Research Possibilities in Phonetics

Peter LADEFOGED*

The start of a distinguished new research instutute in one's own field is a time not only for wishing them well, but also for trying to define what the field is, and how one's own research fits into it. Many people have suggested that there is a field which may be called Communications. If there is, it is certainly of enormous scope, and probably far too large to be a fit object of study within a single research institute. We may try to limit the field a little by referring to Human Communication, thus indicating that although there may be interesting things to be said about the herd calls of porpoises, and exchanges of information among birds and among bees, these are not our central concern. It is probably also a good idea to restrict ourselves to communication by speech; the study of writing, gesture, painting, music and other forms of human interaction are obviously enormously complex in themselves. But are we then left with a sufficiently small, well-defined subject? For many people the answer is still no, and it has been suggested that we should restrict ourselves to the study of the Communication Sciences. The meaning of this phrase is often a little vague; but the intention is to limit the field of study to those aspects of human communication which can be investigated by making experimental observations. Perhaps it would be better to say that we are interested in constructing scientific theories concerning speech. If we take this to be an appropriate interpretation of the phrase Communication Sciences, then we are left with a field which overlaps to a very large extent with linguistics, with the new discipline of psycholinguistics, and with what has traditionally been called speech science (or phonetics) in departments of speech, as well as being concerned with parts of the field of physiology, the techniques of engineering, and the overall viewpoint of the science of cybernetics. This definition of Communication Sciences provides a reasonably accurate description of the field of research of the group associated withe the UCLA Phonetics Laboratory. Most of us are associated with the Department of Linguistics and/or the Department of Speech; and some of us are working in psycholinguistics, either as part of the English as a Second Language group, or in conjunction with the Department of Psychology. Some members of the Laboratory are also electrical engineers; and we have a great deal of interaction with members of the Department of Physiology, with whom we teach a joint course in " Cybernetics of Human Communication."

It is difficult to sum up our current research in a few words. The best that can be done is to discuss some of the general questions which are underlying the particular problems which we are investigating. We would like to know what

* University of California, Los Angeles, California, U.S.A.

kinds of information are conveyed by speech, and what are the physiological and acoustic correlates of the various information bearing aspects of speech. Putting this another way, we may say that we are concerned with how the information conveyed by speech is represented within the brain, how this stored information is encoded so that speech may be produced, and how speech sounds are interpreted by a listener as representations of the information encoded by the speaker.

Much of our research is concerned with linguistic aspects of phonetics. Whenever a linguist describes a language, at some point he has to talk about the sounds; and when he does this he normally uses phonetic categories, such as voiced, alveolar, and fricative (or perhaps acute and strident). The aim of this part of our research is to formalize a possible set of categories for describing the languages of the world, and to test the adequacy of these categories by reference to data from several hundred languages. We also hope to try out some new methods of testing linguistic phonetic descriptions with the aid of a computer controlled speech synthesizer. Our preliminary thinking indicates that this involves some interesting considerations concerning the nature of the bundles of simultaneous categories which specify segments, and the form of the conjoining rules necessary for turning a sequence of segments into continuous speech.

Another large section of our research is conerned with physiological aspects of phonetics. In our earlier work we concentrated on investigations of subglottal activity; but for the last three or four years we have been studying other parts of the vocal apparatus, including the vocal cords, the tongue and the lips. We have conducted electromyographic investigations in all these areas, although so far only our work on the lips has led to publishable results. We have been especially concerned with the relationship between linguistic units and muscular activity. At present we are unable to find a simple correspondence between phonemes and motor commands; there must be different gestures for producing /b/ in initial and final position. Nor can distinctive features such as Grave or Labial be given a simple interpretation; the gesture for final /p/ must be different from that for final /b/. We are currently engaged in experiments attempting to show whether (1) at the level of the motor commands, the minimal linguistic unit involved in the production of speech is larger than the phoneme, perhaps more the size of a syllable; or (2) motor commands correlated with units the size of phonemes or distinctive features are used in speech production but are altered, i.e., context restricted, either by feedback information concerning the existing state of the muscle, or by stored information in the short term memory.

