SPEECH SYNTHESIZERS

FRANKLIN S. COOPER

Although speech synthesizers are not new,[1] it is only within the past ten to fifteen years that they have become important tools for research in phonetics. This is due in part to the sound spectrogram which presents acoustic data as visual patterns that are easy to manipulate conceptually; in part, it is due to an upsurge of research on voice communication, which has provided many of the electronic tools used in synthesis. Thus, as so often happens, research has burgeoned as soon as the necessary tools, both conceptual and instrumental, have become available.

A survey of the new tools for speech synthesis and their application to research in phonetics should answer the following questions: (1) What are the major uses for speech synthesizers? (2) What operating characteristics should they have? (3) What kinds of synthesizers are available, and for what uses are they best adapted? Fortunately, we can take advantage of the excellent and extensive reviews by Eli Fischer-Jørgensen[2] and by C. G. M. Fant[3]; the latter review has been ably supplemented by Fant's presentation at this same session.

Clearly, there is no point in duplicating the factual content of these reviews; moreover, they cover almost all of the first-generation speech synthesizers, and the second-generation devices have yet to appear. There is, therefore, an opportunity – rare for reviewers – to comment on known devices rather than describe new ones. These comments are directed, not to my colleagues who are specialists in acoustic phonetics, but to those who would like to undertake experiments with synthetic speech but are somewhat dismayed by the variety of synthesizers and conflicting accounts of their respective virtues. In thus trying to adopt the point of view of a prospective user, I shall give most attention to the kinds of things that can be *done* with the various synthesizers and to those characteristics of the devices that have most to do with their suitability for one kind of research or another. This is not a very ambitious objective, but it may be a useful one.

[1] Dudley, Homer, and Tarnoczy, T. H., "The Speaking Machine of Wolfgang von Kempelen," *J. Acoust. Soc. Am.*, 22 (1950), pp. 151–166; Dudley, Reisz, and Watkins, "A Synthetic Speaker," *J. Franklin Inst.*, 227 (1939), p. 739; Paget, R., *Human Speech* (London, 1930); Helmholtz, *Sensations of Tone* (1875, Trans. by Ellis).

[2] Fischer-Jørgensen, Eli, "What can the New Techniques of Acoustic Phonetics Contribute to Linguistics?" *Proceedings of the VIII International Congress of Linguistics* (Oslo, 1958), pp. 433–499.

[3] Fant, C. G. M., "Modern Instruments and Methods for Acoustic Studies of Speech," *Proceedings of the VIII International Congress of Linguistics* (Oslo, 1958), pp. 282–362.

THE USES OF SPEECH SYNTHESIZERS

Let us turn first to the uses that can be served by speech synthesizers. In general, the techniques of synthesis complement those of analysis; usually, both are necessary. But, you may ask, is synthesis really necessary if one is interested primarily in a more precise description of speech sounds? Why not rely entirely on the new tools of analysis? Certainly, analysis will be necessary, but a purely descriptive phonetics would be self-defeating, especially when armed with the data-gathering powers of modern instruments. As Fant[4] has commented in his review, "One of the greatest problems in speech analysis is the mass of data to be dealt with." – and again, in speaking of statistical studies of consonant spectra, " – the investigator easily drowns in a sea of details. The major difficulty is to recognize one and the same pattern detail in the spectra from various voices. A formant may for instance be too weak to be observed or merge with adjacent formants. If the investigator is not very well acquainted with possible pattern variations of the visible sound substance, he runs the risk of making errors in the labelling of data which will invalidate the statistics. It is safer to discuss in detail a few samples and wait with the statistics until a reliable specificational frame has been established and can be mastered."

The "reliable specificational frame" is the crux of the matter. In order to establish it, one needs not only skill in dealing with spectrograms, but also information about what aspects of the acoustic pattern are significant carriers of information. Here is where a speech synthesizer is extremely useful, if not, indeed, indispensable. It can be used to convert the spectrum analysis (or a simplified version of it) back into sound for phonetic evaluation by ear. The experimenter can test his hypotheses about significance by *manipulating* the spectrum and *hearing* the result. The speech synthesizer is, if you will, his informant, and it is an extremely versatile one.

