W. A. Ainsworth: **Mechanisms of Speech Recognition.** Oxford, Pergamon Press, 1976, pp. 139.

The book under review, Mechanisms of Speech Recognition, published in a well-known International Series in Natural Philosophy, Volume 85, is a study which will — no doubt — be highly appreciated by both research workers and language learners. Its author, W. A. Ainsworth, member of the Department of Communication, University of Keele, Staffordshire, is right in his hope that any reader who is interested in the topic of speech recognition will find the book highly stimulating. His clear thinking as well as his ability of lucidly summarizing the facts cannot fail to attain the intended purpose.

The book is divided into eleven chapters: 1. Speech Production (pp. 1—11), 2. The Auditory System (pp. 12—21), 3. Auditory Psychophysics (pp. 23—31), 4. Speech Analysis (pp. 34—56), 5. Speech Synthesis (pp. 59—67), 6. Vowel Recognition (pp. 68—73), 7. Consonant Recognition (pp. 75—87), 8. Perception of Prosodic Features (pp. 89—93), 9. Perception of Distorted Speech (pp. 96—103), 10. Automatic Speech Recognition (pp. 104—117) and 11. Models of Speech Perceptions (pp. 120—124). The very titles of the chapters and the amazing number of references (289) give ample evidence of the wide range of the author's interest as well as of the multi-disciplinary character of the study of speech.

In the first two chapters the reader finds a brief but precise description of the movements of the articulators and the resulting acoustic changes involved in the production of speech sounds. Next comes the description of parts of the ear and the auditory nerve as well as their function. Numerous schematic diagrams of parts of the ear, parts of the neurone and the main ascending auditory pathways are of great help for those who do not possess the presupposed skills in this discipline. One also gets reliable information on the methods used in order to examine the auditory system (most of the experiments being based on the hearing systems of various animals) as well as on different techniques employed by various researchers. The results obtained are certainly interesting for those whose domain lies in the theories of speech recognition, cf. e. g. the existence of a few units in the auditory cortex of squirrel monkeys which appear to respond only to specific monkey vocalizations and not to other sounds, or the fact that some units respond best to a frequently repeated stimulus. Most of them, however, habituate responding to the first presentations of the stimulus and rapidly cease to respond when the stimulus is repeated.

Chapter 3 is devoted to the data obtained from human listeners by presenting them with acoustic stimuli and recording their responses. Both methods, close-field and open-field are dealt with, the second one, especially when applied in anechoic chamber, being — in our opinion justly — preferred as more natural. Due attention is given to the questions such as the threshold of hearing, intensity discrimination, frequency thresholds, frequency discrimination, pitch and loudness. Critical bands, perception of duration, fatigue and masking are taken into consideration, all the phenomena being illustrated by aptly chosen diagrams. In the last paragraph the author gives a rough outline of the existing theories of hearing, Helmholtz's „resonance theory" and that of Rutherford taking as a starting point for all the forms of the two which are now known as the „place" and „volley" theories, with all their main advantages as well as drawbacks.

In the fourth chapter the various techniques of analysing speech sounds is discussed. These range from the simple display of the speech waveform through spectrum analysis to statistical analysis of the occurrence of the sounds of speech in language. Sonagrams of the English vowels are included to show not only the differences in the vowels themselves but also the different formant frequencies between the vowels produced by the female and male speaker. Using the measurements of vowels performed by Peterson and Lehiste, Ainsworth is right in his comment that the preceding consonant has hardly any effect on the duration of the vowel (or dipthong) while the effect of the following consonant is considerable. A question arises, however, why at this place Ainsworth speaks of the American English vowels which differ both in quality and quantity from those of British English. Because of its openness the phoneme (ae) is certainly longer than any other vowel in the same surrounding, yet its classing among „long" vowels is rather disputable. The reader is also misled by the bibliographical data (1960 in the text, 1959 in the head of the tables 4.1, 4.2 and 4.3 (pp. 42—43). As for the consonants, due attention is devoted to their classification, both from the phonetic and phonemic point of view, as well as thier mutual combinations and distribution in English.

