Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 142–171
(S. Karger, Basel/New York 1965).

From the Haskins Laboratories, New York City

Instrumental Methods for Research in Phonetics

By FRANKLIN S. COOPER

Instruments have long had a part in the development of phonetics. Their use has been both praised and derogated, depending in part on individual preferences as to whether phonetics should become a science or remain an art. The trend is clearly toward quantitative description and experimental method; hence, the purpose of this report is to review briefly the instrumental developments of the past few years.

A total review would surely exceed your patience and my abilities. Fortunately, the task can be reduced: (1) We can take only passing note of several areas of instrumentation that are of concern primarily to the communications engineer. This is hazardous, however, since these same tools may so easily be adapted to research on phonetic problems. (2) We can rely on existing reviews that cover most of the important developments until quite recently, in particular, the excellent and extensive reviews by *Eli Fischer-Jørgensen*[1] and *C. G. M. Fant*[2] presented at Oslo in 1958 and the reviews by *Fant*[3] and *Cooper*[4] given in Helsinki in 1961. What remains is to update the existing reviews and then perhaps to take note of some new trends in phonetic research and the related developments in instrumentation. There are hazards, to be sure, in presuming to point out significant new trends while they are yet trending: one's biases will surely be revealed, and quite possibly one's follies.

I. Instruments and Methods

Sound Recorders and Oscillographs

The use of magnetic tape for recording and reproducing the sounds of speech is so common and convenient that one tends to

forget his almost total dependence on recordings for every kind of phonetic research. Indeed, some types of studies can be done with little more than tape recordings, a razor blade, and some splicing tape.

Research in phonetics is well served by conventional magnetic tape recorders; even so, some of the many variants on the familiar recorder may prove useful for special purposes. Thus, multitrack recorders have the obvious advantage of allowing two or more distinct signals to be stored and recovered without loss of synchronism. Frequency modulated recordings preserve the audio waveform to zero frequency and are almost indispensable for such specialized tasks as recovering the glottal waveform by inverse filtering. Digital recording has advantages that may outweigh its cost when the experiments require an extremely good signal-to-noise ratio or a high degree of reproduceability on successive playings. A rather simple variant of the conventional magnetic recorder uses cards instead of tape on reels. This permits sentence-long utterances to be recorded, annotated, and filed away with other flat records; thus, a sound spectrogram and the speech it represents can be kept together for easy reference to either pattern or sound. Psychoacoustic tests that require the pairwise comparison of brief sounds are much simplified, as are comparisons of informants' utterances. Card-type recorders seem to have been used less than one might expect, perhaps because they have not been commercially available, at least in the United States, until the last year or two[5].

The display of sound waveforms by the oscillograph is, like sound recording, so commonplace that it is taken for granted. Although the waveform has proved to be less useful in phonetic research than spectrum information derived from it, there are many purposes for which the oscillogram is highly useful. Recent developments in instrumentation and recording media make it quite easy to record speech waveforms; several channels of information containing frequencies up to 5,000 cps or more can be registered on papers that require no processing except brief exposure to a fluorescent light[6].

Spectrograms and Spectrographs

Although the speech waveform contains all the acoustic information about a speech sound, it has long been known that the spectrum of the sound provides a more useful display. This was evident already from Fourier analyses of vocalic sounds that had been made even before this century. Nevertheless, the publication by the staff of the Bell Telephone Laboratories[7] of a series of articles and a book on sound spectrograms – Visible Speech, as it was called – caused much excitement among experimental phoneticians. This happened about 1945 to 1947, so that some of you may be a little surprised to find a much earlier spectrogram in the literature, showing the now-familiar dynamic patterns of speech in the same time-frequency-intensity display that is used for visible speech. This spectrogram appeared just over thirty years ago in a paper by *John C. Steinberg*[8] entitled "Application of sound measuring instruments to the study of phonetic problems" – a timely topic to this day.

Why should phoneticians have waited more than a decade to respond? Only a part of the reason is to be found in the intervening political events. Other reasons are, no doubt, that the hand-drawn patterns of *Steinberg's* spectrogram lacked the wealth of detail and artistic appearance of the machine-generated variety, and that his product also lacked the promotion given to visible speech, first by the Bell Telephone Laboratories, and then by *Martin Joos*[9] in his remarkable little book directed specifically to linguists. But we may suspect that, despite these reasons, the sheer labor of generating the patterns would have discouraged their use in research. *Steinberg's* spectrogram was in fact prepared from separate Fourier analyses of each vocal cord period, done by hand at something like two hours of work for each period. This single spectrogram, then, represents about a month of full-time effort devoted to measurement and calculation; today's spectrograph does the same job – better – in a few minutes. The moral is clear; a method, to be useful in phonetic research, must have both conceptual power and instrumental convenience, or, put in another way, new techniques that open old doorways may be almost as valuable as new insights.

The sound spectrograph, of the type manufactured by the Kay Electric Company under the trade name Sona-graph[10], based on the original Bell Telephone Laboratories' design, is now to be found in a number of laboratories. Although the instrument has its frailties, as most of its users will attest, it has proved extremely useful and is, indeed, the central piece of research instrumentation in most phonetics laboratories. This wide acceptance of a single instrument has had an important consequence in standardizing the pictorial aspects of the sound spectrogram; it is a further tribute to the original design

that even the improvements announced within the past year do not change the basic presentation, but only the mechanisms and circuitry. The new transistorized model of the Sonagraph has features of improved flexibility, ease of calibration and speed of analysis. Another new spectrograph, reflecting a thorough redesign to achieve an improved research instrument, has been made by Prestigiacomo and co-workers at the Bell Telephone Laboratories. This spectrograph[11] retains one of the best features of an earlier model, namely, a rotating magnetic head scanner that allows spectrograms to be made directly from original tape recordings; in addition, the tape system of the new spectrograph can serve as a conventional recorder. Modular electronics permit easy modification of signal conditioning circuits, filters, frequency scales, and other display characteristics. Even more important for routine use are the rugged mechanical design, rapid analysis, and the provisions for automatic calibration and for advancing the input tape. Another research instrument with many supplementary features has been designed and built by *Peterson*[12] at the University of Michigan, primarily for use in his own laboratory.

Despite the wide acceptance of the "standard" spectrograph for research in phonetics, a number of special-purpose instruments have been built. The principal objective has been to obtain real-time spectrograms of long samples of speech, or more generally to obtain running analyses whether or not the information is displayed as a conventional spectrogram. Banks of filters have long been used for this purpose and are now to be found in several laboratories; only two recent additions will be mentioned here. The 51-channel analyzer at the Royal Institute of Technology[13] is an outstanding example; it constitutes, in fact, the central unit in a very versatile research facility with capabilities that include the display and registration of real-time spectrograms and the production of digital tapes for computer use in further analyses of spectrum information. At Columbia University, *Harris*[14] has constructed a real-time audio-frequency spectrum analyzer consisting of a bank of 54 Gaussian filters. The outputs are scanned for presentation as running spectrograms on a radarscope, or converted to digital form for computer processing. Gaussian filters were employed to obtain good response to transient aspects of the speech signal.

The use of filter banks to generate real-time spectrograms has the disadvantage of poor frequency resolution or a requirement for very many filters. Another solution is to use methods that recirculate a short sample of speech at very high speed so that the requisite number of analyses can be completed during the duration of the sample, i.e., in real time. One such device, using digital delay lines and an analyzing filter has been described by *Gill*[15]; another that recirculates the analog signal in a quartz delay line for correlation analysis has been described by *Weiss* and *Harris*[16]. Equipment of this kind, though complex and costly, can provide high resolution spectrograms in quantity.

The nature of the display itself, when it is intended primarily for the human observer, has also been the subject of recent research. *Prestigiacomo*[17], following up the earlier work by his colleagues at the Bell Telephone Laboratories, has published contour spectrograms that seek to retain information about relative intensities that is often lost in the usual spectrogram because of the restricted grey scale of facsimile paper. The contour lines of these spectrograms mark off regions that differ by 6 db in marking level, and a rough identification of each such region is provided by the darkness of shading between the contour lines. Patterns of this kind have received rather wide publicity because of the claims made by *Kersta*[18] about their usefulness as voice prints for speaker identification.

Several variants of the conventional spectrographic display have been employed by *Wood* and *Hewitt*[19] at the General Electric Company in Schenectady. Their real-time analyzer is basically a filter bank which is scanned rapidly "by a unique process that interpolates the filter outputs to create continuous cross-sections of amplitude vs. frequency". The cross-sections are recorded from a cathode-ray tube onto film; in one version, the signal representing spectrum level is used not only to brighten the recording spot but also to deflect it very slightly along the time dimension. This results in some enhancement of the intensity variations across formants. In another version, the film is marked only at peaks of the spectral cross-section; this serves to delineate the course of the formants, as in the much earlier resonagrams described by *Huggins*[20].

