CAMBRIDGE

Cambridge University Press
978-0-521-51954-0 - Applied Speech and Audio Processing: With Matlab® Examples
Ian McLoughlin
Frontmatter
More information

Applied Speech and Audio Processing: With MATLAB® Examples

Applied Speech and Audio Processing is a MATLAB-based, one-stop resource that blends speech and hearing research in describing the key techniques of speech and audio processing.

This practically orientated text provides MATLAB examples throughout to illustrate the concepts discussed and to give the reader hands-on experience with important techniques. Chapters on basic audio processing and the characteristics of speech and hearing lay the foundations of speech signal processing, which are built upon in subsequent sections explaining audio handling, coding, compression and analysis techniques. The final chapter explores a number of advanced topics that use these techniques, including psychoacoustic modelling, a subject which underpins MP3 and related audio formats.

With its hands-on nature and numerous MATLAB examples, this book is ideal for graduate students and practitioners working with speech or audio systems.

Ian McLoughlin is an Associate Professor in the School of Computer Engineering, Nanyang Technological University, Singapore. Over the past 20 years he has worked for industry, government and academia across three continents. His publications and patents cover speech processing for intelligibility, compression, detection and interpretation, hearing models for intelligibility in English and Mandarin Chinese, and psychoacoustic methods for audio steganography.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-51954-0 - Applied Speech and Audio Processing: With Matlab® Examples
Ian McLoughlin
Frontmatter
More information

Applied Speech and Audio Processing

With MATLAB® Examples

IAN MCLOUGHLIN

School of Computer Engineering
Nanyang Technological University
Singapore

CAMBRIDGE
UNIVERSITY PRESS

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-51954-0 - Applied Speech and Audio Processing: With Matlab® Examples
Ian McLoughlin
Frontmatter
More information

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521519540

© Cambridge University Press 2009

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2009

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-51954-0 hardback

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs for external or third-party internet websites referred to
in this publication, and does not guarantee that any content on such
websites is, or will remain, accurate or appropriate.

CAMBRIDGE

Cambridge University Press
978-0-521-51954-0 - Applied Speech and Audio Processing: With Matlab® Examples
Ian McLoughlin
Frontmatter
More information

Contents

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-51954-0 - Applied Speech and Audio Processing: With Matlab® Examples
Ian McLoughlin
Frontmatter
More information

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-51954-0 - Applied Speech and Audio Processing: With Matlab® Examples
Ian McLoughlin
Frontmatter
More information

Preface

Speech and hearing are closely linked human abilities. It could be said that human speech is optimised toward the frequency ranges that we hear best, or perhaps our hearing is optimised around the frequencies used for speaking. However whichever way we present the argument, it should be clear to an engineer working with speech transmission and processing systems that aspects of both speech and hearing must often be considered together in the field of vocal communications. However, both hearing and speech remain complex subjects in their own right. Hearing particularly so.

In recent years it has become popular to discuss psychoacoustics in textbooks on both hearing and speech. Psychoacoustics is a term that links the words *psycho* and *acoustics* together, and although it sounds like a description of an auditory-challenged serial killer, actually describes the way the mind processes sound. In particular, it is used to highlight the fact that humans do not always perceive sound in the straightforward ways that knowledge of the physical characteristics of the sound would suggest.

There was a time when use of this word at a conference would boast of advanced knowledge, and familiarity with cutting-edge terminology, especially when it could roll off the tongue naturally. I would imagine speakers, on the night before their keynote address, standing before the mirror in their hotel rooms practising saying the word fluently. However these days it is used far too commonly, to describe any aspect of hearing that is processed nonlinearly by the brain. It was a great temptation to use the word in the title of this book.

The human speech process, while more clearly understood than the hearing process, maintains its own subtleties and difficulties, not least through the profusion of human languages, voices, inflexions, accents and speaking patterns. Speech is an imperfect auditory communication system linking the meaning wishing to be expressed in one brain, to the meaning being imparted in another brain. In the speaker's brain, the meaning is encoded into a collection of phonemes which are articulated through movements of several hundred separate muscles spread from the diaphragm, through to the lips. These produce sounds which travel through free air, may be encoded by something such as a telephone system, transmitted via a satellite in space half way around the world, and then recreated in a different environment to travel through free air again to the outer ears of a listener. Sounds couple through the outer ear, middle ear, inner ear and finally enter the brain, on either side of the head. A mixture of lower and higher brain functions then, hopefully, recreate a meaning.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-51954-0 - Applied Speech and Audio Processing: With Matlab® Examples
Ian McLoughlin
Frontmatter
More information

It is little wonder, given the journey of meaning from one brain to another via mechanisms of speech and hearing, that we call for both processes to be considered together. Thus, this book spans both speech and hearing, primarily in the context of the engineering of speech communications systems. However, in recognition of the dynamic research being undertaken in these fields, other areas are also drawn into our discussions: music, perception of non-speech signals, auditory scene analysis, some unusual hearing effects and even analysis of birdsong are described.

It is sincerely hoped that through the discussions, and the examples, the reader will learn to enjoy the analysis and processing of speech and other sounds, and appreciate the joy of discovering the complexities of the human hearing system.

