SPEECH SYNTHESIS BY RULE

J. N. HOLMES, IGNATIUS G. MATTINGLY and J. N. SHEARME

Joint Speech Research Unit, Eastcote

SPEECH SYNTHESIS BY RULE

J. N. Holmes, Ignatius G. Mattingly and J. N. Shearme
Joint Speech Research Unit, Eastcote

Intelligible English has been synthesized in part " by rule ", with a system consisting of a computer programme and an electronic analogue synthesizer. Methods for the synthesis of the various phonetic classes and their transitions are described.

We have recently made a start towards the synthesis of speech " by rule ". In synthesis by rule, the input to the synthesizing system is a phonemic transcription of an utterance. Ideally, no subphonemic information need be included, because all such information has been embodied in the design of the system itself. Moreover, the criterion of success for a synthesized utterance is not whether it is a faithful reproduction of an actual utterance of a native speaker of the dialect under study, but whether it is acceptable to such a speaker as a possible utterance in his dialect.

The system to be discussed here consists of a computer programme which, given an input " sentence ", calculates appropriate values for nine parameters for each 10 msec. unit of time, and an electronic analogue synthesizer which produces the sounds specified by successive sets of parameter values. The calculation of the parameter values could be, and in the early stages of the project was, done by hand, but it is a laborious task, and the use of a computer programme means that much larger amounts of speech can be synthesized than would otherwise be practical. It would also have been possible to dispense with the analogue synthesizer, and to embody virtually the entire system in a computer programme; in fact Kelly and Gerstman, who were the first investigators to synthesize speech by rule with a computer, used such a programme (Kelly, 1961, 1962). But the analogue synthesizer makes it possible to get along with a much less powerful computer than would otherwise be needed.

The synthesizer

Fig. 1 is a block diagram of the synthesizer. The excitation of the vocal cords is simulated by a variable-frequency pulse generator; turbulent excitation by a random-noise generator. The output of one of the two generators is selected by a switch and passed through three parallel resonant circuits of variable frequency, which correspond to the first three formants of the vocal tract. Energy of higher frequencies is represented by a fourth resonant circuit fixed at 3,500 c.p.s., if the pulse generator has been selected; or by a band-pass filter at 3,600 - 4,000 c.p.s., if the noise generator has been selected. The amplitudes of the inputs to the variable resonant circuits and the high-frequency circuits are independently variable. The sum of the outputs of these circuits is the output of the synthesizer.

To control the synthesizer, it is necessary to specify, for each time-unit, the rate of the pulse generator, F_0; the condition of the switch S, which selects the generator

18	(	A		ə
19	?	B		b
20		ER	3(20 21 21 22 22 23 4(	ɜ
20		D		d
19	(	I		ɪ
17		N		n
16		DH		ð
16	(	A		ə
18		H		h
20	%%%	AA		æ
12	%%%	N		n
12		D		d
12	(	I		ɪ
13		Z		z
15		W		w
14	(	ER		ɜ
12	=	TH	10 8 2(	θ
12		T		t
16		UU		u
13	(	I		ɪ
13		N		n
14		DH		ð
13	(	A		ə
13	?	B		b
10	%%%%	OO		ʊ
2	%%%%	SH		ʃ
2		END		

Fig. 2. An input sentence. The phonetic elements are represented by literal symbols. The left-hand column gives the F_0 values. Special F_0 values for the element [ER] are specified between the computer trigger symbols "3(" and "4(". The other symbols are modifiers. A conventional transcription of the sentence is given at right.

values; and certain auxiliary "modifier" symbols. A phonetic element is the operational equivalent of a phoneme, and is the basic input unit for the programme. The literal symbols, which are used because conventional phonetic symbols are impractical for computer input, approximate ordinary spelling. The series of F_0 values is derived by measuring a spectrogram of a human version of the utterance to be synthesized, or is simply estimated. A typical input sentence is shown in Fig. 2.

The other input is a collection of tables, one for each distinct element of the ensemble required to synthesize a particular dialect. Entered in the input table for an element are its duration in 10 msec. time-units, the values of the parameters (other than F_0) during its steady-state period, and the specifications for transitions to and from this steady state. The information included in the input tables for the Received Pronunciation of English is summarized in Appendix 1. In compiling the first version of the tables, spectrogram measurements were used, as well as a great deal of invaluable data from the various articles listed as references. The tables have since been substantially modified as evaluation of synthesized utterances has suggested, and will doubtless be further modified in future.

It will be obvious from Appendix 1 that the correspondence between phonemes and phonetic elements is only approximate. In some cases a sequence of two or three elements is needed to represent a single phoneme, and allophones of one phoneme may be represented by distinct elements. There are three elements which do not correspond to phonemes at all : [Q] which is simply a 100 msec. period of silence ; [END], which signals the end of an input sentence ; and [QQ], an element of zero duration whose use in the synthesis of voiced fricatives will be discussed later. And of course, the segmentation of speech implied by the assignment of definite durations to the elements is just a useful programming fiction.

Each cycle of the programme generates the successive sets of parameter values needed to synthesize one element, the number of such sets being equal to the duration of the element. The parameter values (except those for F_0) depend upon the information entered in the input tables for that element and for the immediately preceding and following elements in the input sentence.

The condition of the switch S is given in the input table for the current element, and remains the same for the duration of that element.