C.P. Smith[21] at the Air Force Cambridge Research Laboratories has used a digital print-out of time- and frequency-quantized spectrograms to record numerical data without losing the essential pattern aspects of the spectrogram. The spectrum information in digital form can, of course, be used directly for quantitative experimental studies of speech.

Other Methods of Speech Analysis

There has been renewed interest in analyzing voiced segments on a period-by-period basis, employing variants of the inverse filtering method introduced by *R. L. Miller*[22] of the Bell Telephone Laboratories. The method is basically one of removing or cancelling the lower formants, one by one, from a high-fidelity recording o the wavetrain for a single glottal period. The residue then represents the glottal pulse itself, while the adjustments required to cancel the formants give formant frequencies and bandwidths. Inverse filtering is inherently a method of adjustment, though efforts have been made to automate it. *Lawrence*[23] at the Signals Research and Development Establishment, Christchurch, has referred to his equipment for cycle-by-cycle analysis as a "speech microscope". He has been concerned principally with the extraction of formant information and with adaptation of the method to the tracking of formants on a continuous basis. *Holmes*[24] at the Joint Speech Research Unit, and other workers as well, have used the method to study the details of the volume velocity waveform at the glottis.

The interest in events at the glottis includes, of course, the repetition rate or voice pitch, as well as the details of single glottal cycles. An exciting new method for obtaining the short-time average pitch has been described by *Noll*[25] of the Bell Telephone Laboratories. The method employs a spectrum analyzer designed to operate in real time and to produce high resolution spectra without using either heterodyning methods or bandpass filter banks – a device of considerable interest in its own right. The logarithms of consecutive amplitude spectra from this analyzer serve as the input to a second similar spectrum analyzer, the output of which is the "cepstrum", or power spectrum of the logarithmic spectrum. The cepstrum shows the period of the glottal pulses as peaks that are clearly separated from the peaks due to formants; thus, variations in voice pitch observed at analysis intervals of 15 msec. may be followed throughout a voiced stretch. The method has been tested by computer simulation with initial results that seem extremely promising. The term "cepstrum technique" has been applied to this method of vocal pitch detection.

The extraction of pitch information is a classic problem in the development of bandwidth compression equipment for voice communication. A very substantial amount of engineering effort has been devoted to it recently, with some success; however, none of the resulting devices really meets the need for a simple and thoroughly reliable pitch meter of the kind that could be so useful in phonetic research. One line of engineering development has, in fact, been to evade the problem by transmitting a part of the original speech signal (roughly the first formant region) to the receiving end of a communications link for use in regenerating the speech message. Automatic formant tracking is another active area of research in voice communication; here too, advances have been made. Research on the recognition of connected speech and on speaker identification must, of course, start with information about formant patterns and excitation functions and will probably need to make use of all available linguistic information as well. This is an area in which phonetic and engineering interests overlap, but it is also an area that has shown few striking gains. Quite possibly this is because no adequate rationale is yet available, with the possible exception of a method of analysis-by-synthesis which will be described in a later section. (The discussion in this brief paragraph has merely touched on some of the areas in which extensive engineering and instrumental developments have taken place. Fortunately for the present purpose, this field was reviewed in a Speech Communication Seminar that was organized by Dr. *Fant* and held at the Royal Institute of Technology, Stockholm, in August 1962[26]. The conference was aimed primarily at the field of communications although a number of the papers had a direct bearing on research in phonetics.)

Synthetic Speech and Speech Synthesizers

The great virtue of the sound spectrogram is that it displays the acoustic information as visual patterns that are correlated with the perceived sounds of speech and with the changing configurations of the vocal tract. The latter relationships are the subject of Dr. *Fant's* paper on "Formants and Cavities"; the correspondences between spectrographic pattern and linguistic unit are our present concern.

This concern stems from the need to identify the important aspects of spectrograms if we are to employ them in studying speech sounds and yet escape the fate of the investigator who, as *Fant*[3] has warned, "too easily drowns in a sea of details of unknown significance if he attempts to make use of all observable data". Fortunately, there is a direct way to examine the relationship between pattern

and linguistic unit and, in so doing, to determine the significance of various aspects of the spectrogram. The method uses the spectrographic pattern, often in simplified or derived form, to control the synthesis of sounds that are then presented to listeners for perceptual evaluation, exactly as if they were listening to "real speech". The method, as applied to research in phonetics, was pioneered by my colleagues at Haskins Laboratories[27]; it is now used to good effect in several laboratories to isolate the acoustic cues for speech perception, and to pin-point those spectrographic features that deserve the most careful attention. There will be opportunity at this Congress to hear about work based on these methods in the review paper by *Delattre*, and in a number of other papers as well.

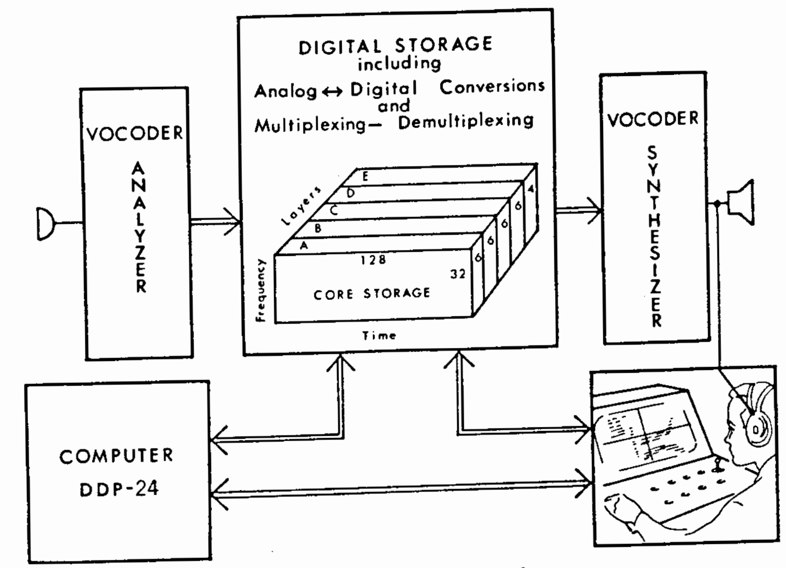
The situation with respect to speech synthesizers differs from that for sound spectrographs: synthesizers are less common, being found only in the larger laboratories, and there is as yet no "standard" instrument. This is due only in part to the complexity and cost of synthesizers. Different users have different needs and so have built synthesizers with widely differing characteristics. The relationships between intended use and machine design were discussed in a report[4] at the Helsinki Congress, with special emphasis on the control characteristics. One of the major considerations in the choice of a synthesizer is the convenience, and hence the overall effectiveness, with which the synthesis can be controlled. Conceptual convenience, as in the use of the spectrographic pattern itself for the control of synthesis, is no less important than mechanical simplicity. The overall complexity of the synthesizer and the difficulties of adequate control are functions also of the required felicity of the synthetic speech. High fidelity is indeed obtainable, but at a price; there is, in short, good reason to use care in selecting the synthesizer to match the problem.

The nature of the problem has been changing subtly as information has accumulated about the major acoustic cues. Much remains to be done at this level and with highly schematized patterns; nevertheless, some of the problems call for syntheses that are substantially closer to normal speech. Thus, in studies of stress and intonation, one may wish to impose various pitch lines on patterns that are otherwise as normal as possible; likewise, investigations of the relative importance of individual cues when more than one serves the same phonemic distinction may require controlled changes in the total spectrographic pattern derived from actual speech.

In studies for which the synthetic speech must approximate real speech, good use can be made of synthesizers of the vocal tract analog type, such as DAVO, or of the formant type that retain some of the inherent constraints of the vocal tract, such as OVE II or the series-connected PAT[3, 4]. Many of you have heard synthetic speech from OVE II that was difficult to distinguish from the real speech on which the synthesis was based. This realism was obtained, however, only by painstaking attention to the accuracy of the control signals and some adjustments by ear as well. Indeed, the virtue of these devices in modeling human speech production carries with it the penalty of interactions among the control signals. These make it more or less difficult (more for DAVO and less for OVE II) to introduce point-by-point changes in the spectrum, when one is seeking to test hypotheses about acoustic cues.