The essential point here, as in all of science, is that we must simplify Nature if we are to understand her. More than that: we must somehow choose a *particular* set of simplifying assumptions from the many sets that are possible. The great virtue of speech synthesizers is that they can help us make this choice.

Let us return to the spectral analysis of speech generated by human speakers. There is, of course, a very close relation between spectral data and the events of articulation; moreover, the correlation of the two descriptions is an important area of research. Here, also, speech synthesizers can be of use, though now we wish them to serve as working models of the vocal tract. There would be little point in building a vocal tract analog to duplicate the three dimensional configurations of the human tract; indeed, the purpose of a model is to substitute simple structures for complex ones. The adequacy of the model can, of course, be tested by comparing the synthetic speech it produces and actual speech. Within their limitations, these analog devices can serve also to clarify the purely physical relationships between shape of tract and resulting sound spectrum.

[4] See Ref. 3, pp. 286 and 315.

A third use for speech synthesizers is in the communications industry. There are situations in which the physical or economic constraints on communications channels are severe enough to justify the complexity and cost of analyzing speech at the transmitter and then reconstructing it at the receiver. Intensive research is underway on such equipment, and on phonetic typewriters and voice-controlled devices. The phonetic interest is indirect, but real. On the one hand, most of the techniques for speech synthesis, and a number of the complete devices, originated in such work; moreover, as Eli Fischer-Jørgensen[5] has observed, "What is of importance for linguists is particularly the research preceding the construction of these machines, since the simplification of speech signals presupposes knowledge of what is relevant to speech communication and what is not." On the other hand, much of the technology is not relevant to experimental phonetics; those who are interested will find excellent reviews[6] in the literature.

We have, then, identified three general areas of usefulness for speech synthesizers: these devices can be characterized as versatile informants, which help us to isolate perceptual cues; as articulatory models, which help us to understand the relationships between articulation and sound; and as sophisticated telephones. There are many types of synthesizers and they differ widely in manner of operation. But we are less interested, as experimental phoneticians, in how the synthetic speech is produced by the device than in how the synthesis is controlled by the experimenter. Let us examine, in reverse order, the three classes of synthesizers from the point of view of how synthesis is controlled.

ANALYSIS-SYNTHESIS DEVICES

In the general class of devices used in analysis-synthesis telephony, the control is, of course, via signals extracted from spoken sounds. The Vocoder[7] is a well-known example. We can take advantage of the fact that the sound source and the sound spectrum of its synthetic speech are controlled by separate signals. No doubt most of you have heard a Vocoder used to impose an advertising message on music or such everyday sounds as traffic noise or a steamboat whistle.

Likewise, the voice pitch can be separately controlled. This use of a Vocoder makes it a potent research tool for studies in intonation and stress. The Intonator,[8] as this particular variant has been called, is a normal Vocoder in every respect except that the loop of magnetic tape which carries the original voice recording moves in step with a

[5] See Ref. 2, p. 443.

[6] Proceedings of Seminar on Speech Compression and Processing, Bedford, Mass., 1959 (AFCRC–TR–59–198; vols. 1–2); Fant, C. G. M., and Stevens, K. N., "Systems for Speech Compression," *Fortschritte der Hochfrequenztechnik*, 5 (1960), pp. 229–262.

[7] Dudley, H., "Remarking Speech," *J. Acoust. Soc. Am.*, 11 (1939), pp. 169–177.