The fundamental frequency F_0 is normally specified in the input sentence once for each element. This value is the value of F_0 at the boundary between the current element and the preceding element. The programme determines the values of F_0 between successive boundary values by linear interpolation. Since the design of the synthesizer does not permit the omission of parameter values, F_0 is always calculated regardless of the condition of S. Any F_0 contour which approximates a straight line for the duration of an element can be synthesized by a proper selection of boundary values. Other contours can be synthesized by means of a programme option which permits the F_0 values for some or all time-units of an element to be specified explicitly in the input sentence.

The values for each of the remaining seven parameters, F_1, F_2, F_3, A_1, A_2, A_3, A_{HF}, must in general be determined for an initial transition, a steady-state period and a final transition. The steady-state value is in the input table and can be read out as required, but since it is not practical to tabulate the values for every possible transition, these must be calculated.

To simplify this task, a " rank " between 1 and 31 is assigned to each element and entered in its input table. This rank is high if the transitions of the corresponding phoneme are characteristic of the phoneme itself ; it is low if the transitions depend upon the character of the adjacent phonemes. Thus stop consonants rank high, vowels low, and the other phonemic elements somewhere between these extremes. The highest

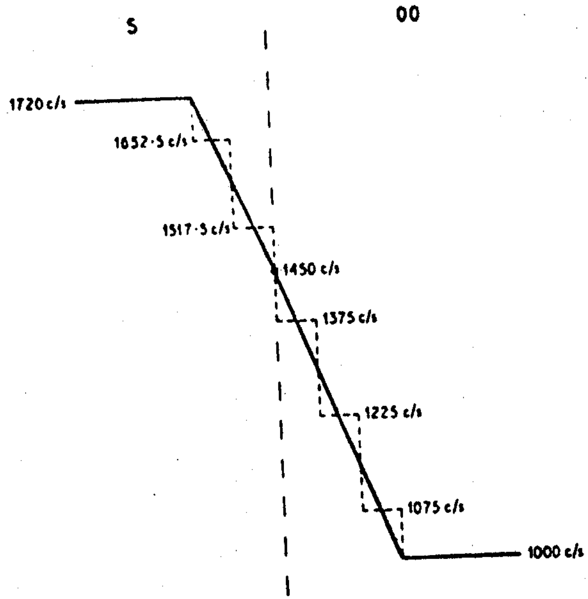

Fig. 3. F_2 parameter transitions for the sequence [S OO]. The ideal transition, represented by the solid line, is approximated by the series of time samples, represented by the dotted line. The vertical dashed line represents the boundary between the two elements.

ranks are assigned to the non-phonemic elements [Q, QQ, END]. The assigned ranks are deliberately spread over the available range so as to leave gaps which can be filled later if the present rank order is revised.

Of two elements which are adjacent in the input sentence, the one of higher rank is "dominant"; the entries in its input table control the character of the transitions on both sides of the boundary between the elements. If the two elements are of equal rank the first element is dominant (with this exception the order of the elements does not affect the transitions; for the sequence of elements *a b a*, the *a b* transitions are symmetrical with the *b a* transitions). This ranking of the elements normally results in acceptable transitions, but there are a minority of cases where it does not, and the programme must eventually be modified to recognize those cases and treat them specially.

In the input table for the dominant element are entered, for each of the seven parameters, the duration in time units of the transition during that element (the "internal" transition); the duration of the transition during the adjacent element (the "external" transition); a quantity called the "fixed contribution"; and the proportion of the steady-state value of the adjacent element which is added to the fixed contribution to get the value of the parameter at the boundary between the

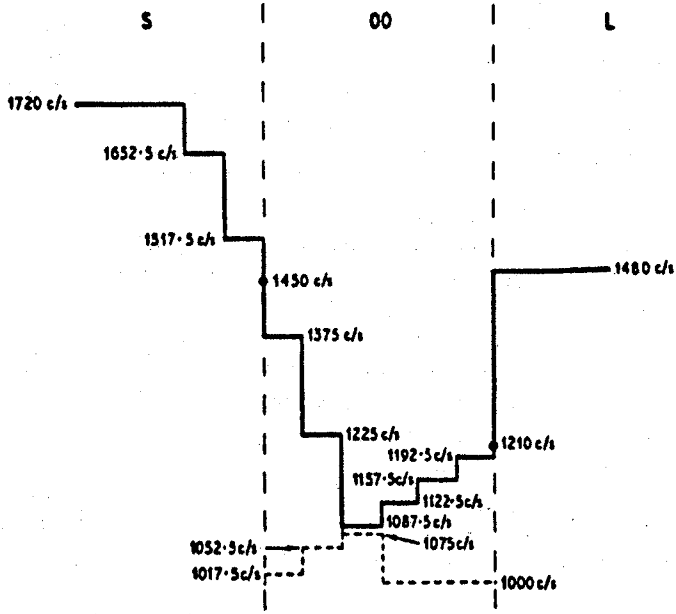

Fig. 4. F_2 parameter transitions for the sequence [S OO L], showing the intersection of the transitions for [S OO] and for [OO L]. The solid line represents the transition actually used; the dotted lines represent values calculated but later discarded. The vertical dashed lines represent boundaries between adjacent elements.

two elements. The values for the parameter during a transition are calculated by linear interpolation between the boundary values and the steady-state values.