In orientation, this book is unashamedly practical. It does not labour long over complex proofs, nor over tedious background theory, which can readily be obtained elsewhere. It does, wherever possible, provide practical and working examples using MATLAB to illustrate its points. This aims to encourage a culture of experimentation and practical enquiry in the reader, and to build an enthusiasm for exploration and discovery. Readers wishing to delve deeper into any of the techniques described will find references to scientific papers provided in the text, and a bibliography for further reading following each chapter.

Although few good textbooks currently cover both speech and hearing, there are several examples which should be mentioned at this point, along with several narrower texts. Firstly, the excellent books by Brian Moore of Cambridge University, covering the psychology of hearing, are both interesting and informative to anyone who is interested in the human auditory system. Several texts by Eberhard Zwicker and Karl D. Kryter are also excellent references, mainly related to hearing, although Zwicker does foray occasionally into the world of speech. For a signal processing focus, the extensive Gold and Morgan text, covering almost every aspect of speech and hearing, is a good reference.

Overview of the book

In this book I attempt to cover both speech and hearing to a depth required by a fresh postgraduate student, or an industrial developer, embarking on speech or hearing research. A basic background of digital signal processing is assumed: for example knowledge of the Fourier transform and some exposure to discrete digital filtering. This is not a signal processing text – it is a book that unveils aspects of the arcane world of speech and audio processing, and does so with MATLAB examples where possible. In the process, some of the more useful techniques in the toolkit of the audio and speech engineer will be presented.

The motivation for writing this book derives from the generations of students that I have trained in these fields, almost each of whom required me to cover these same steps in much the same order, year after year. Typical undergraduate courses in electronic and/or computer engineering, although they adequately provide the necessary foundational skills, generally fail to prepare graduates for work in the speech and audio

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-51954-0 - Applied Speech and Audio Processing: With Matlab® Examples
Ian McLoughlin
Frontmatter
More information

signal processing field. The coverage in this book is targeted toward filling the gap. It is designed to educate, interest and motivate researchers working in this field to build their skills and capabilities to prepare for research and development in the speech and audio fields.

This book contains seven chapters that generally delve into deeper and more advanced topics as the book progresses. Chapter 2 is an introductory background to basic audio processing and handling in MATLAB, and is recommended to those new to using MATLAB for audio work. It also contains justifications for, and explanations of, segmentation, overlap and windowing, which are fundamental techniques in splitting up and handling long recordings of speech and audio.

Chapter 3 describes speech production, characteristics, understanding and handling, followed by Chapter 4 which repeats the same for hearing. Chapter 5 is concerned with the handling of audio, primarily speech, and Chapter 6 with analysis methods for speech and audio. Finally Chapter 7 presents some advanced topics that make use of many of the techniques in earlier chapters.

Arrangement of the book

Each section begins with introductory text explaining the points to be made in the section, before further detail, and usually MATLAB examples are presented and explained. Where appropriate, numbered citations will be provided to a reference list at the end of each chapter. A bibliography is also provided at the end of each chapter, containing a set of the most useful texts and resources to cover the major topics discussed in the text.

Infobox 0.1 Further information

Self-contained items of further interest, but not within the flow of the main text, are usually placed inside an infobox like this one for rapid accessibility.

Commands for MATLAB or computer entry are written in a typewriter font to distinguish them from regular text:

```
type this in MATLAB
```

All of the MATLAB commands are designed to be typed into the command window, or included as part of an m-file program. This book will not use Simulink for any of the examples, and will attempt to limit all examples to the basic MATLAB without optional toolboxes wherever possible.

It is my sincere hope that academics and industrial engineers alike will benefit from the practical and hands-on MATLAB approach taken in this book.

MATLAB® is the registered trademark of MathWorks, Inc. All references to MATLAB throughout this work should be taken as referring to MATLAB®.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-51954-0 - Applied Speech and Audio Processing: With Matlab® Examples
Ian McLoughlin
Frontmatter
More information

Acknowledgements

Kwai Yoke, Wesley and Vanessa graciously gave up portions of their time with me whilst I worked on this text. My parents encouraged me, not just for the writing (it's not as easy as it may appear), but also for my career in research and my education in general.

For my initial interest in speech and hearing, I must thank many friends and role models from HMGCC, the University of Birmingham and Simoco Telecommunications in Cambridge. In particular, Dr H. Ghafouri-Shiraz who guided me, helped me, encouraged and most of all, led by example. His own books on laser diodes and optical fibres are essential reading in those fields, his research skills are exceptional and his teaching exemplary. I would also like to thank Jim Chance for his guidance, help and supervision during my own PhD studies.

More recently, sincere thanks are due to Doug McConnell of Tait Electronics Ltd, Christchurch, and management guru Adrian Busch, for more than I could adequately explain here. The multitalented Tom Scott and enthusiastic Stefan Lendnal both enriched my first half decade in New Zealand, and from their influence I left, hopefully as a better person.

Hamid Reza Sharifzadeh kindly proofread this manuscript, and he along with my other students, constantly refined my knowledge and tested my understanding in speech and audio. In particular I would like to acknowledge the hard work of just a few of my present and past students in this field: Farzane Ahmadi, Fong Loong Chong, Ding Zhongqiang, Fang Hui, Robertus Wahendro Adi and Cedric Tio.

Moving away from home, sincere thanks are due to the coffee growers of the world who supported my writing efforts daily through the fruits (literally) of their labours.

Above all, everything in me that I count as good comes from the God who made me and leads me: all honour and glory be to Him.

© Cambridge University Press www.cambridge.org