Considerations of this kind – and, I must admit, a strong bias for the use of spectrographic patterns in the control of such experiments – have led to the development of a new synthesizer at Haskins Laboratories that combines realtime spectrum analysis and display with a facility for performing "microsurgery" directly on the spectrogram, and of course for hearing the sounds of both the original and modified patterns. Because digital methods are used for storing the spectrum information, and because the primary purpose is to permit manipulation of the spectrum, we refer to the device as a Digital Spectrum Manipulator.

The diagram in Fig. 1 shows the mode of operation: essentially, a vocoder analyzer and synthesizer are used to provide the spectrum and to synthesize speech from it. Between analyzer and synthesizer, there is a magnetic core memory and the conversion equipment needed to get vocoder information into and out of it. The arrangements for addressing the core memory make it look to the outside world as if it were organized in layers, like a cake, with each layer in the form of a digital spectrogram very much like the ones employed by *Smith*[21]. Each spectrogram is about two-and-one-half seconds long; the operator can look at any half-second portion of it on a cathode-ray tube that presents a flickerless spectrographic pattern. He can use a "joy stick" control to move the pattern backward and forward at slow or normal rates or to move an indicator to any frequency channel. The particular data cell marked by this indicator and a vertical line in the center of the screen is available to him and can be changed instantaneously. Two other features deserve



Digital Spectrum Manipulator

Fig. 1. Digital Spectrum Manipulator: a device that allows the experimenter to generate and examine real-time spectrograms on a cathode-ray tube, make changes in the spectrum at any point in time or frequency, and see and hear both the modified and original versions.

mention: one is that several different versions of the same pattern can be stored in the several layers of the memory for easy comparison with each other; another is that more complex changes than point-by-point modification of the spectrum can be done by a small digital computer that communicates directly with the data store and the operator.

The principal disadvantages of the system, as it now stands, are the frequency resolution and voice quality limitations inherent in a 19-channel vocoder and a corresponding graininess, or checkerboard pattern, in the display. We expect that both these limitations can be considerably eased by using 32- or 64-channel filter banks for analysis and by employing either filters or formant generators for synthesis. The advantages of the system lie not only in the fact that one works with the full spectrum of a real utterance, but also in the ease with which speech can be entered from a microphone or tape recorder and then manipulated, with immediate access to both the visual pattern and the corresponding sound of either the original or the modified utterance. The system, though designed for speech, is readily usable for other research in which multi-channel analog information is put into a computer and manipulated under the direct scrutiny and control of the experimenter. Some of the work on physiological measures of articulation, to be mentioned in a following section, is of this kind.

A device somewhat like the Digital Spectrum Manipulator was used by *Smith*[21] to generate the digital spectrograms that have already been mentioned; it, too, makes use of digital storage for the automatic processing of spectrum information about real speech. Although the design had, as its primary objective, the evaluation of a bandwidth compression system based on coded spectral cross-sections, there were provisions for manipulating and modifying the spectrum in numerous ways so that the device could serve as a general-purpose research tool. Analysis and synthesis are performed by a vocoder, with the digital spectra stored in a drum memory. The principal outputs are from a high-speed printer and a loudspeaker. A substantial part of the equipment is devoted to matching spectral sections and compiling various kinds of statistical data, some of which may be quite relevant to problems in phonetics.

The very promising development by *Stevens, Dennis* and their colleagues[28] at the Massachusetts Institute of Technology of synthesizers comprising dynamic analogs of the vocal tract is being continued actively. One phase of their work, following the addition of a nasal analog (DANA) to the original device (DAVO), was the introduction of digital-to-analog conversion equipment to permit control of the synthesizers by the TX-O computer. This step was aimed at the greatly increased flexibility of control by computer as compared with that provided by the original circuitry. Another phase is the construction of a new synthesizer with circuits designed for improved stability and direct digital control. Attention is being given also to computer programs for controlling the synthesis in terms that are natural and convenient for the experimenter; this involves an intimate interplay of constraints imposed by the computer, the synthesizer, and the articulatory process.

Speech stretchers are special variants of synthesizers, but of more than passing interest for research in phonetics. One type, using tape recordings and rotating magnetic heads to repeat brief segments of the recording, has been available commercially for several years; such devices yield good speech when the time expansion ratio is small. The operation of other types of speech stretchers involves both automatic extraction of speech parameters and storage of this information for later read out in slow time. Digital devices with storage, such as those already mentioned, can serve as stretchers although analog methods are quite adequate and usually simpler. Such a speech stretcher has been built at Haskins Laboratories using a channel vocoder and a multi-track magnetic tape recorder to store and re-play the spectrum information that usually flows directly from analyzer to synthesizer. Figure 2 indicates the nature of the

Channel Vocoder

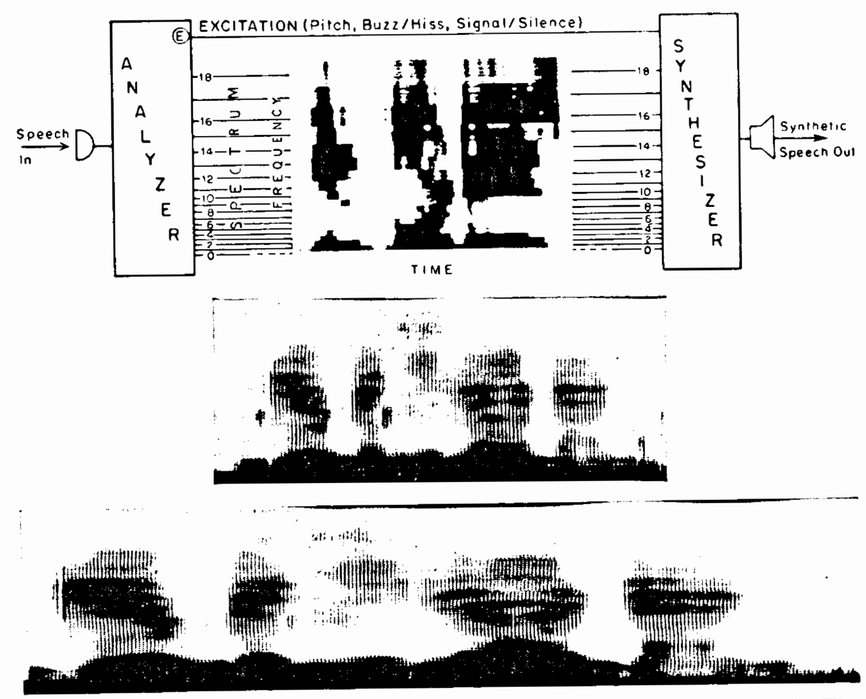


Fig. 2. One type of speech stretcher employs a vocoder for analysis and synthesis. The slowly-varying information that flows from analyzer to synthesizer, shown in essentially spectographic terms in the upper half of the figure, can be recorded on multi-track magnetic tape at one tape speed and then played into the synthesizer at a lower speed. The resulting speech will have the same spectral variations, except for the slower rate, and the same pitch as the original speech. A sample of original and stretched speech is shown in spectrographic form in the lower half of the figure.

information that is stretched, or compressed, and an example of the same speech at the normal rate and at half that rate. The quality of the stretched speech is unchanged by stretching, though it has the faults usual to vocoded speech. Stretching by factors of two-, four-, or eight-to-one, or compression by two-to-one, is done quite easily by changing motor speeds on the tape recorder. Other factors, ranging up to ten-to-one are possible, and have been used in making slow-motion X-ray movies of speech articulation[29].

II. Current Trends in Research on Phonetics

The developments mentioned in the preceding sections on instrumentation for speech analysis and speech synthesis represented mainly further progress along well established directions. Some of the devices that were described were new enough to have escaped mention at the Helsinki Congress in 1961, but not very many. There was increased emphasis on automated analysis and on synthesis applied to the complete speech spectrum. There was, in short, progress without important change of direction.

Let us turn now to some trends in phonetic research that represent significant differences in objective or major changes in research method. Here, too, we shall look in vain for abrupt departures but, over the past three to five years, the shift in direction is unmistakable.

The Use of Computers as Research Tools

Just as the mere automation of Fourier analysis by the sound spectrograph opened the way, in a practical sense, for most of the experimental advances of the past two decades, so the dexterity of the general-purpose digital computer seems about to open new paths for research. There is this difference: the spectrograph was a single-purpose instrument; the computer is extremely versatile, especially when coupled to analog devices at input and output. A survey of the applications of computers to speech, primarily for engineering purposes, would be beyond the scope of this paper; however, a sampling may serve to indicate some of the potential – and actual – uses of computers in research on phonetics.

