The preface to the original edition of this book remains largely relevant and accurate, since this issue of the book has been changed only in minor ways—the majority of which are corrections to typographical and other errors in the earlier work. We are indebted to many readers of the first edition, too numerous to name, for sending us corrections and suggestions for improvement. Your attention to detail is remarkable and is the source of great and well-deserved embarrassment.

The most significant new information to be noted concerns the means for accessing speech data for use in the end-of-chapter problems. Reflecting the remarkable changes in the Internet since the first edition was published in 1993, a simple and flexible procedure for acquiring these data is now possible. The interested reader can simply access the World Wide Web home pages of either author: Deller (URL:http://www.egr.msu.edu/~deller) or Hansen (URL: http://www.cslu.colorado.edu/rspl), and then follow the link to information about this text. Procedures for downloading data are described there.

The original preface expresses our gratitude to Editor John Griffin for painstakingly nurturing this book into existence in his former editorial life at the original publishing company. The glue on the 1993 binding was not even dry before corporate mergers and reorganizations in the publishing world severed the business relationship between the authors and Mr. Griffin, but certainly not the enduring appreciation for his sometimes heroic efforts that made this book possible. In a fortuitous turn of professional events, the authors were given a second chance to work with Mr. Griffin in his current editorial position at the outstanding IEEE Press. Before the authors allowed him to commit to a second round, it seemed inhumane not to remind him of what we put him through in 1993. Realizing that a good challenge now and then will keep a professional sharp, we simply welcomed the renewed opportunity and continue to appreciate John’s support and confidence in this project.

John R. Deller, Jr.
Michigan State University
Preface

Purposes and Scope. The purposes of this book are severalfold. Principally, of course, it is intended to provide the reader with solid fundamental tools and sufficient exposure to the applied technologies to support advanced research and development in the array of speech processing endeavors. As an academic instrument, however, it may also provide the serious student of signal processing an opportunity to strengthen and deepen his or her understanding of the field through the study of one of the most important and interesting contemporary applications of signal processing concepts. Finally, by collecting a large number of contemporary topics with an extensive reference list into a single volume, the book will serve as a convenient resource for those already working in the field.

The book is written by three professors of electrical engineering. This has two implications. First, we view the book as a pedagogical tool. This means we have attempted to keep the student in mind with each sentence and with each chapter. Notation, approach, and general level have been made as uniform as possible across developments and across chapters. Second, the text is written with a clear bias toward the topics and approaches of modern electrical engineering curricula—especially signal processing, systems, and communications. Speech processing is inherently multidisciplinary, and we occasionally indicate to the reader where topics are necessarily treated superficially or not at all, and where the reader can find more information. This occurs principally in areas that would probably be labeled “Speech Science” or “Computer Science.”

Level of the Text. A certain level of sophistication with topics in signal processing and stochastic processes is assumed in this book. This background is typical of a solid senior-level course in each discipline at many American universities. Accordingly, the book is intended for use in one or more graduate-level courses. The book could conceivably be used with advanced seniors, but the instructor is urged to consider the degree of maturity that is required in these areas. A good gauge of this factor is
available in Chapter 1. Sections 1.1 and 1.2 should be comfortably considered review material for anyone who is to succeed with the text. Sections 1.3 and 1.4 need not be review, but the typical EE senior will have at least some exposure to these topics, even if specific courses in pattern recognition and information theory have not been taken. Nevertheless, Sections 1.3 and 1.4 do provide sufficient background in their respective topics for the remaining chapters, whereas Sections 1.1 and 1.2 are not intended as substitutes for relevant coursework. Section 1.5 is simply a review of some concepts that should be quite familiar to any engineering student, and it is included principally as a means for establishing notation.