With the help of Appendix 1 and Fig. 3 the method of calculating transitions can be illustrated. Suppose it is required to calculate the F_2 transition for the sequence [S OO] representing /su/. The steady-state values, in cycles per second, are 1720 for [S] and 1000 for [OO]. Since [S] has a rank of 18 and [OO] a rank of 2, [S] is dominant. The fixed contribution for [S] is 950; the proportion of the steady-state value of [OO] to be added is 0.5; the boundary value is therefore 1450. The duration of the internal transition is 2 time-units; the duration of the external transition is 3 time-units. Interpolation from the steady-state value of [S] to the boundary value, and thence to the steady-state value of [OO], gives the internal transition 1652.5, 1517.5 as the final transition of [S] followed by the external transition 1375, 1225, 1075 as the initial transition for [OO] (the effects of quantization and rounding at various stages of the programme are neglected in this example).

If the sum of the durations of the initial and final transitions which have been so calculated is less than the duration of the element, the steady-state value entered in the input table for the element is used for the time-units between the two transitions. If the sum equals the duration of the element, the initial transition is followed immediately by the final transition. It can easily happen, however, that this sum is greater than the duration of the element. In this case, if the paths of the two transitions do not intersect, they are discarded, and the actual values to be used for the entire duration of the element are calculated by linear interpolation between the initial and final boundary values. If the paths of the transitions do intersect, the values for time-units between the boundaries and the intersection are used and the others discarded. For example, suppose the initial and final F_2 transitions for [OO] in the sequence [S OO L], representing /sul/, have been calculated (see Fig. 4). The initial transition, as already shown, is 1375, 1225, 1075. The final transition, determined by the dominant [L] is 6 time-units long : 1017.5, 1052.5, 1087.5, 1122.5, 1157.5, 1192.5. The duration of [OO] is 6 time-units, which is less than the sum of the duration of the two transitions. The transitions intersect between the second and third time-units of the element. Therefore, the last value of the initial transition and the first two values of the final transition are discarded, and the F_2 values during [OO] are actually 1375, 1225, 1087.5, 1122.5, 1157.5, 1192.5. The steady-state value 1000 is approached but not attained. It is suggested that the transitions and steady states of natural speech behave in a way not unlike this (Shearme, 1961 ; Lindblom, 1963).

The programme will construct a considerable variety of transitions, depending upon the values entered in the input tables. The paths of the amplitude and frequency transitions between two vowels are straight lines, for the external and internal transition durations are equal and the boundary value is the average of the steady-state values of the dominant first vowel and the second vowel. The quality of [H] (representing /h/), is heavily modified by the adjacent vowel, so [H] has external transitions of zero length and internal transitions which are nearly as long as its overall duration. The external F_2 transitions for a stop or a nasal terminate at a boundary value which is the average of the " locus " (Delattre, 1955) characterizing the element and the steady-state value of the adjacent element. Finally, there are many cases, such as the amplitude transitions of stops, nasals and fricatives, in which the durations of internal and external transitions are zero, resulting in a simple discontinuity at the boundary point between the steady-state values of the dominant and adjacent elements.

It has already been mentioned that some phonemes are treated as sequences of elements. For example, a diphthong is represented by a sequence of two vowel-like elements ; in the input table for the first, which dominates the second, a long transition is specified. A stop is represented by a sequence of three elements corresponding to the period of closure, the period of high energy at the beginning of the release, and the period of diminished energy at the end of the release. The second element dominates the first and third and specifies amplitude and frequency transitions of zero duration. Thus the input tables for the first and third elements control the transitions into and out of the stops without causing internal complications. The voiced fricatives are treated as three-element sequences in an attempt to synthesize these sounds convincingly even though the noise and pulse generators cannot operate simultaneously. The first element, accounting for half the total duration, is always voiced. The second element is of zero duration, and like the second element of a stop, inhibits transitions between the first and third elements, which it dominates. The third element is devoiced by a special routine of the programme if the following element is voiceless : otherwise it is identical to the first element. The voiceless affricate /tʃ/ is represented by a two-element sequence. The table entries for the first are the same as those for [T], the first element of the sequence representing /t/ ; the entries for the second are the same as those for [SH], which represents /ʃ/. Both elements, however, have been somewhat shortened. The voiced affricate /dʒ/ is represented by a four-element sequence constructed in the same way from [D] and the fricative sequence [ZH, QQ, ZI].

An entry in the input table for the first element of each of these sequences indicates that one or more supplementary elements is to follow. Only this first element need be included in the input sentence ; the programme inserts the others automatically. Element sequences are bracketed in Appendix I.

Synthesis by rule of the utterance

/ə bɜd ɪn ðə hænd ɪz wɜθ tu ɪn ðə bʊʃ/

will demonstrate the operation of the system. The input sentence for this utterance is shown in Fig. 2. Element symbols are in the central column ; the associated F_0 values are in the left-hand column. The symbols between these two columns are modifiers. A modifier alters certain table entries for an element for the particular occurrence of the element with which the modifier is associated. Modifiers can be used to take care of allophonic variations, to test proposed changes in the input tables, and to make any desired *ad hoc* variations. In this sentence, alternative durations for vowels in unstressed position are selected by the modifier " (". The [D] (representing /d/) at the end of [H AA N D] requires a lengthening of the preceding vowel and nasal ; this is accomplished by the " % " modifiers, each of which lengthens the associated element by two time-units. The last two elements of [B OO SH] are lengthened in the same way because they are in sentence-final position. The release for [B] is not entirely satisfactory and it has been eliminated in the synthesis of this sentence by the use of the " ? " modifier, which inhibits the insertion of the second and third elements of the [B] sequence. Any of these modifications could also have been accomplished somewhat less conveniently by the use of the " = " modifier, which indicates that the element symbol is followed by data for a temporary change of one or more tables entries. Here this modifier is used to shorten the [TH] (representing /θ/) in [W ER TH]. The [TH] is followed by " 10 ", specifying the position in the table where the duration of the element is stored ; " 8 ", the new value for the duration ; and " 2(", a computer trigger symbol indicating that there are no more table entry changes. For the first six time-units of the [ER] (representing /ɜ/) in [B ER D], special F_0 values are given between the two computer triggers " 3(" and " 4(".