[8] Borst, J. M., and Cooper, F. S., "Speech Research Devices Based on a Channel Vocoder," *J. Acoust. Soc. Am.*, 29 (1957), A, p. 777; Borst, J. M., "The Use of Spectrograms for Speech Analysis and Synthesis," *J. Audio Eng. Soc.*, 4 (1956), pp. 14–23.

transparent tape on which the experimenter can draw hill-and-dale pictures of the voice pitch that he wishes to impose on the recorded utterance. (Recording 1: two spoken sentences on which quite different intonation patterns were imposed; the synthetic versions were followed by the original utterance. Recording 2: five words that are minimally distinguished by tone, spoken by a Thai informant; this was followed by a set of synthetic words in which the first word of the spoken set had the five tonal patterns imposed on it[9] by the Intonator.)

A somewhat different application provides a fairly good speech stretcher. The Vocoder, as you know, divides the speech spectrum into a number of narrow frequency bands and, after analysis, represents each of them by a slowly varying voltage. To stretch the speech, it is necessary only to record all these voltages (twenty in the present case) on the separate tracks of a multitrack tape recorder, and then play the recording into the Vocoder synthesizer at reduced tape speed. Each part of the spectrum, and the pitch also, is reproduced just as it would have been at normal speaking rates, but now the time scale has been stretched. (Recording 3: a brief passage played back at normal, half-, and quarter-speed.) The possible applications to phonetic research are obvious.

SYNTHESIZERS AS VERSATILE INFORMANTS

Some of the above uses of equipment designed primarily for analysis-synthesis telephony are hardly distinguishable from the overt use of speech synthesizers as robot informants. The primary difference has been that the experimenter's concern with control signals was limited to a single variable, and the synthesizer took care of itself in all other respects. When the synthesizer is to be used as a general-purpose informant, the control signals pose a much greater problem, since the experimenter must now assume direct responsibility for the entire acoustic spectrum. This would be an intolerable chore if one could not make simplifying assumptions. The nature of these assumptions, and the level of simplicity at which one can operate, depend, of course, on the research objectives, which may cover a very wide range. The point is an important one, even if obvious: there are different kinds of synthesizers just as there are different kinds of research objectives, and one needs to choose the right synthesizer for his particular job; further, this choice will depend primarily on how the synthesis is controlled.

This focusing of attention on the *control* of synthesis, is not intended to imply that one control system is good and another bad; rather, that there is a close relation between the control system and the simplifying assumptions that are built into the machine, and that the machine, in turn, imposes on the research done with it. These relationships can be discussed to best advantage in terms of concrete examples: we

[9] The experimental procedures illustrated by the recording are described by A. S. Abramson, *The Vowels and Tones of Standard Thai*; *Acoustical Measurements and Experiments* (Ph.D. dissertation, Columbia University, 1960).

need to consider for this purpose only a few synthesizers, all with proven capabilities, namely, the Pattern Playback and Voback of Haskins Laboratories, PAT of the Signals Research and Development Establishment and the University of Edinburgh, OVE II of the Royal Institute of Technology, and DAVO of the Massachusetts Institute of Technology. Let me begin with the Pattern Playback.

PATTERN PLAYBACK (PB-2)

Much of the research done by my colleagues at Haskins Laboratories has been concerned with the question, "What is significant in the spectrographic pattern – and what is not?" For this purpose, it has been very convenient to use the spectrogram itself as the control information and to make trial simplifications directly in pattern terms.[10] The Pattern Playback,[11] with which much of this work has been done, scans the spectrographic patterns with a line of light that is modulated at multiples of 120 cycles, and thereby generates a form of monotone speech. A spectrogram of this synthetic speech is esentially identical with the painted control pattern.

The device has its faults, but it has three characteristics that deserve our attention. It is, for one thing, extremely flexible. There are almost no constraints on the kind of pattern that can be painted. The pattern can be very detailed if one wishes to approximate a real spectrogram, or it can be extremely simple with a minimum number of formants, with angular transitions – or what you will. Second, the device is easy to use in a conceptual sense; that is, one can think about speech phenomena in spectrographic terms and manipulate the device in exactly these same terms. Third, it is easy to use in a mechanical sense; the patterns can be painted very simply and quickly at the machine, and they can be heard, revised, and heard again within the minute. A somewhat subtler advantage that we did not fully appreciate at first is that several patterns, up to a total of about ten seconds of speech, are readily available for comparative listening. For all these reasons, we have been willing to forgive the Pattern Playback its rough voice quality, monotone pitch, and fricative sounds that are inclined to "twitter", especially if one scans through the speech in slow motion.