Course Planning. The general topical content of the speech processing field as reflected in the book is described in Section 1.6. The instructor might wish to review that section in planning a course around the book (and to have the students read this section as an introduction to the course). Clearly, it will be impossible to cover the entire book in a single quarter or semester. We have found that in a typical semester course (15 weeks), the following can be covered: Chapter 1 (Background material—Brief review of Sections 1.1 and 1.2); Chapter 2 (Speech science topics—Rely heavily on student reading of qualitative material and focus on issues necessary to engineering modeling); Chapter 3 (Modeling—The main goal is the digital model. Highlight the mathematics of the acoustic tube theory and stress the physical significance without extensive in-class formal development of the results); Chapter 4 (Short-term processing—This is often the students' first real exposure to short-time processing. Cover the basics carefully and have the student use the computer.); Chapter 5 (Linear prediction—Cover this central topic thoroughly except for some of the details of the solution methods in Section 5.3.3 that the instructor may choose to omit.); Chapter 6 (Cepstral analysis—The instructor may choose to omit Section 6.3 as explained in the reading notes at the beginning); Chapters 10, 11, and 12 (Recognition basics—Many details will need to be omitted, for example, in Sections 12.2.4–12.2.8); Chapters 13 and 14 (Language modeling and neural network approaches—These topics can be covered only superficially as time permits). Alternatively, the instructor may choose to include Chapters 7, 8, and 9 of Part IV on coding and enhancement, rather than recognition, at the end of the course. These three chapters could be covered in some detail. If time and resources permit, an ideal approach is to thoroughly cover material in Parts I, II, and III in an introductory course, and Parts IV and V in an advanced applications course.

Obtaining Speech Data for End-of-Chapter Problems. It will be noted that many of the problems in this book require real speech files. Although most universities have the facilities to create such files, we offer the following options to instructors who wish to avoid the details of col-
lected speech files on-site. Several standard databases have been compiled by the U.S. National Institute of Standards and Technology and are available on compact disk from the U.S. Department of Commerce, National Technical Information Service (NTIS). These databases are described in Section 13.8 of the book. For ordering information, call the NTIS at (703) 487-4650, or fax (703) 321-8547. Alternatively, the authors of this book have made available some speech samples that can be downloaded over the electronic mail network using the instructions below.

Speech files can be downloaded to a personal computer (PC) through the WWW internet computer network.¹

Acknowledgments. We appreciate the support of many people who have contributed to the writing and production of this textbook. We are indebted to a number of colleagues and graduate students who provided critical reviews and suggestions that enhanced and improved the presentation. Among those are Professor Mark Clements, Georgia Tech; Professor Jerry Gibson, University of Texas at Austin; Professor Paul Milenkovic, University of Wisconsin; Professor Larry Paarmann, Wichita State University; Dr. Joseph Picone, Texas Instruments Corporation; Dr. Dale Veeneman, GTE Laboratories, Waltham; and Professor Greg Wakefield, The University of Michigan. We also thank the members of the EE 801 class of Spring Quarter 1991 at Michigan State University, who struggled through the early versions of the problem sets and found (the hard way) the stumbling blocks. We also appreciate the diligent efforts of Mr. Sudhir Kandula in assisting with graphics and problem solutions. We also thank the students of the Robust Speech Processing Laboratory (T. W. Pai, S. Nandkumar, S. Bou-Ghazale, and L. Arslan) for their graphics and proofreading assistance. We also greatly appreciate the encouragement and support of Editor John Griffin of the Macmillan Publishing Company whose patience and good sense of humor were extraordinary. We also wish to thank Senior Production Supervisor Elaine Wetterau of Macmillan for many painstaking hours with the manuscript that contributed greatly to the quality of the finished work.

J. D. would also like to acknowledge research support during the writing of this book from the Whitaker Foundation, the National Science Foundation, and the Office of Naval Research. The research supported by these sponsors enriched the author’s understanding of many important topics.

                      J.R.D.
                      J.G.P.
                      J.H.L.H.

¹See Preface to the IEEE Edition for information on speech data, (pg. xvii).