Statistical analyses of data about speech sounds is, of course, an obvious use for the computer's extensive memory and its sorting capabilities. Such a study was reported, at an early phase, by *Denes* at the Helsinki Congress; details have since been published[30]. The initial objective of this study was to examine the frequency of occurrence of pairs of phonems in a large sample of running text as a basis for applying linguistic constraints to the operation of a phonetic typewriter. Quite aside from this practical objective, the frequencies of phoneme digrams and minimal pairs reveal some very interesting

characteristics of the English language and point, in particular, to the heavy functional load carried by certain phonemic distinctions. Comparable uses of computers with higher-order linguistic units have been especially numerous: the fields of information retrieval and of mechanical translation are largely computer-based; indeed, some of you may already have made plans to attend next year an International Conference on Computational Linguistics, a label that includes, according to the call for papers, "all uses of computers to manipulate natural or artificial languages".

Direct manipulation of the speech waveform by computers equipped with suitable analog-to-digital conversion devices and data buffers has been employed for a wide variety of tasks involving both analysis and synthesis. Such applications put heavy demands on the computer facility, in part because the speech waveform requires so much memory capacity per second of speech, and in part because the real-time inputs and outputs involve both precise timing and high data rates. The use of computer methods for making decisions about pitch periods in the waveform of voiced sounds has been described by *Gold*[31] of the Lincoln Laboratories, MIT. Several criteria are employed, related to those that might be used by a human observer in hand-marking the pitch periods; the computed period is a composite measure that is more reliable than the result from any single test. *Mathews*, *Miller* and *David*[32] at the Bell Telephone Laboratories have carried out Fourier analyses of individual pitch periods, obtaining measures of both the transfer function of the vocal tract and the glottal excitation function. Pitch-synchronous analysis has the virtue that results are directly comparable for a wide range of voices without regard to their pitch. *Pinson*[33], also at the Bell Telephone Laboratories, has performed comparable analyses by fitting a set of damped sinusoids to the speech waveform. By using only the portion of the vocal cord cycle during which the cords are closed, he obtained the transfer function for the vocal tract uncontaminated by information about glottal excitation. Computer methods for calculation of the volume velocity at the glottis is being reported by *Flanagan* at this Congress.

When the speech spectrum rather than the speech waveform is to be manipulated, the usual procedure is to use a bank of filters or some form of real-time spectrum analyzer ahead of the computer rather than to compute the analysis, primarily because such computations are time consuming and therefore expensive. A recent example of spectrum manipulation by computer is the work of *Harris* and *Weiss*[34] of Columbia University and the Federal Scientific Corporation who used a high-resolution, real-time spectrum analyzer of the correlation type and an IBM 7090 computer. Their programs yield information about the spectrum power, the amplitudes and frequencies of the first three formants, and a short-time average of the voice pitch derived from harmonic spacings. The speech spectrum has been used by a number of workers as a basis for speech recognition; in particular, the *Forgies*[35], working with the TX-2 computer at Lincoln Laboratories, have made extensive studies of the automatic recognition of individual phonemes. *Sandra Pruzansky*[36], at the Bell Telephone Laboratories, has used the method of matching spectral patterns to achieve automatic talker recognition; this was an extension of an earlier experiment by *Denes* and *Mathews*[37] on the recognition of spoken digits by pattern matching.

The simulation of devices rather than their construction is another field in which computers can be especially useful. An example is provided by *Golden*[38], working with the IBM 7090 computer at the Bell Telephone Laboratories, in simulating a voice-excited vocoder. Although it required nearly 200 times as long to compute the speech from the simulated vocoder as would have been required by an actual device, the simulation was not only of high quality, but also allowed a number of design changes to be tested quickly without the cost of building and replacing equipment. Another example, already cited, is the cepstrum pitch analyzer; here, too, construction of the actual device would have been laborious and costly.

The use of computers to control the operation of other devices may require about the same peripheral equipment that is employed in simulation. Nevertheless, there is a significant difference and some of the most exciting prospects appear to lie in this area of application to phonetic research. The reason is simple enough: by divorcing the control function from the signal processing operation, it becomes easier for the human experimenter to intervene directly in the experiment. He can use a computer as a lively and tireless assistant, as my colleagues and I plan to do with our Digital Speech Manipulator, or he can guide the computer's eager efforts by supplying an element of judgment and direction. One example of the latter type is the use by *Denes*[39] of a digital computer in recovering the glottal pulse waveform by inverse filtering. The technique is essentially

that used by *Miller* and others[22–24], but without the requirement for special equipment. Another application, likewise not strictly dependent on a computer but making good use of it, is the analysis-by-synthesis procedure that will be discussed in a later section.

The production of speech directly from words or phoneme strings is another area in which computers have an advantage over analog methods in speed and facility. Various methods have been described[40], for generating speech by rules from dyads, from diphones, or from stored data in various other forms. The first use of a general-purpose digital computer in generating speech by rule from a phonemic transcription was made by *Gerstman* and *Kelly*[41]; many of you have heard the famous soliloquy from Hamlet as rendered by the IBM 7090.

These are a few of the ways computers have been used in processing speech information. The advantages have been stressed rather than the disadvantages, though there are some of the latter. Computers are expensive, both in money and in programming effort; also, for speech use, they are likely to require more analog accessories than might be supposed. Finally, they may prove to be surprisingly slow in overall performance despite the enormous speed with which their internal transactions are conducted; speech programs that run in real time are, as of today, the exception rather than the rule.

The trend to computers is nevertheless real and significant; indeed, an awareness of computer capabilities is becoming a minimal requirement for following research in experimental phonetics. The reasons for this trend were explained by *Denes*[42] in his paper at Helsinki on "The use of computers for research in phonetics". He points out that, although extensive facilities are required, "these facilities are available without the need for any electronic design on the part of the experimenter: that has been taken care of by the designer of the computer. All the phonetician has to do is to write down a series of instructions that embody the logical sequence of steps that he wants to have carried out: the computer will do the rest." In short, somebody else has done the work, insofar as equipment is concerned. *Denes* points out also the advantage of programming over construction in allowing the experimenter to keep his attention more closely focussed on the logic of the experiment. He concludes, as I would, by saying: "...there is every indication, therefore, that the availability of computers is about to produce a profound change in the way in which the experimental phonetician approaches his problems, in the range of experiments open to him and in the kind of training he requires to enable him to carry out his work." I would add only that the course of research over the last three years bears out this prediction.

Emphasis on the Changing Configurations of the Vocal Tract

A concern with the shape of the vocal tract is certainly not new in phonetics; nevertheless, the resurgence of interest in this topic is one of the significant trends in current research. What is new is not so much the area of interest as the reasons for that interest. One of these is the adoption as a working hypothesis of the idea that speech perception is somehow closely linked to articulation. This is entirely consistent with the view that sound spectrograms are so useful precisely because of an underlying relationship between the articulatory sequences they display and the linguistic units of the utterance. A second reason is the development of a powerful new method for studying speech.

The method of analysis-by-synthesis as developed by *Stevens* and his co-workers at MIT [43] was used initially as a means for locating the formant frequencies of slowly-changing vocalic sounds by matching the spectrum of the actual speech against a computed spectrum. The computation makes use of variable parameters with constraints set by a "resonance" model of speech generation. Fig. 3 illustrates the analysis-by-synthesis procedure: the incoming speech information (from a bank of filters) is held in temporary storage and presented to the comparator as the input spectra. Computed spectra are also presented to the comparator; these are formed in the spectrum generator by some strategy that will minimize the error score, i.e., the difference between the two spectra. This strategy may be carried out either by a computer or by a human observer working from a cathode-ray display. The principal limitations of the method are the adequacy of the model for speech production (incorporated into the program of the spectrum generator) and the strategy for minimizing errors, though the latter affects principally the speed and level of automation that can be obtained.

