Book Reviews

Computational Models of American Speech

M. Margaret Withgott and Francine R. Chen
(Interval Research Corporation and Xerox Palo Alto Research Center)

Stanford: Center for the Study of
Language and Information (CSLI
Lecture Notes 32), 1993, viii + 143 pp.
Distributed by the University of
Chicago Press.
Hardbound ISBN 0-937073-97-0, $39.95;
Paperbound, ISBN 0-937073-98-9, $15.95

Reviewed by
Judith Markowitz
J. Markowitz, Consultants

This small, well-written book tackles one of the great challenges facing automatic speech recognition systems: handling "variation in pronunciation, associated with the recognition of ordinary speech" (p. vii). The authors present the results of their work on variation in American English and, more importantly, describe a methodology for future work in that field.

In an effort to "combine desirable properties from each approach" (p. vii and p. 2), the authors create an interesting blend of linguistic, computer science, and statistical techniques overlaid with good empirical scientific method. The result is a dictionary of phonological baseforms and computer models designed to demonstrate

> that it is possible to model differences in pronunciation using a good set of contextual descriptors without relying on hand-created rule ordering schemes. (p. 2)

1. Outline of the Book

Most of the chapters are introduced by a citation designed to orient the reader to the topics in the chapter and at or near the start of each chapter, the authors state its purpose(s). Most chapters conclude with chapter summaries.

Chapters 1 and 2 set the stage for the detailed description of the authors' work presented in Chapters 3 through 5. Chapter 1 includes literature reviews for rule ordering, contextual variation, and approaches to the selection of units to represent variation. Chapter 2 describes traditional linguistic methodology and examines problems related to reliance on hand-ordering of rules. The human informants of linguistic research are replaced with an on-line lexical database, TIMIT, which is described in terms of its usefulness for research on phonetic variation.

Chapters 3 through 5 describe the authors' computational models in detail, accompanied by numerous examples. Chapter 3 focuses on rule creation, ordering, and interaction; Chapter 4 describes vocabulary coverage and their handling of confusable words. (Detailed data are presented in appendices found at the end of the book.)

Chapter 5 describes how they integrate statistical information and computational tools, such as decision trees, with the other components of their system.

In contrast to much of the current work being done in automatic speech recognition, the unit of representation in the authors' system is the phoneme rather than the triphone. Their dictionary "baseform" representations are "coarse" phonetic representations that are converted to pronunciation variants through the application of unordered rules. The contextual descriptors required to model phonetic variation are divided into two classes: sound structure and composition (e.g., location of the phonetic element in the linguistic structure); and lexical category (e.g., function/content word). Variants are derived from baseforms using recursive application of a single set of unordered rules implemented as a finite-state grammar. Those rules contain contextual, derivational, process control, and probabilistic information. The derived forms are organized into a probabilistic pronunciation network.

In Chapter 6 the authors highlight the major achievements of the work described in this book:

1. it proves that transcribed data of phonetic variation can be modeled without using hand-ordered rules; and
2. it contains a methodology for "transferring theoretical advances to the domain of engineering practice" (p. 111).

2. Usefulness for Computational Linguistics

The self-described accomplishments listed above are not off the mark. The authors do an excellent job of demonstrating how intelligent combination of linguistic, computational, and statistical tools can advance speech processing technology. The descriptive elements (e.g., phones, contextual descriptors) come largely from linguistics, and the procedural components are primarily computational (e.g., decision trees, finite-state grammars) and statistical.

Their handling of rule ordering represents one of the most interesting examples of this melding of approaches. The authors begin their work by using empirical testing to assess the results of all possible orderings of the rules extracted from the TIMIT data. Finding that

> rule environments and memory-efficient algorithms count for more than rule ordering schemes, (p. 39)

the authors elect to encode obligatory rule information directly into their dictionary baseforms and optional rules into a finite-state machine. Rules are applied using a recursive process called "machine iteration," which entails feeding the output of the finite-state machine back into itself until no new forms are produced. Machine iteration is constrained by augmenting the rules with additional contextual information, restrictions on the number of times some rules can apply, and probabilities. Finally, experiments were designed to test the resulting system.

The writing is clear, well-organized, and pedagogically motivated. The reader is shown the reasoning underlying the authors' theoretical and implementation decisions. System descriptions are accompanied by a large number of examples.

The authors accurately claim that their audience includes "speech engineers, linguists, and anyone who wishes to understand symbolic systems and communication" (p. vii). Technical linguistic terminology is clearly defined for nonlinguists, and computational tools, such as on-line lexical databases and decision trees, are described

in detail. Their explanation of linguistic methodology and the descriptions of their own methodology foster better cross-disciplinary understanding. As a result, this book would be useful as a supplementary text for graduate-level courses in phonetics, computational linguistics, and speech processing.

Some of the findings described in the book refer to isolated word recognition, and it is not clear that they can automatically be applied to continuous speech despite the authors' stated concern with "fluent speech." Even so, this work is an interesting addition to the literature of computational linguistics.

Judith Markowitz is president of J. Markowitz, Consultants, a software consulting company specializing in speech recognition, information retrieval, and knowledge-based systems. Her address is: J. Markowitz, Consultants, Technology Innovation Center at Northwestern University, 1840 North Oak, Evanston, IL 60201; e-mail: markwitz@steve.iit.edu.