

# Briefly Noted

## Using Speech Recognition

Judith A. Markowitz

Upper Saddle River, NJ: Prentice Hall PTR, 1996, xxii+292 pp; paperbound, ISBN 0-13-186321-5, \$34.00

*Using Speech Recognition* is written for people from various professional backgrounds who have recently come into contact with speech recognition and possibly need to make a decision as to how it can most profitably be applied in their field. Accordingly, the book makes an effort to be easily readable and tries to avoid technical detail while still informing the reader about the crucial issues in speech recognition systems.

The largest part of the book is taken up by the chapters: "What is a Speech Recognition System?," "Representing the Vocabulary," "Adding Structure," "Speaker Modeling," "The Flow of Speech," and "The Speaking Environment." Each of these is divided into two parts, called "Technology Focus" and "Application Focus." The technology part introduces the options that researchers have pursued, while the application focus discusses what choices should be made for a given application and what approaches have resulted in commercial products. For example, the technology focus of Chapter 4, "Adding Structure," states that there are essentially three ways to restrict the number of legal word sequences that a recognition system will accept: finite-state grammars, statistical  $n$ -gram models, and linguistic grammars. The application focus then proceeds to clarify that finite-state grammars are the correct choice for small-vocabulary command-and-control tasks, while  $n$ -grams are used in most large-vocabulary systems. It is also stated that linguistic grammars have not found a way into the marketplace as a part of speech recognition systems, and that true language understanding systems that allow natural discourse between humans and computers will not be commercially available in the foreseeable future.

While all this sounds fairly technical, the author is prepared to avoid mathematics at any price. As far as I have noticed, there is not a single formula even of the simplest kind in the text. At the same time, there is indeed a section called "Transfer Functions in Neural Networks." I think it is a

waste of time to talk about nodes in a network and how information is propagated between them if you are not in a situation to give some mathematical facts. In my opinion, large parts of the technology focus suffer from this defect. Readers will not really be able to understand what, for example, "transfer functions" are. They will just accumulate a useful set of catchwords and phrases that will come in handy in the next meeting.

The strong point of the book, and the reason why I would recommend it to application developers and decision makers, is its up-to-date survey of commercial applications of speech recognition. You will find a clear taxonomy of applications (data entry, command and control, dictation, etc.) in the chapter entitled "Application Types" and numerous examples of successful commercial systems in the chapter "Applications at Work." Somewhere in the book, you will find the name of every important company in the market and what their products are. Also, if the technical sections are too vague for your tastes, there is always a reference to a fundamental scientific paper. The bibliography is excellent and contains most of the classic papers in the field.

Most importantly, the book is easy reading. If you always wanted to know something about speech recognition and you don't have the time or nerve to fight your way through a lot of scientific papers, read this book on the plane while on a business trip.—*Andreas Hauenstein, pc-plus GmbH, Munich*