The acoustical theory [44] implicit in the operation of the spectrum generator treats the spectrum of the vocal tract output (in decibels) as the sum of a source spectrum, a transfer function, and a radiation

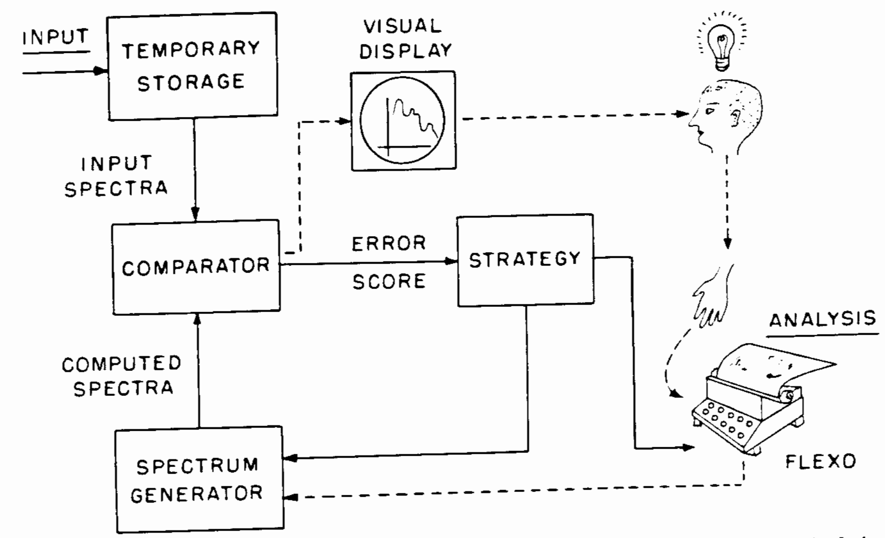


Fig. 3. Schema for the method of analysis-by-synthesis [Reproduced from *Speech Analysis and Synthesis Final Report* (Ref. [28]) with the kind permission of Prof. *K. N. Stevens*].

characteristic. Most of the spectrum variation, within a given class of speech sounds, is assignable to the transfer function which, in turn, is determined by the articulatory configuration and source location. In the case of non-nasal vowel and vowel-like sounds, the transfer function comprises only poles, or resonances; for nasal and fricative sounds, the transfer function will have zeros, or antiresonances, as well as poles. Thus, the analysis-by-synthesis technique based on the resonance model for speech generation yields a description of the speech spectra in terms of the poles and zeros of the vocal tract transfer function (and also of generalized source and radiation characteristics). The description is, in effect, an analytic version of the sound spectrogram; ongoing articulation is represented by movement of poles and zeros over the complex plane in the one case and by formant transitions and spectral shifts in the other.

The method can be extended, however, to deal with relationships between spoken sound and vocal tract configuration, if use is made of an "articulatory" model for the generation of the computed spectra. The box marked strategy in the figure, or the human operator, must now provide information to the spectrum generator in terms of cross-sectional areas of the vocal tract at successive points along its length. The computation then proceeds in two stages. First, the poles and zeros of the transfer function are computed from the shape of the acoustic tube, and second, the computed spectra are generated from these pole-zero specifications. Both strategy and computation are now more complex, but an important advantage has been gained, particularly for the consonant sounds with their more complicated spectra. The articulatory configurations must change smoothly and continuously in proceeding from vowel to consonant, whereas the pole-zero configurations need not show comparable continuity. Also, it seems reasonable to expect that the constraint on the shape parameters, due to physical constraints on the organs of speech, might be simpler than constraints that operate at the acoustic level.

It may be feasible to extend the general method of analysis-by-synthesis still further to include a spectrum generator that makes use of rules governing the generation of speech sounds in accordance with a set of phonetic parameters "which are considered to be the result of a set of neuromuscular instructions or excitations that operate on the mechanical system consisting of lips, tongue, velum, and so forth. It is hypothesized that, for a given speech sound, these excitations assume a set of steady values, and jump discontinuously from one set of values to another as the sequence of speech sounds is generated. It is assumed that a unique set of excitations is associated with each speech sound. Relations between the excitations and the resulting articulatory motions are undoubtedly not linear, but it is convenient to characterize the mechanical properties of the articulatory structures in an approximate manner by a set of system functions[45]." Thus, it may eventually be possible to arrange the strategy box so that its signals to the spectrum generator (and to the outside world) are in the form of segmental phonemes. The assumption implicit in this extrapolation from current capabilities is that a simple relationship exists between the shape of the tract, described in phonetic parameters, and the linguistic units of the message. Whether or not this expectation is realized, the power of the analysis-by-synthesis procedure more than justifies the resurgence of interest in vocal tract configurations and their relation to the speech that is produced.

The major instrumentation used thus far in analysis-by-synthesis studies has been a filter bank, a general-purpose digital computer and its peripheral cathode-ray display that permits the operator to observe and control the spectrum matching operation.

An obvious further requirement in studying articulatory configurations is, of course, to obtain as direct information as possible about events in the human vocal tract during speech. X-ray motion pictures are a principal source of such information[46], though high-speed photography of the lips and of the vocal cords is also useful[47]. X-ray techniques were discussed at length in a review paper by the *Subtelneys* at Helsinki, and their synchronization with speech and spectrogram were described by *Truby*[48] at the same Congress. There have been extensions in X-ray technology within the past few years, primarily in the direction of higher frame rates and the use of pulsed emission, both of which improve time resolution. Viewing equipment that permits flickerless projection at very low frame rates is extremely useful, though not new. The combination of high-speed photography and speech stretching brings to slow-motion projection the added dimension of synchronous sound.

We may note, at this point, not only a trend toward studies of articulation, but also one away from an overriding concern with relationships between acoustic signal *per se* and linguistic unit. Undeniably, the acoustic signal must be heard to be understood, and it is useful to characterize the sounds of speech by the acoustic cues that ensure their recognition; nevertheless, the research trends discussed here and in the following sections do reflect a shift in emphasis from acoustic to articulatory aspects of the speech event.

Emphasis on the Perceptual Nature of Linguistic Units

The trend toward greater attention to the act of speaking is due in considerable part to gains in our understanding of the perceptual nature of linguistic units (at the level of phonemes or distinguishing features). The very special qualities of language, most notably its exceptional efficiency as an acoustic signalling system, are of course intimately bound up with a categorizing process that reshapes and regroups incoming sound into discrete units of the same kind as those that control the production of speech. Clearly this categorizing process and the perceptual nature of the categorical units themselves are of primary interest in understanding the relationship of sound to language.

The nature of this relationship has been studied intensively. One line of investigation, although it concerned itself initially with acoustic cues for perception, was led by the experimental evidence to a motor theory of speech perception. The involvement of motor reference in the perception of speech sounds, and the evidence for it, was discussed in some detail by *Liberman* in 1957 and again in 1962[49]; according to this theory, the perceptual conversion of incoming sound stuff to linguistic units can, and often does, make use of the greater distinctiveness of the articulatory gestures that produce the sounds than exists in the sounds *per se*. What concerns us here, however, is not so much the theory, or the evidence for and against it, as the questions it poses about the nature of linguistic units and the rationale it provides for certain kinds of experiments.

Much of the research on the perceptual nature of linguistic units is concerned with their distinctiveness and with the differences in distinctiveness of different classes of units. It has been found, for example, that the stop consonants are perceived in a highly categorical manner, which is roughly equivalent to saying that even when the sound stimulus differs substantially from the "norm" a listener will still hear the same phoneme and may, indeed, be unable even to detect the deviation. The situation with vowels is quite different: the listener readily detects small differences in the acoustic signal but his identifications of deviant stimuli are less sure and are strongly influenced by context[50]. It may be noted in passing that this marked difference in the perceptual nature of vowels and consonants could have been predicted from the theory. The line of investigation indicated here is currently a very active one and it seems likely to lead to additional insights into some of the fundamental problems of phonology[51]. As to techniques and instrumentation, this kind of research differs from much that has been discussed thus far in not being instrument-bound; rather, a wide variety of techniques is employed. The principal instrumental requirements are speech synthesizers for the production of controlled test stimuli and devices for measuring reaction time and muscle activity.

Emphasis on Neuromuscular Aspects of Speech Gestures

The hypothesis that speech perception is somehow closely related to articulation has served to motivate research on neuromuscular aspects of speech gestures, just as it has prompted increased attention to the changing configurations of the vocal tract. Are these essentially the same approach, or is there a significant difference? Certainly they are related, since muscle contractions and cavity shapes stand to each other as cause to effect; but muscle contractions, being the prior events, may stand in the more direct relation to linguistic units. Thus, the sequence of events in production, which presumably starts with the intended message organized in appropriate linguistic units, proceeds by successive transformations of the message into motor command signals on the neural pathways to the articulators, then into contractions of selected sets of muscles in suitable time patterns, then into the articulatory configurations caused by these contractions, and finally into the stream of sounds that we call speech. Each of these conversions blurs still further the relationship with the initial linguistic units of the message. The encoding of motor commands into shapes and movements is an especially complex transformation because of the constraints inherent in the bone-and-muscle mechanism of the vocal tract; since the motor commands operate ahead of these complications, they escape this kind of recoding. Hence, of all the speech events – acoustic signal, articulatory shape, or neuromotor command – about which we can reasonably expect to collect information, the neural commands to the articulators would seem to provide the simplest relationship to the intended and perceived units of the message.