A spectrographic display of the output produced from this input is shown in Fig. 5b. This display was prepared from the synthesizer control tape by an auxiliary

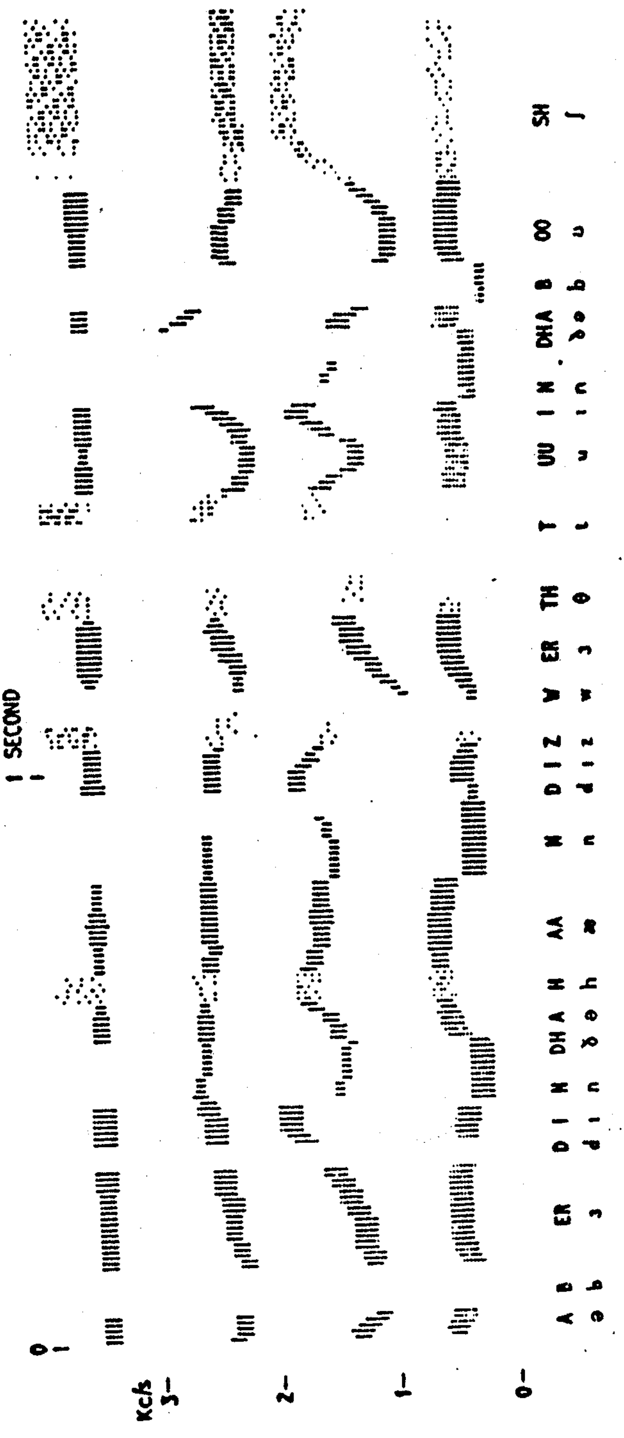

Fig. 5a. Spectrographic display of directly synthesized utterance.

