Speech technology, in one form or another, has been available commercially since the mid-1970s, and still has little mass market penetration, despite considerable improvements in performance in recent years. Reading these article is soon becomes clear why: there is an enormous mismatch between the way we are accustomed to using speech as a free communication medium between people who understand each other and the way we have to use speech to communicate with a machine.

Two generic problems are the necessity of providing feedback to overcome the errors made by a recognizer and the fact that HCI has evolved to make use of other, conventional technology, so that existing interfaces don't work well when adapted for voice. Anyone who has used a speech-driven telephone database enquiry service will appreciate the tedium of menu-filling this way. In addition, speech is a very poor device for pointing, inferior to a mouse. The advantages frequently claimed for 'hands-and-eyes-free' interfaces often turn out to be less than crucial in practice.

Of necessity, many of the studies reported in Baber and Noyes were performed using equipment which is now obsolete: some of the difficulties uncovered may be of only temporary importance. There is considerable funding committed over the next 10 years, particularly in the USA, Japan and Germany, to the development of 'Spoken Language Systems', for instance for airline reservations and for 'Interpreting Telephony' (speak in English and your Japanese colleague hears you in his native language). It remains to be seen whether this research, with the benefit of better recognition techniques and attempts to 'understand' the interaction, at least in limited domains, are able to produce habitable speech-driven interfaces.

P. GREEN
*University of Sheffield*

MARTIN COOKE
*Modelling Auditory Processing and Organization.* Cambridge University Press, 1993, 122 pp, hardbound, £22.95, ISBN 0-521-45094-2

Published by the Cambridge University Press in their 'distinguished dissertations in computer science' series, this book deals with a model of the peripheral auditory system.

Being a dissertation, it presents a model for auditory processing developed by the author and is not, as the title might suggest, a treatise on modelling of auditory processing in general. It concerns itself with one—albeit important—aspect of auditory perception: integration of the component parts of the complex auditory signal that the ear receives into segregated structures that can be attributed to specific sources. While the title suggests organization, the model developed concerns itself with segregation.

The human auditory system, while listening to complex signals such as speech, perceives the stimulus as an integrated whole and not in terms of its individual component parts. If there are several sound sources, it is able to attribute the component elements to the respective sources, providing the listener with a proper perception of the auditory scene. This involves decomposing the signal into its component parts, followed by segregating and reintegrating them source wise. It is known that the human listener groups the components on the basis of commonality of attributes such as spatial location of the source (binaural hearing), temporal properties (onset and offset times of individual events) or spectral attributes (e.g. being harmonics of a common fundamental). Thus, temporal and spectral attributes are both important. The author presents a computer model which seeks to accomplish this task.

Computer models can be representational (concerning themselves with abstractions at various levels), functional (aimed purely at yielding the desired performance) or structural (based on the structure of the original system). While there is a fair amount of knowledge regarding the peripheral auditory system, the underlying mechanisms are not well understood. Even less is known regarding the manner in which the auditory system identifies the properties of the acoustic signal, associates them with individual sources and forms an auditory space. Thus, the best guiding principles currently available for the modelling task are from psycho-acoustics. The approach here has therefore been to use the results of psychoacoustic experiments as guide lines for developing a functional model.

The author is thus guided by psychoacoustics and is pragmatically motivated: his stated goal is to make the organization of synchronous auditory activity in time and frequency explicit; the algorithm employed is of secondary importance.

Earlier work suggests that auditory scene analysis can be accomplished by decomposing the scene into time frequency objects which characterize onsets, offsets and movement of spectral components in time. These are perceptually grouped if they are harmonically related, start and end at the same times, share common rates of amplitude modulation and if they are proximate in time and frequency.

This processing is performed on the outputs of a model of the auditory periphery. The author accomplishes this in a two stage framework: the first step consists in the decomposition of the signal into a collection of time frequency descriptions called synchrony strands. These are captured using the property that contiguous sections of a filter bank have synchronous responses just as reasonably large sections of the basilar membrane vibrate in unison. The strands extracted from speech signals correspond to the individual harmonics (at low frequencies) and formants (at high frequencies).

In the second stage, the components are recombined to form coherent 'groups' attributable to auditory objects, using auditory principles. The strategies of auditory grouping are generally presented as heuristics.