The experimental objective, therefore, is to get a description of speech in terms of motor commands or, since muscle contraction follows so directly from neural signal, in terms of patterns of muscle contractions and their relation to the phonemes (or distinguishing features) of the message. The total pattern of contractions may, however, contain much that is irrelevant to the phonemic distinction so that the task, put more sharply, is to seek for those component parts of the total gesture, the "characteristic gestures", that relate most directly to linguistic units[52].

The experimental procedure depends primarily on electromyography, i.e., on the fact that muscle contraction is accompanied by electrical signals that can be collected by putting electrodes on the surface of the face or tongue, or inserting small needles into the muscles. Although electromyography gives the most direct information about muscle activity, other techniques are valuable whenever reliable inferences can be drawn from the data they provide: pressure measurements both above and below the glottis, using the procedures developed by *Ladefoged*[53] give valuable information about the timing of occlusions and releases; throat microphones and equipment for transillumination[54] both yield valuable data about activity at the glottis; X-rays, particularly in slow motion with synchronous sound, can be helpful in suggesting electrode placements; and sound spectrograms can serve as a guide in planning experiments and as a check on inferences drawn from other data.

The instrumentation for research of this kind is mostly conventional though some of it must be designed for the specific task. It may be useful to describe the facilities developed at the Haskins Laboratories in starting this field of research[55, 52]. One of the first, and most troublesome, problems was to devise an electrode system that could be used inside the mouth and on the tongue, as well as on the face. A solution was found in the use of the small suction cup electrodes shown in Fig. 4. The electrode is hemispherical, about

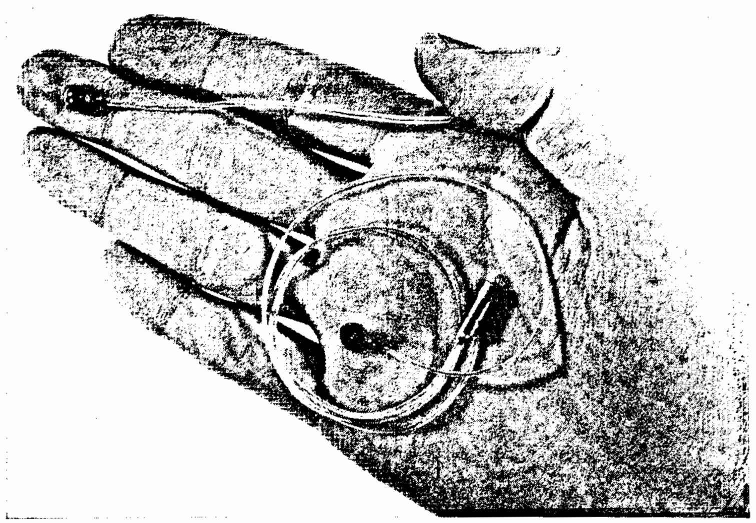


Fig. 4. Suction cup electrodes used in studies of muscle contraction by surface electromyography.

6 mm in diameter, and connected to a brass plug by a flexible plastic tube about 50 cm long. A thin steel wire threaded through the tube provides electrical connection to the plug, which can be inserted into a manifold that applies vacuum to all the electrodes as well as electrical connection from each electrode to its own preamplifier. The preamplifiers are commercial units designed for low noise, high common-mode rejection and high amplification. The system, shown diagrammatically in Fig. 5, includes eight electromyographic channels and eight other channels for varied uses. Each channel has additional amplification and facilities for monitoring signal levels

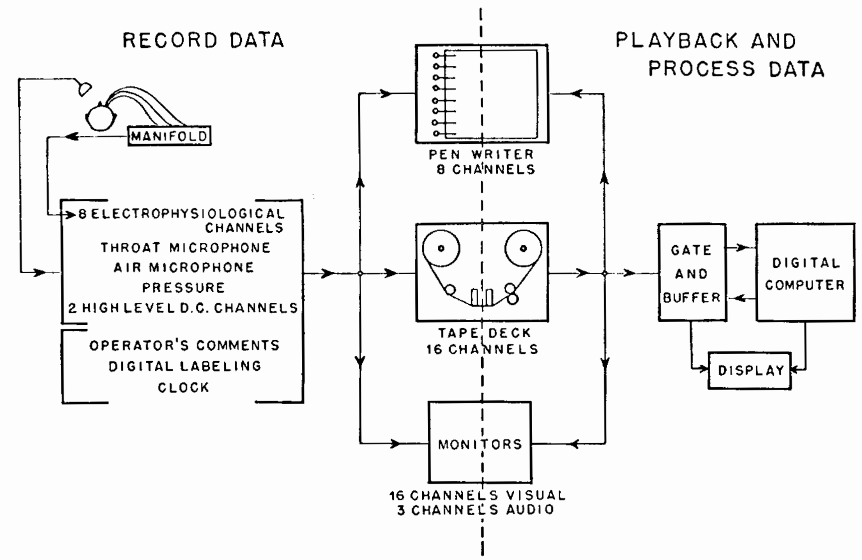


Fig. 5. Diagram of equipment and procedures for physiological studies of speech articulation. Data collection is indicated on the left of the central dashed line and data processing on the right.

during an experiment. At present, the non-myographic channels are used for a conventional microphone signal, a throat microphone, two pressure transducers, two digital signals for timing and identification of events, a pre-recorded sequence of utterances used to guide the subject's productions, and a microphone for comments by the experimenter. All sixteen channels are recorded on 1 inch magnetic tape as the experiment proceeds; parallel use is made of an 8-channel pen recorder, primarily to monitor the signals from electrodes and throat microphone. After the experimental session has been concluded, the monitoring amplifiers and pen recorders can be used to transcribe the experimental data and the event codes from the magnetic tape. Thus far, data reduction has all been done by hand methods – a slow and laborious procedure. Arrangements are almost complete to shift this burden to a small computer which will read the data directly from the magnetic tape and, with the aid of event codes and clock signals, sort and process it. The experimenter can control the operation by prior examination of the pen traces or by observing a cathode-ray display of the data as it arrives at the computer.

The trend toward physiological problems and methods in research on phonetics is not, of course, limited to work aimed primarily at a motor command description of speech production and perception. *Ladefoged*[56] and his co-workers at the University of California are following up the work reported at Helsinki and elsewhere on subglottal activity during speech with an extensive program aimed at describing and quantifying not only subglottal events in speech production but activities of the entire vocal tract. X-ray methods have been used to obtain parameters of characteristic lip positions in American English vowels, and analog models of the vocal tract are under construction. *Philip Lieberman*[57], at the Air Force Cambridge Research Laboratory, has combined techniques for high-speed photography of the glottis, waveform recording on the high-speed film, and computer-aided processing of the pictures to conduct several studies of vocal cord function. *André Malecot*[58], at the University of Pennsylvania, has developed some very ingenious equipment with which to test the ability of subjects to estimate intraoral pressures such as they themselves might generate in producing stop consonants. This bears, of course, on the question of whether or not direct awareness of such pressure differences as exist between fortis and lenis productions could provide a speaker with usable feedback information about his productions. The apparatus transmits pressure pulses of controlled magnitude and duration to the subject's mouth or to pressure plates at lips or palate. Mention has been made already of methods for transilluminating the glottis[54]; in the main, these require an open-mouth position, but it seems likely that extensions of this technique to running speech will soon be possible. A recent article by *Kozhevnikov et al.*[59] on instrumentation for phonetic research includes a description of an artificial palate with a number of electrodes that permit a running characterization of the patterns of tongue contact.

A substantial part of the physiological research on speech is, of course, medically oriented, especially to the problems of cleft-palate speech[60]. Several X-ray installations designed primarily for speech research are now available in the United States; in Sweden there is the excellent facility at the Wenner-Gren Foundation originally adapted to speech research by *Truby* and used by him for studies of the vocalization of infants.

Summary

In summary, this discussion of instrumental methods for research in phonetics has attempted, first of all, to report the principal new

instruments and techniques that have appeared since the Helsinki Congress three years ago. There were not many, though the tempo of development in related fields such as the bandwidth compression of voice signals suggests that some new tools for phonetic research may be expected soon.

The discussion then turned – or was bent – to the broader topic of the directions in which phonetic research is moving. It seems to this reviewer that some changes have occurred within the last few years and that they and the reasons for them are significant. One such trend is toward a major role for computers in many phases of research on phonetics, not because computers generate exciting new ideas, but simply because they are versatile, fast, and ready to be used on problems that were heretofore unapproachable.