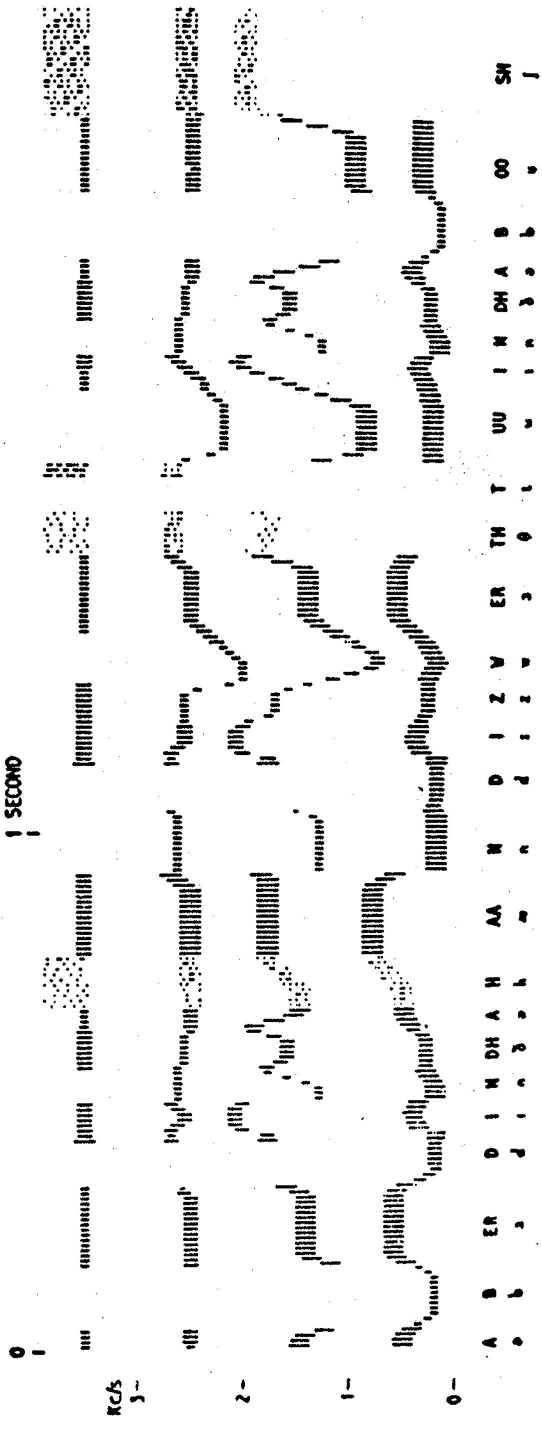

Fig. 5b. Spectrographic display of utterance synthesized by rule.

computer programme, and is similar to a conventional wide-band spectrogram. Noise generator excitation is represented by randomly-spaced dots ; pulse generator excitation by vertical lines (the display does not, however, indicate the value of F_0). The length of the lines and the density of the dots increase with formant amplitude.

Fig. 5b may be compared with Fig. 5a, which displays in similar fashion the same utterance, directly synthesized : its parameter values, instead of being calculated by the computer programme, were taken from measurements on spectrograms of a human utterance of the sentence. Both utterances exhibit certain features resulting from the design of the synthesizer which are not found in natural speech : the forced choice between the two kinds of excitation, the use of a fixed resonance to represent the fourth and higher formants, and the lack of proper nasal formants. These features aside, comparison of Figs. 5a and 5b illustrates the similarities and differences between natural speech and speech synthesized by rule. The timing of the two utterances is quite similar, though of course the directly synthesized utterance is not as self-consistent, as can be seen by comparing occurrences of the same sound, e.g. [B] (representing /b/) in [B ER D] and [B OO SH] or [D] (representing /d/) in [B ER D] and [H AA N D]. There are also interesting differences in the treatment of voiced fricatives in the two utterances. [DH] (representing /ð/) is quite explicit in both of its occurrences in the version synthesized by rule, but in the directly synthesized version the only obvious acoustic correlate is the lengthening of the preceding [N]. And the [Z] (representing /z/ of [I Z]), entirely voiced in the version synthesized by rule, consists, in the directly synthesized version, of a voiced period followed by an unvoiced period which decays to silence. The two versions differ most strikingly, however, in the behaviour of the steady states and transitions of the vowel formant frequencies. In the version synthesized by rule the steady-state values tend to be further away from the central or schwa values, the duration of the steady states tends to be longer, and the transitions consequently more abrupt. As a result this version, while perfectly intelligible, is not as natural-sounding as the directly synthesized version. The reason for this situation is that in general the steady-state values in the input tables derive ultimately from experiments with isolated words, or words spoken in special contexts. In running speech the formant frequencies tend to be centralized owing to the inertia of the articulators (Shearme, 1961). Centralization of the formant frequency values in future versions of the tables should increase the naturalness of the output.

The synthesis system has been used to produce a variety of material including simple nonsense syllables, lists of words, and ordinary English sentences such as the one just discussed. Any English utterance can be synthesized. The output of the system is usually clearly intelligible even without *ad hoc* modifications of the tables.

Conclusions

Substantial progress towards successful synthesis by rule has been made. An electronic synthesizer intrinsically capable of producing high-quality speech has been constructed. A computer programme has been written to perform the laborious task of calculating the parameter values which control the synthesizer, making it practical to synthesize relatively large amounts of speech. This programme embodies a technique for calculating parameter transitions which is sufficiently versatile to deal with the various types of transition which occur in human utterances. The phonetic information required by the programme is stored in a separate set of input tables which can be quickly and easily revised as critical listening suggests. In this way tables have been developed for English which are thought to be substantially correct, though no doubt some of the current table entries can be considerably improved upon.

Two problems, however, remain. First, the present system does not deal adequately with allophones in systematic variation. They are either ignored entirely, or taken care of by using modifiers, or represented by distinct elements. Secondly, the system includes no rules for translating the stress and intonation patterns of language into variations of fundamental frequency, amplitude and duration. F_0 values are copied directly from human speech, and amplitude and duration are fixed for each element unless modifiers are used. If practical rules can be written for allophonic variation and for stress and intonation, something close to genuine synthesis by rule will have been achieved. Work along these lines is in progress.

References

Delattre, P. C., Liberman, A. M. and Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *J. acoust. Soc. Amer.*, 27, 769.

Fant, G. (1959). Acoustic analysis and synthesis of speech with applications to Swedish (Stockholm).

Harris, K. S., Hoffman, H. S., Liberman, A. M., Delattre, P. C. and Cooper, F. S. (1958). Effect of third-formant transitions on the perception of the voiced stop consonants *J. acoust. Soc. Amer.*, 30, 122.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1, 1.

