the necessary basic aspects of digital audio signal processing, the book presents basic physical properties of sound and propagation of sound waves, as well as a review of microphone technologies, providing the reader with a strong understanding of key aspects of digitized sounds. The book discusses the fundamental problems of noise reduction, which are usually solved via techniques based on statistical models of the signals of interest (typically voice) and of interfering signals. An important discussion of properties of the human auditory system is also presented; auditory models can play a very important role in algorithms for enhancing audio signals in communication and recording/playback applications, where the final destination is the human ear.

Microphone arrays have become increasingly important in the past decade or so. Thanks to the rapid evolution and reduction in cost of analog and digital electronics in recent years, it is inexpensive to capture sound through several channels, using an array of microphones. That opens new opportunities for improving sound capture, such as detecting the direction of incoming sounds and applying spatial filtering techniques. The book includes two excellent chapters whose coverage goes from the basics of microphone array configurations and delay-and-sum beamforming, to modern sophisticated algorithms for high-performance multichannel signal enhancement.

Acoustic echoes and reverberation are the two most important kinds of signal degradations in many sound capture scenarios. If you're a professional singer, you probably don't mind holding a microphone or wearing a headset with a microphone close to your mouth, but most of us prefer microphones to be invisible, far away from our mouths. That means microphone will capture not only our own voices, but also reverberation components because of sound reflections from nearby walls, as well as echoes of signals that are being played back from loudspeakers. Removing such undesirable artifacts presents significant technical challenges, which are well addressed in the final two chapters, which present modern algorithms for tackling them.

A key quality of this book is that it presents not only fundamental theoretical analyses, models, and algorithms, but it also considers many practical aspects that are very important for the design of real-world engineering solutions to sound capture problems. Thus, this book should be of great appeal to both students and engineers.

I have had the pleasure of working with Ivan on research and development of sound capture systems and algorithms. His enthusiasm, deep engineering and mathematical knowledge, and pragmatic approaches were all contagious. His work has had significant practical impact, for example the introduction of multichannel sound capture and processing modules in the Microsoft Windows operating system. I have learned a considerable amount about sound capturing and processing from my interactions with Ivan, and I am sure you will, as well, by reading this book. Enjoy!

Henrique Malvar
Managing Director
Microsoft Research
Redmond Laboratory

Preface

Capturing and processing sounds is critical in mobile and handheld devices, communication systems, and computers using automatic speech recognition. Devices and technologies for proper conversion of sounds to electric signals and removing unwanted parts, such as noise and reverberation, have been used since the first telephones. They evolved, becoming more and more complex. In many cases the existing algorithms exceed the abilities of typical processors in these devices and computers to provide real-time processing of the captured signal.

This book will discuss the basic principles for building an audio processing stack, sound capturing devices, single-channel speech-enhancement algorithms, and microphone arrays for sound capture and sound source localization. Further, algorithms will be described for acoustic echo cancellation and de-reverberation – building blocks of a sound capture and processing stack for telecommunication and speech recognition. Wherever possible the various algorithms are discussed in the order of their development and publication. In all cases the aim is to try to give the larger picture – where the technology came from, what worked and what had to be adapted for the needs of audio processing. This gives a better perspective for further development of new audio signal processing algorithms.

Even the best equations and signal processing algorithms are not worth anything before being implemented and verified by processing of real data. That is why, in this book, stress is placed on experimenting with recorded sounds and implementation of the algorithms. In practice, frequently a simpler model with fewer parameters to estimate works better than a more precise but more complex model with a larger number of parameters. With the latter one has either to sacrifice estimation precision or to increase the estimation time. This balance of simplicity, precision, and reaction time is critical for real-time systems, where on top of everything we have to watch out for parameters such as latency, consumed memory, and CPU time.

Most of the algorithms and approaches described in this book are based on statistical models. In mathematics, a single example cannot prove but can disprove a theorem. In statistical signal processing, a single example is . . . just a sample. What matters is careful evaluation of the algorithms with a good corpus of speech or audio signals, distributed in their signal-to-noise ratios, type of noise, and other parameters – as close as possible to the real problem we are trying to solve.

The solution of practically any signal processing problem can be improved by tuning the parameters of the algorithm, provided we have a proper criterion for optimality. There are always adaptation time constants, thresholds, which cannot be estimated and their values have to be adjusted experimentally. The mathematical models and solutions we use are usually optimal in one or another way. If they reflect properly the nature of the process they model, then we have a good solution and the results are satisfactory. In all cases it is important to remember that we do not want a "minimum mean-square error solution," or a "maximum-likelihood solution," or even a "log minimum mean-square error solution." We do not want to improve the signal-to-noise ratio. What we want is for listeners to perceive the sound quality of the processed signal as better – improved – compared to the input signal. From this perspective, the final judge of how good is an algorithm is the human ear, so use it to verify the solution. Hearing is an important sense for humans and animals. In many places in this book are provided examples of how humans and animals hear and localize sounds – this explains better some signal processing approaches and brings biology-inspired designs for sound capture and processing systems.

In many cases the signal processing chain consists of several algorithms for sound capture and speech enhancement. The practice shows us that a sequence of separately optimized algorithms usually provides suboptimal results. Tuning and optimization of the designed sound capturing system end-to-end is a must if we want to achieve best results.

For further information please visit http://www.wiley.com/go/tashev_sound

Ivan Tashev
Redmond, WA
USA

Acknowledgements

I want to thank the Book Program in MathWorks and especially Dee Savageau, Naomi Fernandes, and Meg Vulliez for the help and responsiveness. The MATLAB® scripts, part of this book, were tested with MATLAB® R2007a, provided as part of this program.

I am grateful to my colleagues from Microsoft Research Alex Acero, Amitav Das, Li Deng, Dinei Florencio, Cormac Herley, Zicheng Liu, Mike Seltzer, and Cha Zhang. They read the chapters of this book and provided valuable feedback.

And last, but not least, I want to express my great pleasure working with the nice and helpful people from John Wiley & Sons, Ltd. During the long process from proposal, through writing, copyediting, and finalizing the book with all the details, they were always professional, understanding, and ready to suggest the right solution. I was lucky enough to work with Tiina Ruonamaa, Sarah Hinton, Sarah Tilley, and Catlin Flint – thank you all for everything you did during the process of writing this book!

Contents