Visual Representations of Speech Signals, edited by Malcolm Cooke, Steve Beet and Malcolm Crawford. Chicester: Wiley, 1993. 385 pages. $83.95 ISBN 0 471 93537 9

Language is one of the evolutionary hallmarks of *Homo sapiens.* Virtually unique among human behaviors, there is no truly corresponding animal model for studying its physiological and anatomical bases. Until recently, most of our knowledge pertaining to its neurological foundations stemmed from the clinical literature, fraught with ambiguity and difficult to interpret. The advent of sophisticated computer-based visualization techniques (such as MRI and PET) are beginning to yield impressive results, but to date have provided information pertaining principally to the anatomical localization of cortical activity associated with linguistic processing. Little is known about the physiological processes underlying the perception of speech and this sad state of affairs is likely to continue until the development of non-invasive techniques with sufficient spatial and temporal resolution to measure the activity of small neural populations.

However, the situation is not completely hopeless. For the past decade and a half a small coterie of researchers have painstakingly recorded the activity of single neurons from the mammalian auditory pathway (mostly from cat) to speech and speech-like signals under the assumption that general processing strategies will be similar across species [e.g. *Journal of Phonetics*, 16: 1-151]. By necessity, experimental investigations have focused on the pre-phonetic stage of speech processing, pertaining to spectral computation and temporal segmentation. This research has recently become of interest to computer scientists and psychologists investigating language processing in the expectation that understanding the physiology of speech processing will provide important insights into the neural mechanisms underlying language understanding.

Besides the general interest in speech understanding a pragmatic attitude drives the interest in auditory processes. Computer-based speech recognition systems perform poorly when the acoustic environment changes. Speech recognition scores of ninety percent or better plummet when the speech is embedded in a realistic acoustic environment containing such background interference as machine noise, door slams, laughter or speech. Humans generally experience little difficulty understanding speech under such conditions, so the question naturally arises as to how the auditory system deals with such adverse acoustic circumstances.

The present volume brings together the work of more than 80 researchers (most of whom are computer scientists or electrical engineers) from Europe and North America presented at a conference sponsored by the European Speech Communication Association in April, 1992. The meeting sought to compare the various analytical and visualization techniques used to process and display speech. In order to make the results as comparable as possible, participants were asked to focus their presentations on four sentences provided by the organizers. These speech materials were designed to be representative of "real-world" conditions. Different speakers, of separate gender, were used and the speech, in two instances, was recorded in very noisy environments (an appendix provides details concerning the speech materials and recording conditions).

The book is divided into two parts. The first contains five tutorial length chapters, each pertaining to an important topic in speech research. Subjects covered include pitch extraction, neural networks, time-frequency analysis, acoustic-articulatory interactions and auditory models. The second, far lengthier, section consists of brief papers summarizing the thirty-six conference presentations.

Why should the readership of *The Physiologist* take note of a volume squarely aimed at an audience immersed in the mathematical technicalities of speech analysis? The answer lies in the manner in which many of the participants applied physiological data to develop higher-order representations of the speech signal. "Visual Representations of Speech Signals" amply demonstrates the power and utility of the computational approach applied to the study of a complex human behavior.

Nearly half of the studies utilized some form of auditory modeling. Traditionally, the sound spectrograph has been the analysis tool of choice among speech researchers. It displays time running along the abscissa and frequency on the ordinate, with a third dimension, energy, represented as the density of the display as a function of the other two dimensions. Most of the auditory representations described in this book adopt a similar convention, where frequency is replaced by cochlear location (frequency channel). In most models the temporal information contained in the auditory nerve firing distributions is converted into a "place" represention at some level of auditory analysis (e.g. the spiral periodicity detector of Patterson and colleagues). Many of the models assume that the frequency analysis of auditory-nerve spike patterns is accomplished through some form of autocorrelation.

However, the most interesting portion of the book concerns the strategies used to integrate the fine detail of neural firing patterns into representational forms commensurate with higher-order (i. e., perceptual) analysis of these sounds. For example, Patterson and colleagues look for global patterns corresponding to acoustic "images." The Sheffield group (Cooke, Beet, and Crawford) emphasize "synchrony strands" through which the auditory system constructs maps of related temporal activity through time as a means of grouping neural activity related to a coherent sound source. Their goal is to construct rules for mapping neural events onto the acoustic "objects" that produced them, similar in approach in vision pioneered by James Gibson and David Marr. Because so little is known about the physiology of speech processing in the upper reaches of the auditory pathway, these papers can serve as a valuable point of departure for those interested in studying the mechanisms by which the peripheral representation of the speech signal is transformed into higher-order, more abstract constructs. For this reason many of the papers in this volume will be extremely useful to auditory researchers working with speech and other complex sounds.

The tutorial chapter of greatest interest to physiologists is "On the importance of time - a temporal representation of sound," by Slaney and Lyon. They provide a clear description of their non-linear cochlear model, which produces spikes trains distributed across the tonotopically organized auditory nerve. At some more central, anatomiclly unspecified location, these spike patterns are processed by an autocorrelator, which serves to extract information pertaining to both pitch and speech timbre. Hermes' tutorial on "Pitch analysis" will also be of interest to auditory researchers. Although focused on computational issues, the chapter discusses some of the more recent auditory research on pitch. Yet a third tutorial of interest describes recent innovations in signal processing techniques that circumvent the time-frequency trade-off characteristic of traditional spectral-analytic approaches (Loughlin and colleagues).

Of the thirty six papers in the second part of the book, fifteen pertain in some fashion to how the auditory system processes speech signals. Of special note are the papers from the conference organizers (Cooke, Crawford and Beet) and the MRC Hearing group (Allerhand, Patterson and Holdsworth) describing their recent work on auditory scene analysis of speech. Their research builds upon experiments performed by various researchers in North America (e.g., Albert Bregman) and England (e.g. Chis Darwin and Quentin Summerfield) delineating the perceptual grouping strategies employed by the auditory system to construct a coherent image of the acoustic world.

This volume will prove a valuable reference for those seeking to apply computational methods to their experimental and theoretical studies of speech processing in the auditory pathway.

Steven Greenberg
University of California, Berkeley