House, A. and Fairbanks, G. (1953). Influence of consonant environment upon the secondary acoustic characteristics of vowels. *J. acoust. Soc. Amer.*, 25, 105.

House, A. (1961). On vowel duration in English. *J. acoust. Soc. Amer.*, 33, 1174.

Hughes, G. W. and Halle, M. (1956). Spectral properties of fricative consonants. *J. acoust. Soc. Amer.*, 28, 303.

Ingemann, F. and Anthony, J. (1960). The specification of speech sounds by means of acoustic parameters (Phonetics Department, University of Edinburgh).

Jassem, W. (1962). The formant patterns of fricative consonants. *Quarterly Progress and Status Report* 3/1962 (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm).

Kelly, J. L. and Gerstman, L. J. (1961). An artificial talker driven from a phonetic input. *J. acoust. Soc. Amer.*, 33, 835 (A).

Kelly, J. L. and Lochbaum, C. (1962). Speech synthesis. *Speech Communication Seminar, Stockholm*, Session F.

Liberman, A. M., Delattre, P. C., Cooper, F. S. and Gerstman, L. J. (1954). The rôle of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychol. Monogr.*, 68, No. 8, 1.

LIBERMAN, A. M., INGEMANN, F., LISKER, L., DELATTRE, P. and COOPER, F. S. (1959). Minimal rules for synthesizing speech. *J. acoust. Soc. Amer.*, 31, 1490.
LINDBLOM, B. (1963). On vowel reduction (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm).
LISKER, L. (1957). Minimal cues for separating /w, r, l, y/ in intervocalic position. *Word*, 13, 256.
LISKER, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, 33, 42.
O'CONNOR, J. D., GERSTMAN, L. J., LIBERMAN, A. M., DELATTRE, P. C. and COOPER, F. S. (1957). Acoustic cues for the perception of initial /w, j, r, l/ in English. *Word*, 13, 24.
PETERSON, G. E. and LEHISTE, I. (1960). Duration of syllable nuclei in English. *J. acoust. Soc. Amer.*, 32, 693.
POTTER, R. K., KOPP, G. A. and GREEN, H. C. (1947). Visible Speech (New York).
SHEARME, J. N. and HOLMES, J. N. (1961). An experimental study of the classification of sounds in continuous speech according to their distribution in the formant 1-formant 2 plane. *Proc. IVth Int. Cong. of Phonetic Sciences* (Helsinki), 234.

APPENDIX 1

The information entered in the input tables is summarized below. The key to the column numbers is as follows :

1. Phoneme symbol.
2. Element symbol. Brackets indicate the correspondence of a single phoneme to a sequence of elements.
3. The rank of the element.
4. The standard duration of the element.
5. The duration of the element in unstressed position (vowels only).
6. The condition of the switch S which selects the noise generator ($S = 0$) or the pulse generator ($S = 1$).

7, 9, 11. The frequencies of F_1, F_2, F_3 respectively.

8, 10, 12. The fixed contribution to the boundary value for F_1, F_2, F_3 respectively.

13, 14. The proportion of the steady-state value of the adjacent element which is added to the fixed contribution to derive the boundary value for F_1 and F_2, and for F_3, respectively.

15. The duration of the external transitions of F_1, F_2, F_3.

16. The duration of the internal transitions of F_1, F_2, F_3.

17, 19, 21, 23. The amplitudes in decibels of A_1, A_2, A_3, A_{HF}, respectively.

18, 20, 22, 24. The fixed contribution to the boundary value for A_1, A_2, A_3, A_{HF}, respectively.

25. The proportion of the steady-state value of the adjacent element which is added to the fixed contribution to derive the boundary value for A_1, A_2, A_3, A_{HF}.