The other trends for which significance was claimed depend less on new technology than on ideas about the nature of phonological units. Because these units depend on perceptual operations for their very existence as categorical entities, the linkage of perception to speaking or to listening is a crucial one. The idea that perception may be closely tied to the act of speaking represents something of a departure from the views that guided phonetic research during the 1950's. It carries with it the implication that there is a better chance of finding one-to-one relationships between the things one measures and the atomic units of language when working with articulation than when working with the acoustic signal. But measures of articulation can be either in terms of the shape of the vocal tract or of the neuromotor commands that control it. The latter are upstream with respect to the complex transformation that relates command to shape, and so may provide the simpler description – though quite possibly the ideal of a one-to-one correspondence may not be realizable short of the central nervous system.

So much for motivations. The consequence, in some laboratories at least, has been a shift in emphasis toward studies of the articulatory process by various means: notably, by the method of analysis-by-synthesis; by the use of X-ray movies; by the methods of experimental psychology applied to the perceptual nature of consonants and vowels; and by electromyographic and other physiological measures of speech gestures.

It is yet too early to exhibit the results of these lines of research and thereby to demonstrate their significance but, by the same token, your reviewer is not yet subject to the full penalties for false prophecy.

References

1. *Fischer-Jørgensen, Eli:* What can the new techniques of acoustic phonetics contribute to linguistics? Proc. 8th Int. Congr. Linguists, pp. 433–499 (Oslo 1958).
2. *Fant, C. G. M.:* Modern instruments and methods for acoustic studies of speech. Proc. 8th Int. Congr. Linguists, pp. 282–362 (Oslo 1958).
3. *Fant, C. G. M.:* Sound spectrography. Proc. 4th Int. Congr. Phonetic Sciences, Helsinki 1961, pp. 14–33 (The Hague 1962).
4. *Cooper, F. S.:* Speech synthesizers. Proc. 4th Int. Congr. Phonetic Sciences, Helsinki 1961, pp. 3–13 (The Hague 1962).
5. Language Master: Audio Instruction Device, mfg. by Bell and Howell, 2100 McCormick Rd, Chicago 45, Illinois, USA.
6. The *Visicorder:* mfg. by Heiland Division, Minneapolis-Honeywell Regulator Co., Denver 10, Col., is typical of several such recorders made in the USA; Other manufacturers include Century Electronics & Instruments, Tulsa, Okla., Consolidated Electrodynamics, Pasadena, Calif., and Midwestern Instruments, Tulsa, Okla.
7. *Potter, R. K.:* Visible patterns of sound. Science *102:* 465 (1945). – Six articles by *Potter, R. K. et al.* in J. acoust. Soc. Amer. *18:* 1–89 (1946).
   Potter, R. K.; Kopp, G. A. and *Green, Harriet C.:* Visible Speech (New York 1947).
8. *Steinberg, J. C.:* Application of sound measuring instruments to the study of phonetic problems. J. acoust. Soc. Amer. *6:* 16–24 (1934). The spectrogram referred to in the text is Fig. 9.
9. *Joos, M.:* Acoustic phonetics. Lang. Monogr. 23, Language *24:* 2. Suppl. (1948).
10. Kay Electric Company, Pine Brook, N. J., USA.
11. A spectrograph, based on a recent Bell Telephone Laboratories' design, has been announced for sale by Voiceprint Laboratories, Somerville, N. J.; for an account of the earlier spectrograph mentioned in the text, see *Prestigiacomo, A. J.:* Plastic tape sound spectrograph. J. Speech Dis. *22:* 321–327 (1957).
12. Personal communication; a description is to be published.
13. Quarterly Progress and Status Report, Speech Transmission Laboratory, R. I. T., Stockholm, STL-QPSR *4:* 1–6 (1963).
14. *Harris, C. M.* and *Waite, W. M.:* Gaussian-filter spectrum analyzer. J. acoust. Soc. Amer. *35:* 447–450 (1963). – Display of sound spectrographs in real time. J. acoust. Soc. Amer. *35:* 729 (L) (1963).
15. *Gill, J. S.:* A versatile method for short-time spectrum analysis in "real time". Nature *189:* 117–119 (1961).
16. *Weiss, M. R.* and *Harris, C. M.:* Computer technique for high-speed extraction of speech parameters. J. acoust. Soc. Amer. *35:* 207–214 (1963).
17. *Prestigiacomo, A. J.:* Amplitude contour display of sound spectrograms. J. acoust. Soc. Amer. *34:* 1684–1688 (1962).
18. *Kersta, L. G.:* Voiceprint identification. J. acoust. Soc. Amer. *34:* 725 (A) (1962). – Voiceprint-identification infallibility. J. acoust. Soc. Amer. *34:* 1978 (A) (1962).
19. *Wood, D. E.* and *Hewitt, T. H.:* New instrumentation for making spectrographic pictures of speech. J. acoust. Soc. Amer. *35:* 1274–1278 (1963).
   Wood, D. E.: New display format and a flexible-time integrator for spectral-analysis instrumentation. J. acoust. Soc. Amer. *36:* 639–643 (1964).
20. *Huggins, W. H.:* A phase principle for complex-frequency analysis and its implications in auditory theory. J. acoust. Soc. Amer. *24:* 582–589 (1952).

21. *Smith, C. P.:* Voice-communication method, using pattern matching for data compression. J. acoust. Soc. Amer. *35:* 805 (A) (1963). – Pitch intensity and spectrum stereotypes in vocoder-analysis data for readings of phonetically balanced word lists. J. acoust. Soc. Amer. *35:* 807 (A) (1963). – An approach to speech bandwidth compression. Proc. Seminar on Speech Compression and Processing. Electronics Research Directorate, Air Force Cambridge Research Center (1959). Avail. fr. U.S. Dept. of Commerce, Office of Technical Services, Washington, D.C.
   Kerby, A. E. and *Schoene, L. D.:* A digital voice data processer, ibid.
22. *Miller, R. L.:* Nature of the vocal cord wave. J. acoust. Soc. Amer. *31:* 667–677 (1959).
23. *Lawrence, W.:* Formant tracking by self-adjusting inverse filtering. J. acoust. Soc. Amer. *35:* 1113 (A) (1963).
24. *Holmes, J. N.:* An investigation of the volume velocity waveform at the larynx during speech by means of an inverse filter. Proc. Speech Communications Seminar, Speech Transmission Laboratory, RIT (Stockholm 1963).
25. *Noll, A. M.:* Short-time spectrum and "cepstrum" techniques for vocal-pitch detection. J. acoust. Soc. Amer. *36:* 296–302 (1964).
26. Proc. Speech Communications Seminar, Stockholm 1962, Speech Transmission Laboratory, RIT (3 vols.) (Stockholm 1963).
   Summaries of the papers presented at Stockholm are to be found in J. acoust. Soc. Amer. *35:* 1112–1117 (1963); a brief announcement about the meeting, ibid., *34:* 1793 (1962).
27. *Liberman, A. M.:* Some results of research on speech perception. J. acoust. Soc. Amer. *29:* 117–123 (1957).
   Delattre, P.: Les indices acoustiques de la parole: premier rapport. Phonetica *2:* 108–118; 226–251 (1958).
   Liberman, A. M.; Ingemann, F.; Lisker, L.; Delattre, P. and *Cooper, F. S.:* Minimal rules for synthesizing speech. J. acoust. Soc. Amer. *31:* 1490–1499 (1959).
28. Speech Analysis and Synthesis Final Report, Massachusetts Institute of Technology, Research Laboratory of Electronics, Cambridge, Mass. (1963), doc. no. AFCRL-64-300, avail. fr. U.S. Dept. of Commerce, Office of Technical Services, Washington, D.C.
   Various papers in Proc. Speech Communication Seminar, Speech Transmission Laboratory, RIT, Stockholm (1963).
   Dennis, J. B.; Whitman, E. C. and *Tomlinson, R. S.:* On the construction of a dynamic vocal-tract model. J. acoust. Soc. Amer. *36:* 1038 (A) (1964).
   Dennis, J. B.: Control language for experimental speech synthesis. J. acoust. Soc. Amer. *36:* 1039 (A) (1964).
29. *Abramson, A. S.* and *Cooper, F. S.:* Slow-motion X-ray pictures with stretched speech as a research tool. J. acoust. Soc. Amer. *35:* 1888 (A) (1963).
30. *Denes, P. B.:* On the statistics of spoken English. J. acoust. Soc. Amer. *35:* 892–904 (1963).
31. *Gold, B.:* Computer program for pitch extraction. J. acoust. Soc. Amer. *34:* 916–921 (1962).
32. *Mathews, M. V.; Miller, Joan E.* and *David, E. E., Jr.:* Pitch synchronous analysis of voiced sounds. J. acoust. Soc. Amer. *33:* 179–186 (1961).
33. *Pinson, E. N.:* Pitch-synchronous time-domain estimation of formant frequencies and bandwidths. J. acoust. Soc. Amer. *35:* 1264–1273 (1963); for another study employing comparable techniques, see:
   Manley, H. J.: Analysis-synthesis of connected speech in terms of orthogonalized exponentially damped sinusoids. J. acoust. Soc. Amer. *35:* 464–474 (1963).
34. *Harris, C. M.* and *Weiss, M. R.:* Pitch extraction by computer processing of high-resolution Fourier analysis data. J. acoust. Soc. Amer. *35:* 339–343 (1963).