Experimental results: The effectiveness of the strategy is evaluated in the first instance on the basis of how well an utterance can be characterized. This is possible by means of resynthesis of the original signal from the characterization accomplished by the model. Where there are intrusive sources, the criteria relate to how well the group represents all the evidence for the first source and how effectively it separates speech from the intruding signal, say other speech.

When no intrusive noise is present, the group formed by the model captures 67–79% of the strands characterizing the whole of the utterance. Generally, the high frequency components are missed. In the presence of intrusive noise (such as noise bursts, siren, telephone, other speech, etc.), 84–96% of the elements in the group pertain to the speech source (and the rest from the intruding noise). (On the other hand, when groups are assembled by picking up strands purely randomly, this figure varies between 46–88%.) Intrusion is worst where the other source is music and is least for tone burst sequences. Intrusion remains under 18% for groups formed by the model but rises upto 46% for random grouping. The system performed well in separating speech from other speech. With laboratory noise, only the first few harmonics of the pitch frequency could be grouped. The quality of speech resynthesized from the groups ranged from totally intelligible to almost incomprehensible (particularly in the presence of white noise). Segregating a rhythmic structure from a mixture left a set of rhythmic gaps which were noticeable.

To sum up, the model evolves groups which are better than randomly assembled ones, but overall, its performance falls short of promise. There is, however, much scope for improvement, since the field is as yet evolving. The author suggests several avenues for improvement.

For instance, the author uses a bottom up or data driven strategy; a schema driven strategy could be tried. Competition for elements between groups is not incorporated because no detailed perceptual investigations have been done in this area. Of all principles discussed for grouping, only two are implemented: subsumption (removing groups whose elements are contained in larger groups) both before and after fusion of groups based on pitch contours and simultaneous integration (of groups whose derived properties are similar).

It is possible to take recourse to sequential organization or relating primitives or groups across time. Better grouping strategies such as pitch contour continuity and spectral continuity could be used. The assumption regarding atomicity of auditory primitives is a limitation, since it may become necessary to share properties associated with the same object between streams. The model can be extended to allow the investigation of auditory as opposed to acoustic phonetic coding. This is still to be done.

The synchrony strands are a vehicle of convenience and do not have any basis in human speech processing. The process of forming synchrony strands assumes that there is a single dominant resonance in each frequency channel and that a summary of the dominance can be obtained from a contiguous section of the filter bank. This is an approximation. Also, synchrony strands do not make explicit every kind of acoustic source component. Whispered speech is adequately represented by strands for resynthesis but this is achieved through using many objects; a more descriptive representation may be better. Offsets and onsets are not explicitly modeled.

Because the model summarizes a single dominance in each frequency region, it will tend to represent the most dominant harmonics in any harmonic series. Thus, there would be no difficulty in determining pitch contours, even if the fundamental frequencies cross, provided the first formants are reasonably separated. Frame by frame analysis would present problems relating to pitch correspondence in this case.

The thesis seeks organization in representations derived from a model of the auditory system. The link with the auditory model, however, is quite weak; in fact, substantially the same methodology could have been evolved even without inspiration from the auditory model. The author himself states that further development of the model should reflect what is known of auditory processing.

The volume cites extensively from earlier work in related areas and is therefore a rich source of reference material. The facts, concepts and methodologies used are precisely articulated and rigorously treated. The reader is given a clear insight into the process of evolving research strategies and implementing them. He is also provided with an objective analysis of the limitations of the approach and an outline of future directions for research. The book would therefore be a good addition to a researcher's library but might disappoint a reader looking for exposure to the broad perspective of the field.

Also, a number of questions come up. Would it be easier to group objects because of similarity or remove components from a mixture because of dissimilarities? Should the grouping be schema driven rather than data driven? One might also ask: how elegant is the implementation? Does it show an innovative spark or is it merely a patchwork of motley strategies?

In all, the author raises more questions than he answers, but that is the way research moves ahead.

P. V. S. RAO
*Tata Institute of Fundamental Research, Bombay*

JOHN GORDON (editor)
*Practical Data Security*. Ashgate, 1993, £49.50, 160 pp hardbound, ISBN 1-85742-145-0

This book is a collection of papers, most of which were presented at a Unicom seminar on data security in the summer of 1992. The papers cover a variety of topics falling under the general heading of data security. However, due to the breadth of the subject and the