In chapter 5 the reader is acquainted with the history of a speech synthetizer, the oldest being probably the Kratzenstein's set of acoustic resonators. Of the later, more elaborate machines, that of Kempelen, Wheatstone and Bell are reported on. Ainsworth, however, correctly points out that successful talking machines could hardly have been produced until the development of electronics, the Voder being probably one of the first electrical synthetizers. Another one is the

Vocoder. Its principle, the synthesis of speechlike sounds from a pattern of slowly varying signals, has become to be of great utility in the study of speech preception; it has enabled speech-like stimuli to be generated and the effects of manipulating features of these stimuli to be listened to. Thus the features of speech sounds which are important for the perception have been deduced. One of the earliest devices used in the study of speech preception was „pattern playback", built at the Haskins Laboratories. The sounds produced by this device have a constant pitch of 120 Hz, but in other respects they do resemble speech sounds and are fairly intelligible. Formant synthetizers, such as PAT and OVE II are the next described, followed by the Digital synthetizers. They have the advantage of being less prevalent to drift than their analogous counterparts. LPC (Linear Predictive Coding) has proved to be a highly efficient bandwidth compression system. Given a good fundamental frequency tracking device and a sufficient number of predictor coefficients, good-quality speech is generated.

Vowel and consonant recognition, richly documented by various methods, is the subject of chapters 6 and 7. In order to determine whether the hemispheric specialization (i.e. left hemisphere for the speech-like sounds, right hemisphere for other noises) applies to low-level processing, such as phoneme recognition, or only at higher levels of processing, Ainsworth presents many interesting experiments where the listeners are asked to identify the sounds presented to each ear, REA (right ear advantage) being one of today's generally known, discovery.

Perception of prosodic features, such as length, stress, rhythm and intonation is dealt with in chapter 8. Of the particulars, let us mention here at least the fact that durations are generally perceived more accurately for speech sounds (when compared with various non-speech sounds), or the discovery that although intensity and length both contribute to the perception of stress, length is the more important factor. As for rhythm and tempo, the comb model is — as shown by Kozhevnikov and Christovitch — the best model for Russian speech and this may — in Ainsworth's opinion — well apply to other languages. As shown in various theories (e.g. that of Halliday and Pike), the mechanisms employed for perceiving the pitch of tones and complex sounds may be used in the perception of intonation contours. Now that synthesis-by-rule systems are available, Ainsworth expresses his hope that research in this area will develop rapidly.

The author is aware of the fact that a number of operations ocurring in natural circumstances or introduced deliberately, distort the speech wave and thereby affect the recognition of the message. Some of the effects of distortions, such as frequency distortions, aplitude distortions, time distortions, masking of speech, effect of context, cocktail-party effect and verbal transformation effect on the perception of speech are considered in chapter 9.

In chapter 10 the reader finds opportunity to acquaint himself with speech recognizers starting from those the purpose of which was to recognize spoken digits, to more general purpose recognizers which would recognize phonemes rather than words. Various forms of pre-processing are dealt with and major problems of methods applied are shown. In recognition of continuous speech these problems are multiplied. It seems evident that if machines are to approach human performance, at least some linguistic expertise must be built into them. A number of projects in this respect has been developed. Ainsworth mentions the following: hierarchical systems (applying the sources of knowledge in series in a fixed order), top-down systems (functioning according to the „hypothesize-and-test" paradigm) and the heterarchical systems (represented e.g. by HEARSAY; in this structure, each source of linguistic knowledge is modelled as a self-contained „procedure" with three functions: the procedure decides when it has something useful to contribute, it makes contributions by originating hypotheses, and it tests the hypotheses made by the other procedures).

The last chapter offers information on models of speech perception; short-term auditory memory, location of speech detector mechanisms, motor and auditory theory of speech perceptio and models from automatic speech recognition are considered in this connection, all fors and againsts aptly and persuasively argued.

No matter how highly informative the book under review undoubtedly is (and only the lack of space prevents us from going into further details) the methods so far employed to unravel the mysteries of speech-recognition mechaniam are, as underlined by the author himself in Conclusions, by no means exhausted. Techniques from experimental psychology, adaptation and masking, are only just beginning to be applied with speech-like stimuli. Advances are continnally being made in neurophysiology which have implications for speech recognition. Developments in electronics and the availability of computers is helping to make the performing of complex experiments more feasible. And, last but not least, linguistics is being expanded so that the importance of the spoken as well as written word is being fully recognized.

Jaroslava Pačesová