35. *Forgie, J. W.* and *Forgie, Carma D.:* Scheme for utilizing formant-transition cues in the recognition of a class of prevocalic consonants; also Automatic method of plosive identification. J. acoust. Soc. Amer. *34:* 1979 (A) (1962).
36. *Pruzansky, Sandra:* Pattern-matching procedure for automatic talker recognition. J. acoust. Soc. Amer. *35:* 354–358 (1963).
37. *Denes, P. B.* and *Mathews, M. V.:* Spoken digit recognition using time-frequency pattern matching. J. acoust. Soc. Amer. *32:* 1450–1454 (1960).
38. *Golden, R. M.:* Digital computer simulation of a sampled-data voice-excited vocoder. J. acoust. Soc. Amer. *35:* 1358–1366 (1963).
39. Personal communication; see also
Miller, Joan E. and *Mathews, M. V.:* Investigation of the glottal waveshape by automatic inverse filtering. J. acoust. Soc. Amer. *35:* 1876 (A) (1963).
40. *Liberman, A. M.; Ingemann, F.; Lisker, L.; Delattre, P.* and *Cooper, F. S.:* Minimal rules for synthesizing speech. J. acoust. Soc. Amer. *31:* 1490–1499 (1959).
Peterson, G. E.; Wang, W. S-Y and *Sivertsen, Eva:* Segmentation techniques in speech synthesis. J. acoust. Soc. Amer. *30:* 739–742 (1958).
Estes, S. E.; Maxey, H. D. and *Walker, R. M.:* Speech synthesis from stored data. IBM Journal *8:* 2–12 (1964).
Cooper, F. S.: Speech from stored data. IEEE Convention Record *7:* 137–149 (1963).
41. *Gerstman, L. J.* and *Kelly, J. H.:* An artificial talker driven from a phoneme input. J. acoust. Soc. Amer. *33:* 835 (A) (1961); see also
Holmes, J. N. and *Shearme, J. N.:* Speech synthesis by rule controlled by a small, low-speed digital computer. J. acoust. Soc. Amer. *35:* 1911 (A) (1963).
42. *Denes, P.:* The use of computers for research in phonetics. Proc. 4th Int. Congr. Phonetic Sciences, Helsinki 1961, pp. 149–154 (The Hague 1962).
43. *Paul, A. P.; House, A. S.* and *Stevens, K. N.:* Automatic reduction of vowel spectra: an analysis-by-synthesis method and its evaluation. J. acoust. Soc. Amer. *36:* 303–308 (1964); also, the Final Report cited in Ref. 28, with its references.
44. *Stevens, K. N.* and *House, A. S.:* An acoustical theory of vowel production and some of its implications. J. Speech and Hear, Res. *4:* 303-320 (1961); for a complete treatment, see:
Fant, G.: Acoustic theory of speech production (Mouton, 's-Gravenhage 1960).
45. Speech Analysis and Synthesis: Final Report: 74–75. See Ref. 28.
46. Work in this field continues to be active, as evidenced by the research reported at this Congress by *Faure, Öhman* and *Lindblom, Ondráčková, Truby* and *Wierzchowska;* a further example is the work of one group concerned with quantitative measures of movement derived from x-ray movies:
Öhman, S. E. G. and *Stevens, K. N.:* Cineradiographic studies of speech: procedures and objectives. J. acoust. Soc. Amer. *35:* 1889 (A) (1963).
Heinz, J. M. and *Stevens, K. N.:* On the derivation of area functions and acoustic spectra from cineradiographic films of speech. J. acoust. Soc. Amer. *36:* 1037 (A) (1964).
Lindblom, B.E.F.: Articulatory activity in vowels. J. acoust. Soc. Amer. *36:* 1038 (A) (1964).
47. *Fujimura, O.:* Bilabial stop and nasal consonants: a motion picture study and its acoustical implications. J. Speech and Hear, Res. *4:* 233–247 (1961); see also Ref. 57.
48. *Subtelny, Joanne D.* and *Subtelny, J. D.:* Roentgenographic techniques and phonetic research. Proc. 4th Int. Congr. Phonetic Sciences, Helsinki 1961, pp. 129–146 (The Hague 1962).
Truby, H. M.: Synchronized cineradiography and visual-acoustic analysis, ibid. pp. 265–279; see also
Moll, K. L.: Cinefluorographic techniques in speech research. J. Speech and Hear, Res. *3:* 227–241 (1960).

49. *Liberman, A. M.:* Some results of research on speech perception. J. acoust. Soc. Amer. *29:* 117–123 (1957).
Liberman, A. M.; Cooper, F. S.; Harris, K. S. and *MacNeilage, P. F.:* A motor theory of speech perception. Proc. Speech Communication Seminar, Stockholm 1962, Speech Transmission Laboratory, RIT.
50. *Liberman, A. M.; Harris, Katherine S.; Hoffman, H. S.* and *Griffith, B. C.:* The discrimination of speech sounds within and across phoneme boundaries. J. exp. Psychol. *54:* 358–368 (1957).
Fry, D. B.; Abramson, A. S.; Eimas, P. D. and *Liberman, A. M.:* The identification and discrimination of synthetic vowels. Language and Speech *5:* 171–189 (1962).
Studdert-Kennedy, M.; Liberman, A. M. and *Stevens, K. N.:* Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. J. acoust. Soc. Amer. *35:* 1900 (A) (1963).
Stevens, K. N.; Öhman, S. E. G. and *Liberman, A. M.:* Identification and discrimination of rounded and unrounded vowels. J. acoust. Soc. Amer. *35:* 1900 (A) (1963).
51. *Lisker, L.; Cooper, F.S.* and *Liberman, A.M.:* The uses of experiment in language description. Word *18:* 82–106 (1962).
52. *Cooper, F.S.:* Research techniques: EMG. Proc. Conference on Communicative Problems in Cleft Palate 1963 (Amer. Speech and Hearing Ass.).
MacNeilage, P.F.: Electromyographic and acoustic study of the production of certain final clusters. J. acoust. Soc. Amer. *35:* 461–463 (1963).
53. *Ladefoged, P.:* Sub-glottal activity during speech. Proc. 4th Int. Congr. Phonetic Sciences, Helsinki 1961, pp. 73–91 (The Hague 1962). – Some physiological parameters in speech. Lang. and Sp. *6:* 109 (1963).
Ladefoged, P. and *McKinney, Norris P.:* Loudness, sound pressure, and subglottal pressure in speech. J. acoust. Soc. Amer. *35:* 454–460 (1963).
54. *Sonesson, B.:* On the anatomy and vibratory pattern of the human vocal folds (Lund 1960).
55. *Harris, Katherine S.; Rosov, R.; Cooper, F. S.* and *Lysaught, Gloria F.:* A multiple suction electrode system. J. Electroenceph. clin. Neurophysiol. (1964).
56. *Ladefoged, P.:* Physiological parameters of speech. J. acoust. Soc. Amer. *36:* 1037 (A) (1964).
Fromkin, Victoria A.: Parameters of lip position. J. acoust. Soc. Amer. *36:* 1037 (A) (1964).
Anthony, J.: True model of the vocal tract. J. acoust. Soc. Amer. *36:* 1037 (A) (1964).
57. *Lieberman, P.:* Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. J. acoust. Soc. Amer. *35:* 344–353 (1963).
Soron, H.I. and *Lieberman, P.:* Some measurements of the glottal-area waveform. J. acoust. Soc. Amer. *35:* 1876–1877 (A) (1963).
58. Personal communication.
59. *Kozhevnikov, V.A.; Kuz'min, Yu.I.* and *Shapliakov, V. S.:* A system of transducers and electronic devices for analog recording and data processing in the study of speech articulation. Med. Electronics and Biol. Engineering *2:* 135–144 (1964).
60. Conference on Communicative Problems in Cleft Palate, Washington, D.C. (Amer. Speech and Hearing Ass. 1963).

Author's address: Dr. Franklin C. Cooper, Haskins Laboratories, 305 E. 43rd Street, *New York 17, N.Y.* (USA).