Most of our data on states of the glottis has been obtained through laryngoscopic photography and photoelectric glottography. In the latter we use a photoelectric cell in the lower part of the pharynx and a light source shining on the neck below the larynx. The output of the photo-cell is proportional to the glottal aperture; and as a first approximation we may regard the AC variations which we record as being proportional to the variations in sound pressure at the level of the glottis. Analyzing this " glottal wave " as tape recorded during running speech, we have

found that its spectrum changes during changes in the pressure differential across the glottis, e.g., during voiced stops. We are also using this system to investigate different phonation types such as creaky voice (or vocal fry). It seems that in sounds of this kind the number of glottal openings is greater than or equal to the number of glottal pulses sharp enough to excite the vocal tract. Glottal openings in creaky voice typically occur in groups of two or three, but only the last of each group invariably has sufficient energy to excite the vocal tract.

The third principal area of our research involves paralinguistic studies of speech. Much of our work here is concerned with the study of voice quality, and consequently overlaps with the physiological investigations of the state of the glottis reported above. We are also studying the use of intonation and variations in phrasal stress (or accent) in conveying information. We are trying to develop rules for automatic accent and intonation pattern assignment which would take into account not only the speaker's attitude and emotional state, but also relations often not considered in linguistic description, such as those of anaphora, contrast, thematization, discourse equivalence, and other connections above the level of the sentence.

This third area of research leads us into consideration of the problem of identifying speakers. During the last few months we have become increasingly involved in legal cases in which tape-recorded voices have been identified by means of the so-called " voiceprint " system. We are extremely worried about the claims made by proponents of this technique. Our own experiments indicate that it might sometimes be possible to say, beyond a reasonable doubt (to use the legal phrase), that two tape recordings were made by *different* speakers, but that it is impossible, using the " voiceprint " technique of matching spectrograms, to say (beyond a reasonable doubt) that two recordings were made by the *same* speaker. We have found many cases when spectrograms of different speakers look more alike than spectrograms of the same speaker on different occasions.

The UCLA Phonetics Laboratory is now well equipped for research of the kind that has been described. Undoubtedly, the heart of the laboratory is the LINC-8 computer. This is a general purpose computer with an 8,000 word core memory, 16 channel (multiplexed) A to D conversion, 24 channel D to A conversion, oscilloscope display, dual magnetic tape unit, and a teletypewriter with paper tape reader and punch. The computer has been installed only recently. At the moment we have running (in addition to all the general purpose programs included with the computer software) only our programs for processing electromyographic data. Soon we hope to be processing glottographic and other physiological data, and to be using the computer for controlling speech synthesis systems.

We are experimenting with various forms of speech synthesizers:

(1) The Terminal Analog Speech Synthesizer (TASS) allows us to produce speech-like sounds in which the parameters are the frequencies and amplitudes of the first three formants, the fundamental frequency, and the center frequency

and amplitude of fricative noise. Until the advent of the computer, this synthesizer has been controlled by a function generator in which the control voltages are taken from lines representing the parameters drawn in conducting ink on a plastic sheet which passes under a resistive roller. (2) The Mechano-Acoustic Speech Synthesis system, (MASS), comprises a replica of the vocal tract driven by a glottal analog. The tract shape cannot yet be controlled dynamically, but the system is useful for studies involving the correlation of different shapes of the vocal tract with steady state sounds. (3) The Line Analog Speech Synthesizer (LASS) is nearly completed. This is an electronic simulation of the vocal tract which can be computer controlled. There are 18 dynamically variable circuits representing sections of the vocal tract and there will be a shunt section representing the nasal passages. The fundamental frequency and amplitude of the pulse input, and the center frequency and amplitude of the noise source are also controllable as in the terminal analog speech synthesizer.

Finally, in this discussion of the work of the UCLA Phonetics Laboratory, mention must be made of the teaching commitment. A major aim of the Laboratory is to train people to work in the field. As a result not only does the instrumentation include all that is necessary in this respect, but also (sometimes to the regret of the more advanced research workers) there is a policy that the instrumentation should be made as available as possible to all qualified students. We constantly find that the stimulus of this basic teaching leads us to some of the most interesting research possibilities in phonetics.

(The research reported in this paper is being conducted by the author, Victoria Fromkin, Frank Heny, Stan Hubler, Willie Martin, John Ohala, Rick Ratcliffe, Tim Smith, Ken Truitner, Ralph Vanderslice, Harry Whitaker, and others. It is supported by NSF Grant GS-1147 for Computer Studies of Speech, NSF Grant GS-1426 for Research on Linguistic Phonetics, USPHS Grant NB 04595 for Physiological Parameters for Synthesizing Speech, ONR contract N00014–67–A0111–0005 for Paralinguistic Studies of Speech, and UCLA funds.)