A VOCODER PLAYBACK (VOBACK)

The same kind of control information, namely a spectrographic pattern, can of course be used with other types of sound generating equipment. The photoelectrically

[10] Reviews and references to earlier work are given by P. C. Delattre (elsewhere in this volume); Delattre, P. C., "Les indices acoustiques de la parole: Premier rapport," *Phonetica*, vol. 2 (1958), pp. 108–118, 226–251; Liberman, A. M., *et al.*, "Minimal Rules for Synthesizing Speech," *J. Acoust. Soc. Am.*, 31 (1959), pp. 1490–1499; Liberman, A. M., "Some Results of Research on Speech Perception," *J. Acoust. Soc. Am.*, 29 (1957), pp. 117–123.

[11] Cooper, F. S., Liberman, A. M., Borst, J. M., "The Interconversion of Audible and Visible Patterns as a Basis for Research in the Perception of Speech," *Proceedings of the National Academy of Sciences*, 37 (1951), pp. 318–325; see also Ref. 8.

controlled Vocoder, first described by Schott[12] and developed at Haskins Laboratories as a full-fledged research tool,[13] employs an 18-channel Vocoder to generate the synthetic speech. Spectrographic patterns provide control voltages that modulate the buzz (or hiss) signals flowing in the 18 spectrum channels. Two additional controls are added across the top of the pattern: one determines whether the speech sounds will be voice-like or friction-like, and the other is the hill-and-dale pattern for voice pitch control that was mentioned in connection with the Intonator. Thus, the Vocoder Playback retains the conceptual and instrumental conveniences of the Pattern Playback, though there has been some loss in flexibility, since the spectrum now consists of only 18 discrete zones so that fine adjustments of formant frequencies are not always possible; also, the sound source must change from "voice" to "friction" in all channels at the same instant. In return for these limitations, we have control of voice pitch, realistic frictional sounds, and much less noise behind the speech.

In practice, Voback and the Pattern Playback have been used almost interchangeably as they are in the following recordings. (Recording 4: an example of the same sentence synthesized on the two machines at both normal and slow rates. There is no spoken version for comparison, since the spectrogram was painted by Pierre Delattre solely on the basis of his experience with the acoustic cues for speech sounds.) Other examples of the manipulations of acoustic cues will be given by Delattre in a recorded demonstration, as part of his own paper at a later session.

THE PARAMETRIC ARTIFICIAL TALKER (PAT)

A different kind of simplifying assumption was used by Lawrence[14] in his Parametric Artificial Talker – known to most of you, I am sure, by its nickname and by PAT's now-famous question, "What did you say before that?" The assumption underlying PAT's design is that the significant information in speech is contained in the changing natural frequencies of the vocal resonators. One can, therefore, perform adequate syntheses by controlling the frequencies of the first three formants, together with the kind and amount of excitation (i.e., the intensity and the periodicity of a buzz signal and the intensity of a wide-band noise). In the original PAT, these six parameters were painted as miniature hill-and-dale patterns on a glass slide and were scanned by a cathode-ray tube.

More recent versions of PAT, at Christchurch and Edinburgh, have additional parameters and a more manageable control system consisting of a plastic belt with the parameters plotted in conducting ink. Each parameter has its own zone across the width of the belt, and all are read simultaneously by running the belt through

[12] Schott, L. O., "A Playback for Visible Speech," *Bell Telephone Lab. Rec.* 26 (1948), pp. 333–339.
[13] See Ref. 8.
[14] Lawrence, W., "The Synthesis of Speech from Signals which have a Low Information Rate," *Communication Theory*, ed. W. Jackson (London, 1953).