26. The duration of the external transitions of A_1, A_2, A_3, A_{HF}.

27. The duration of the internal transitions of A_1, A_2, A_3, A_{HF}.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
	Q	29	10		0									0	0	-∞	-10.5	-∞	-10.5	-∞	-10.5	-∞	-10.5	1	3	0
	END	31	80		0									0	0	-∞	-10.5	-∞	-10.5	-∞	-10.5	-∞	-10.5	1	3	0
/p/	P	23	8		0	190	110	760	350	2500	0	.5	1	2^1	2	-∞		-∞		-∞		-∞			0	0
	PY	29	1		0	190		760		2500				0	0	24.5		49		43.75		38.5			0	0
	PZ	23	2		0	190	110	760	350	2500	0	.5	1	2^1	2	24.5		38.5		33.25		28			0	0
/t/	T	23	6		0	190	110	1780	950	2680	2680	.5	0	2	2	-∞		-∞		-∞		-∞			0	0
	TY	29	1		0	190		1780		2680				0	0	-∞		-∞		38.5		50.75			0	0
	TZ	23	2		0	190	110	1780	950	2680	2680	.5	0	2	1^1	-∞		-∞		28		40.25			0	0
/k/	K	23	8		0	190	110	1480	1550	2620	1580	.5	.5	3	3	-∞		-∞		-∞		-∞			0	0
	KY	29	1		0	190		1480		2620				0	0	-∞		50.75		50.75		29.75			0	0
	KZ	23	4		0	190	110	1480	1550	2620	1580	.5	.5	3	3	-∞		40.25		40.25		19.25			0	0
/b/	B	26	12		1	190	110	760	350	2500	0	.5	1	2^1	2	24.5		-∞		-∞		-∞			0	0
	BY	29	1		1	190		760		2500				0	0	24.5		49		43.75		38.5			0	0
	BZ	26	0				110		350		0	.5	1	2^1	0	-∞		-∞		-∞		-∞			0	0
/d/	D	26	8		1	190	110	1780	950	2680	2680	.5	0	2	2	31.5		-∞		-∞		-∞			0	0
	DY	29	1		1	190		1780		2680				0	0	38.5		38.5		35		45.5			0	0
	DZ	26	1		1	190	110	1780	950	2680	2680	.5	0	2	0	38.5		28		24.5		35			0	0
/g/	G	26	12		1	190	110	1480	1550	2620	1580	.5	.5	3	3	35		-∞		-∞		-∞			0	0
	GY	29	1		1	190		1480		2620				0	0	35		45.5		40.25		24.5			0	0
	GZ	26	2		1	190	110	1480	1550	2620	1580	.5	.5	3	2	35		35		29.75		14			0	0
/m/	M	15	6		1	190	110	1000	350	2200	0	.5	1	3	0	45.5		19.25		26.25		-∞			0	0
/n/	N	15	6		1	190	110	1300	950	2620	2680	.5	0	3	0	45.5		19.25		26.25		15.75			0	0
/ŋ/	NG	15	6		1	310	110	820	1550	2800	1580	.5	.5	3	0	50.75		33.25		28		26.25			0	0
/f/	F	18	12		0	400	170	1420	350	2560	980	.5	.5	3	2	-∞		42		36.75		35			0	0
/θ/	TH	18	15		0	400	170	1780	1190	2680	2680	.5	0	3	2	-∞		26.25		28		22.75			0	0
/s/	S	18	12		0	400	170	1720	950	2620	0	.5	1	3	2	-∞		28		28		40.25			0	0
/ʃ/	SH	18	12		0	400	170	2020	1190	2560	0	.5	1	3	2	-∞		31.5		42		31.5			0	0

[1]F_3 value is 0.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
/h/	H	9	10		0	490	0	1480	0	2500	0	1	1	0	7	35	−14	36.75	−14	26.25	−7	22.75	−3.5	1	0	7
/v/	V	20	4		1	280	170	1420	350	2560	980	.5	.5	3	2	29.75		40.25		36.75		33.25			0	0
	QQ	30	0											0	0										0	0
	VZ	20	4		1	280	170	1420	350	2560	980	.5	.5	3	2	29.75		40.25		36.75		33.25			0	0
/ð/	DH	20	4		1	280	170	1600	1190	2560	0	.5	1	3	2	29.75		31.5		26.25		28			0	0
	QQ	30	0											0	0										0	0
	DI	20	4		1	280	170	1600	1190	2560	0	.5	1	3	2	29.75		31.5		26.25		28			0	0
/z/	Z	20	4		1	280	170	1720	950	2560	0	.5	1	3	2	29.75		24.5		24.5		36.75			0	0
	QQ	30	0											0	0										0	0
	ZZ	20	4		1	280	170	1720	950	2560	0	.5	1	3	2	29.75		24.5		24.5		36.75			0	0
/ʒ/	ZH	20	4		1	280	170	2020	1190	2560	0	.5	1	3	2	29.75		26.25		36.75		26.25			0	0
	QQ	30	0											0	0										0	0
	ZI	20	4		1	280	170	2020	1190	2560	0	.5	1	3	2	29.75		26.25		36.75		26.25			0	0
/tʃ/	CH	23	4		0	190	110	1780	950	2680	2680	.5	0	2	2	−∞		−∞		−∞		−∞			0	0
	CI	18	8		0	400	170	2020	1190	2560	0	.5	1	3	2	−∞		31.5		42		31.5			0	0
/dʒ/	J	26	4		1	190	110	1780	950	2680	2680	.5	0	2	2	31.5									0	0
	JY	20	3		1	280	170	2020	1190	2560	0	.5	1	3	2	29.75		26.25		36.75		26.25			0	0
	QQ	30	0											0	0										0	0
	JZ	20	[illegible]		1	280	170	2020	1190	2560	0	.5	1	3	2	29.75		26.25		36.75		26.25			0	0
/l/	L	11	8		1	460	230	1480	710	2500	1220	.5	.5	6	0	36.75		26.25		26.25		21			0	0
/ɫ/	LL	11	8		1	460	230	940	470	2500	1220	.5	.5	6	0	36.75		26.25		26.25		21			0	0
/r/	R	10	11		1	490	0	1180	590	1600	740	.5[1]	.5	5[2]	5	42	21	35	17.5	35	17.5	−∞	0	.5	5	5
/w/	W	10	8		1	190	50	760	350	2020	980	.5	.5	4	4	43.75	21	28	14	21	10.5	−∞	0	.5	4	4
/j/	Y	10	7		1	250	110	2500	1190	2980	1460	.5	.5	4	4	50.75	24.5	33.25	17.5	38.5	17.5	31.5	14	.5	4	4
/ɪ/	I	2	6	6	1	400	170	2080	1070	2560	1340	.5	.5	4	4	50.75	24.5	36.75	17.5	35	17.5	29.75	14	.5	4	4
/ɛ/	E	2	8	4	1	640	350	2020	1070	2500	1220	.5	.5	4	4	50.75	24.5	42	21	38.5	17.5	31.5	14	.5	4	4
/æ/	AA	2	10	5	1	790	410	1780	950	2500	1220	.5	.5	4	4	50.75	24.5	47.25	24.5	38.5	17.5	31.5	14	.5	4	4

