[Published in *Journal of the International Phonetic Society.* 20.2.52-54. (1990).]

Review of Douglas O'Shaughnessy
Speech Communication: Human and Machine
Reading, Massachusetts: Addison-Wesley Publishing Company
1987
ISBN 0-201-16520-1

In addition to an *Introduction* and a *Review of Mathematics for Speech Processing* (useful to those who need a refresher, but probably too compressed to serve as an introduction) this book contains nine chapters:

3. Speech Production and Acoustic Phonetics
4. Hearing
5. Speech Perception
6. Speech Analysis
7. Coding of Speech Signals
8. Linear Predictive Coding
9. Speech Synthesis
10. Speech Recognition
11. Speaker Recognition

Each chapter is accompanied by a set of problems, solutions to some of which are provided at the end of the book.

As can be seen from the chapter titles, the book begins with a three chapter introduction to phonetics, followed by a chapter on the basic signal processing techniques for acoustic analysis. The remaining five chapters are devoted to various engineering topics. As this organization suggests, the book is aimed basically at engineers. The detailed survey of coding techniques in Chapter 7 in particular is long, technical, and unlikely to be of more general interest.

The chapters on speech recognition, synthesis, and so forth provide good overviews of these areas, which with the omission of some of the more technical material will be useful to non-engineers interested in a glimpse at these areas.

It is chapters 3-6 that will be of greatest interest to general phoneticians. These chapters by themselves are not sufficient for a course in phonetics, but they may profitably be used for the topics with which they deal. Except for its emphasis on English, Chapter 3 is typical of many introductory phonetics texts but provides more material on the acoustic theory of speech production than is typical in books on general phonetics. Chapter 4 provides an excellant review of hearing, including both the anatomy and physiology of the ear and an introduction to auditory psychophysics. This chapter especially would make a useful supplement to most introductory phonetics textbooks. Chapter 5 covers the usual topics in speech perception, with emphasis on the acoustic cues for linguistically relevant distinctions

and comparatively little attention to overall models of speech perception. Chapter 6 reviews the basic methods of acoustic analysis, including some, such as cepstral analysis, used in engineering that are less well known to other phoneticians.

Two aspects of the book that struck me were its anglocentrism and its strong emphasis on the lower-level signal-processing aspects of speech communication. Since work on speech recognition, synthesis, and coding has concentrated heavily on English, it is understandable that the book should largely be concerned with that language, but it is occasionally disconcerting to read such unqualified general statements as " ... there are about 1000 diphones (combinations of 32 x 32 phonemes) ... " (p.313), which is evidently restricted to English.

Attention to languages other than English is limited, and not always accurate. For example, OS cites Japanese as a language without consonant clusters (p. 59), which is not true. Even if affricates and *Cy* sequences are treated as single segments, there remain NC clusters, as in *onseigaku* "phonetics", and clusters of voiceless consonants occur phonetically as a result of deletion of intervening high vowels.

I find equally dubious his statement (p. 62) that "Languages usually differ most significantly in the set of vowels used." It is difficult to know how to quantify the extent to which languages differ in phonetic inventory, but I rather suspect that OS's impression is an artifact of the languages with which he is most familiar. The great number of known consonant sounds together with the widely varying sizes of consonantal inventories, from 7 for Hawaiian to more than 80 for some Caucasian languages, make this statement highly questionable.

In a similar vein, OS's statement (p.62) that "Except for trills and ingressive sounds ... English provides good examples of sounds used in various languages." gives the misleading impression that the phonetic inventory of English covers most known speech sounds. In addition to trills and ingressives, English lacks glottal ejectives, rounded front vowels, retroflexes, uvulars, bilabial fricatives, pharyngeal fricatives, nasal fricatives, and pharyngealization, among others. If we consider distinctive oppositions we may add to this list still other categories, such as aspiration, nasalization, and voiceless sonorants.

The emphasis on the lower-level aspects of speech communication manifests itself in the limited attention given to such topics as the integration of speech recognition into more general natural language understanding systems or the higher-levels of text-to-speech systems. I suspect that it is also responsible for what seems to me to be a curious approach to the evaluation of the quality of speech coding techniques. OS insists on a distinction between time domain coding methods, for which he considers Signal to Noise Ratio (SNR) appropriate, and spectral methods, for which "subjective" methods, i.e. those that directly measure intelligibility and naturalness, must be used. Even so he is forced in at least two cases (ATC p.293, ADPCM-HS p.298) to concede the inappropriateness of SNR as a quality measure even for time domain methods. It seems obvious to this reviewer that what OS refers to as "subjective" measures (surely a misnomer in the case of intelligibility measures), are by the nature of the problem the most appropriate in all cases since the purpose of a communication system is to maximize intelligibility and naturalness, and that SNR is a sometimes convenient but inferior quality measure.

Otherwise I found few real problems in the book. A small point that may confuse some readers is OS's definition of metric (p.425), which he defines as a function that is: (a) commutative ($d(a,b) = d(b,a)$); (b) satisfies the Triangle Inequality ($d(a,c) \leq d(a,b) + d(b,c)$); and (c) transitive. I am at a loss to explain what he may mean by "transitive", since this is a property of binary relations whereas a metric is a function of two variables, which is to say a ternary relation. The usual third defining property of a metric is self-adjacency ($d(a,a) = 0$).

The greatest defect of the book is the unfortunate system of citation. References are cited by number, which I find harder to keep in mind than the usual name and date. But the real problem is the fact that the references are grouped separately by chapter at the end of the book, and within the list for each chapter are listed in order of first mention. This order makes the bibliography virtually useless for looking up an incomplete reference, a task for which this book might have been very useful, and the grouping of references by chapter with no headers indicating for which pages of the book the references are to be found on that page makes looking up a reference a tedious task. A unified alphabetically-organized bibliography would have been much easier to use.

The book with which this work invites comparison is the classic but now rather dated Flanagan (1972). Although the overlap in material is not complete, the two books serve very much the same needs, and *Speech Communication* may be expected to replace Flanagan (1972) as the most important textbook and reference in this area.

References

Flanagan, James L. (1972) *Speech Analysis, Synthesis, and Perception.* Berlin: Springer-Verlag.

William J. Poser
Department of Linguistics
Stanford University
Stanford, CA 94305-2150
USA