a "mangle" that serves to convert the wavy lines into dynamic control voltages. This arrangement allows convenient manipulation, both in changing from one belt to another and in revising the control signals by erasure and re-painting. There is some question about the conceptual conveniences of a series of wiggly lines as compared with a spectrographic pattern; the information, to be sure, is basically the same, and there is some visual resemblance, but not very much. There is a distinct loss in flexibility: Frances Ingemann,[15] in describing her experiences with PAT, has commented, "More serious is the difficulty inherent in the parametric approach: All eight parameters, no more and no less, must be represented at all times. For example, even if a formant is not evident on a spectrogram, it must be synthesized. Possible cues such as changes in relative formant amplitude or bandwidth must be ignored. Despite these problems, fairly high quality speech can be obtained." It would be unfair to PAT, and to other synthesizers of the same general type, to enumerate their limitations without calling special attention to the concluding point, namely, that PAT can indeed produce fairly high quality speech.

OVE II

A related device, also of the formant-generator type, is OVE II[16] in the Speech Transmission Laboratory of the Royal Institute of Technology in Stockholm. The simplifying assumptions on which the design is based are essentially the same as for PAT, but with at least one significant addition: the formant generators used to synthesize the vocalics are connected in series (rather than in parallel) to preserve transmission characteristics like those of the human vocal tract during vowel production. There are good reasons why this can provide more natural vowel sounds; there is the further advantage that the relative intensities of the formants are uniquely determined by their frequency positions, so that the buzz intensity control need only be used to manipulate overall intensities. The production of nasals and of fricatives and stops, involves different transmission characteristics in the human vocal tract and is accomplished in essentially separate and parallel synthesizers in OVE II. Thus the conplete synthesizer (OVE II) consists of three parellel subsystems; a total of eight parameters and four gating signals serve to control the synthesis.

It is perhaps worth commenting on the series connection of the formant generators as yet another illustration of the way in which research orientations tend to guide the choice of instrumental characteristics. On the one hand, the series connection copies nature and has advantages[17] if one is principally concerned with naturalness

[15] Ingemann, Frances, "Eight Parameter Speech Synthesis," presented at the meeting of the Acoustical Society of America in San Francisco, Oct. 20, 1960; included in progress report from the Phonetics Dept., Univ. of Edinburgh, Sept.–Dec., 1960.
[16] Fant, C. G. M., "Modern Instruments and Methods for Acoustic Studies of Speech," *Proceedings of the VIII International Congress of Linguistics* (Oslo, 1958), pp. 282–358.
[17] In this connection, Fant has observed (personal communication), "... that the OVE II control

and fewer controls for formant intensities; on the other hand, if one needs maximum flexibility, as in the study of acoustic cues for perception, he would wish to retain independent control of the separate formant intensities. This, and engineering convenience, account for the use of parallel-connected filters in several synthesizers of this general type. For comparable reasons, the factors of convenience are less important in studies largely aimed at improving the naturalness of synthetic speech, since one must, at least at this stage of our understanding of the factors affecting naturalness, work from careful analytic measurements of real speech spectra. Hence, a larger number of control functions can be tolerated, and the control tapes need not be so readily interchanged.

Dr. Fant has very kindly provided a tape recording illustrating the capabilities of OVE II when it is supplied with control parameters based on the analysis of real speech. He recently gave a remarkable demonstration[18] in which many in his audience had difficulty in deciding which was the real speech and which the synthetic. The last sentence of the tape you will hear is from that demonstration. (Recording 5: short sentences as spoken and as synthesized on OVE II.) Now, as Fant would no doubt agree, it is one thing to generate "natural" speech from data about real speech, but quite something else to give a simple description of the essential relationships between control parameters and perceptual effects. Much patient research lies ahead before this objective can be realized, but at least an adequate tool is at hand, and in use.