[1]F_1 value is 10. [2]F_1 value is 0.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
/ʌ/	U	2	9	6	1	700	350	1360	710	2500	1220	·5	·5	4	4	50·75	24·5	43·75	21	31·5	14	24·5	10·5	·5	4	4
/ɒ/	O	2	9	6	1	610	290	880	470	2500	1220	·5	·5	4	4	50·75	24·5	47·25	24·5	22·75	10·5	15·75	7	·5	4	4
/ʊ/	OO	2	6	4	1	370	170	1000	470	2500	1220	·5	·5	4	4	50·75	24·5	42	21	28	14	22·75	10·5	·5	4	4
/æ/	A	2	4	4	1	490	230	1480	710	2500	1220	·5	·5	4	4	50·75	24·5	50·75	24·5	33·25	17·5	26·25	14	·5	4	4
/i/	EE	2	11	7	1	250	110	2320	1190	3220	1580	·5	·5	4	4	50·75	24·5	33·25	17·5	36·75	17·5	31·5	14	·5	4	4
/ɜ/	ER	2	16	16	1	580	290	1420	710	2500	1220	·5	·5	4	4	50·75	24·5	45·5	21	33·25	17·5	26·25	14	·5	4	4
/ɑ/	AR	2	15	15	1	790	410	880	470	2500	1220	·5	·5	4	4	50·75	24·5	49	24·5	29·75	14	22·75	10·5	·5	4	4
/ɔ/	AW	2	16	10	1	490	230	820	470	2500	1220	·5	·5	4	4	50·75	24·5	45·5	21	22·75	10·5	17·5	7	·5	4	4
/u/	UU	2	14	9	1	250	110	880	470	2200	1100	·5	·5	4	4	50·75	24·5	38·5	17·5	17·5	7	10·5	3·5	·5	4	4
/eɪ/	AI	2	9	6	1	640	290	1600	830	2500	1220	·5	·5	5	5	50·75	24·5	45·5	21	35	17·5	29·75	14	·5	5	5
	AJ	2	8	6	1	400	170	2080	1070	2560	1340	·5	·5	4	4	50·75	24·5	36·75	17·5	35	17·5	29·75	14	·5	4	4
/aɪ/	IE	2	9	6	1	790	410	880	470	2500	1220	·5	·5	5	5	50·75	24·5	49	24·5	29·75	14	22·75	10·5	·5	5	5
	IF	2	8	6	1	400	170	2080	1070	2560	1340	·5	·5	4	4	50·75	24·5	36·75	17·5	35	17·5	29·75	14	·5	4	4
/ɔɪ/	OI	2	9	6	1	490	230	820	350	2500	1220	·5	·5	5	5	50·75	24·5	45·5	21	22·75	10·5	17·5	7	·5	5	5
	OJ	2	8	6	1	400	170	2080	1070	2560	1340	·5	·5	4	4	50·75	24·5	36·75	17·5	35	17·5	29·75	14	·5	4	4
/aʊ/	OU	2	9	6	1	790	410	1300	590	2500	1220	·5	·5	5	5	50·75	24·5	47·25	24·5	35	17·5	28	14	·5	5	5
	OV	2	8	6	1	370	170	1000	470	2500	1220	·5	·5	4	4	50·75	24·5	42	21	28	14	22·75	10·5	·5	4	4
/oʊ/	OA	2	9	6	1	490	230	1480	710	2500	1220	·5	·5	5	5	50·75	24·5	50·75	24·5	33·25	17·5	26·25	14	·5	5	5
	OB	2	8	6	1	370	170	1000	470	2500	1220	·5	·5	4	4	50·75	24·5	42	21	28	14	22·75	10·5	·5	4	4
/ɪə/	IA	2	9	6	1	310	170	2200	1070	2920	1460	·5	·5	5	5	50·75	24·5	35	17·5	36·75	17·5	31·5	14	·5	5	5
	IB	2	8	6	1	490	230	1480	710	2500	1220	·5	·5	4	4	50·75	24·5	50·75	24·5	33·25	17·5	26·25	14	·5	4	4
/ɛə/	AIR	2	9	6	1	640	350	2020	1070	2500	1220	·5	·5	5	5	50·75	24·5	42	21	38·5	17·5	31·5	14	·5	5	5
	AS	2	8	6	1	490	230	1480	710	2500	1220	·5	·5	4	4	50·75	24·5	50·75	24·5	33·25	17·5	26·25	14	·5	4	4
/ʊə/	OOR	2	9	6	1	370	170	1000	470	2500	1220	·5	·5	5	5	50·75	24·5	42	21	28	14	22·75	7	·5	5	5
	OP	2	8	6	1	490	230	1480	710	2500	1220	·5	·5	4	4	50·75	24·5	50·75	24·5	33·25	17·5	26·25	14	·5	4	4
/ɔə/	OR	2	9	6	1	490	230	820	470	2500	1220	·5	·5	5	5	50·75	24·5	45·5	21	22·75	10·5	17·5	7	·5	5	5
	OS	2	8	6	1	490	230	1480	710	2500	1220	·5	·5	4	4	50·75	24·5	50·75	24·5	33·25	17·5	26·25	14	·5	4	4