SYNTHESIZERS AS ARTICULATORY MODELS

We have considered at some length the role of speech synthesizers in studying the relationships between acoustic signal and perceptual effect. Let us turn briefly to the use of synthesizers as models of the articulatory apparatus. We find again that simplifying assumptions are needed, but now they take the form of geometrical approximations to the shape of the human tract. True, the geometry may be realized in

parameters are in several respects related to speech production, e.g., the separate nasal and fricative system. Control parameter changes of, for instance, the frequency F_1 have a complicated effect as far as the spectrum is concerned (due to the change in spectrum levels). The effect is simple, however, when translated to the articulatory gestures (opening *vs* closing of the vocal tract) as in the transition from a voiced consonant to a vowel or vice versa.

When several spectrum variables have a conditioned variation in human speech it is of course valuable to be able to make use of this redundancy in synthesis experiments devoted to an investigation of what shifts in linguistic responses are evoked by changes along these natural parameters. (A shift in F_1-position within a consonant in intervocalic position would for example correspond to a shift from b to v to w.) Once the investigator is wholly aware of what he is doing, this is a valid procedure for investigating the perceptional as well as linguistic significance of spectral patterns observed from spectrograms."

[18] The recorded demonstration was part of a paper given at the meeting of the Acoustical Society of America in Philadelphia, May 10, 1961; for abstract see C. G. M. Fant, *et al.*, "Recent Progress in Formant Synthesis of Connected Speech," *J. Acoust. Soc. Am.*, 33 (1961), pp. 834–5.

terms of an electrical transmission line, as it is in the electrical vocal tract[19] in Fant's laboratory, and in the closely related devices[20] at the Massachusetts Institute of Technology and the Bell Telephone Laboratories. Stevens and House,[21] in a notable series of studies, have examined the consequences of simplifying the model to the point that only three parameters are required to describe the vocal tract: the back-to-front position of the tongue hump, the radius of the constriction at this point, and the area-to-length ratio for the constriction at the lips.

It was a general limitation of the first-generation vocal tract synthesizers that they could only deal with static configurations. Recently, a dynamic vocal tract analog (DAVO) was developed by Rosen,[22] in Stevens' laboratory at MIT. The individual electrical sections of this tract can have their equivalent geometries changed by control voltages and these, in turn, can be programmed to produce short dynamic sequences such as consonant-vowel syllables. Still more recently, a "nose" has been added to this synthesizer so that the combination of DAVO and DANA can also produce nasal consonants.[23] The use of such a synthesizer has much to commend it in phonetic studies concerned with the relations between speech sound and articulatory gesture, since the control of synthesis can be directly in terms of articulatory shape and movement. The voice quality is good, and it should be possible to further develop the control system for more convenient manipulation.

SUMMARY OF USES AND CONTROL CHARACTERISTICS OF SYNTHESIZERS

Let us summarize the discussion thus far: Speech synthesizers often serve to complement the techniques of analysis and, for some purposes, synthesis is indispensable. We are, as phoneticians, not very directly concerned with the use of speech synthesizers in telephony, but we can often use the techniques of the communications engineer in our research. We have two principal uses for synthesizers: as versatile informants and as models of the articulatory tract. In their role as informants, synthesizers can help us isolate the acoustic correlates of the perceived speech event. This may involve, at one extreme of abstraction, a search for the major cues for the perception of linguistically significant units; at the other extreme, it may be a study

[19] Fant, C. G. M., *Acoustic Theory of Speech Production*, pp. 79–87 ('s-Gravenhage, Mouton and Co., 1960); also see description in *Proceedings of the VIII International Congress of Linguistics* (Oslo, 1958), pp. 351–3

[20] Stevens, K. N., Kasowski, S., Fant, C. G. M., "An Electrical Analog of the Vocal Tract," *J. Acoust. Soc. Am.*, 27 (1953), pp. 734–742; Dunn, H. K., "The Calculation of Vowel Resonances and an Electrical Vocal Tract," *J. Acoust. Soc. Am.*, 22 (1950), pp. 740–753.

[21] Stevens, K. N., House, A. S., "Development of a Quantitative Description of Vowel Articulation," *J. Acoust. Soc. Am.*, 27 (1955), pp. 484–493; Stevens, K. N., House, A. S., "Studies of Formant Transitions Using a Vocal Tract Analog," *J. Acoust. Soc. Am.*, 28 (1956), pp. 578–585.

[22] Rosen, G., "A Dynamic Analog Speech Synthesizer," *J. Acoust. Soc. Am.*, 30 (1958), pp. 201–209.

[23] Hecker, M. H. L. (to be published, *J. Acoust. Soc. Am.*, Feb. 1962).

of the minimal parametric description necessary for naturalness. There is, between these extremes, much unexplored territory.

In general, these different uses are best served by different synthesizers. The phonetician, in choosing a synthesizer, should give principal attention to how the synthesis is controlled. He will, in every case, need conceptual convenience. For exploratory studies, he will also be concerned with flexibility of control and with instrumental convenience; for studies aimed at improving the naturalness of synthetic speech, he will require, of necessity, a synthesizer that is capable of high quality speech, though he may sometimes lament the effort needed to obtain this quality synthetically. Finally, certain specific synthesizers were described briefly as representative of the types that are now available.

NEWER DEVELOPMENTS AND PROSPECTS

Perhaps "available" is not the right word, since there are only a few research centers that are equipped with synthesizers and the necessary technical staffs to service them. What, you may ask, are the prospects for a small, simple, and readily available synthesizer that can serve, perhaps, as a companion piece to the sound spectrograph? It should be possible to adapt some of the existing types to this purpose, if a demand existed, but I know of no such undertaking. Indeed, the present trend seems to be in the opposite direction.

Thus, although the second generation of speech synthesizers has only begun to appear, there is a clear trend toward the use of digital techniques and large general-purpose computers. The Bell Telephone Laboratories have done much of the pioneering work in using digital computers to process speech signals. This work is largely concerned with communications problems, and most of the processing has been applied to the speech waveform. Of more direct relevance to research in phonetics is a recent demonstration by Kelly and Gerstman[24] of their first efforts in using a computer to synthesize speech from a phonetic transcription. The basis for this synthesis is related in part to the "minimal rules" procedures described by Liberman,[25] though more closely to the "programming" used in an early synthesizer known as Octopus.[26] The synthesis techniques described by Kelly and Gerstman appear to be very flexible and will, no doubt, be used to test various simplifying assumptions about the acoustic structure of speech.

K. N. Stevens[27] has discussed a procedure for analysis-by-synthesis that has already given important results and has far-reaching possibilities for the future, though the evolution of a general-purpose synthesizer from these studies is yet to come. Intensive use of a general-purpose digital computer is being made also by the group with Stevens at MIT.

C. P. Smith[28] at the Air Force Cambridge Research Laboratory has, near completion, a large, special-purpose digital equipment with full facilities for manipulating speech spectra in digital terms and for listening to the results immediately. This device offers exciting possibilities, either for making gross changes in the digital spectrogram or for making detailed changes under such close control that one might appropriately refer to the procedure as "microsurgery".

The prospects for the future are indeed promising but a word of caution may be in order. However glamorous the tools, the facility with which they can be used – and the research objectives to which they can be applied – will be determined very largely by the nature of the simplifying assumptions that the experimenter chooses to make about the speech process.

Haskins Laboratories
New York City

[24] Kelly, J. L., Jr., and Gerstman, L. J., "An Artificial Talker Driven from a Phonetic Input," *J. Acoust. Soc. Am.*, 33 (1961), A, p. 835.

[25] Liberman, A. M., *et al.*, "Minimal Rules for Synthesizing Speech," *J. Acoust. Soc. Am.*, 31 (1959), pp. 1490–1499.

[26] See Ref. 8.

[27] Stevens, K. N., "Toward a Model for Speech Recognition," *J. Acoust. Soc. Am.*, 32 (1960), pp. 47–55.

[28] Smith, C. P., "An Approach to Speech Bandwidth Compression," *Proceedings of Seminar on Speech Compression and Processing*, Air Force Cambridge Research Center, Vol. II (1959).