SYNTHESIS BY RULE OF PROSODIC FEATURES

IGNATIUS G. MATTINGLY
Joint Speech Research Unit, Eastcote

Reprinted from
LANGUAGE AND SPEECH
Vol. 9, Part 1, January-March 1966, pp. 1 - 13

SYNTHESIS BY RULE OF PROSODIC FEATURES

IGNATIUS G. MATTINGLY
Joint Speech Research Unit, Eastcote

Synthesis by rule of a limited set of prosodic features of Southern English has been attempted as an extension of a previously reported system for synthesis of segmental phonemes. Methods used for synthesis of intonation features, pausal features and prominence are described.

Recently the J.S.R.U. system for speech synthesis by rule described by Holmes, Mattingly and Shearme (1964) has been extended to provide for the synthesis of certain prosodic features of Southern English. Previously the segmental "phonetic elements" had been synthesized by rule, but the variations in duration, fundamental frequency and intensity which are the acoustic correlates of the prosodic, or "suprasegmental" features were either copied directly from natural utterances, or estimated, or simply ignored.

THE PROSODIC MODEL

The first problem in the synthesis of prosodic features was to decide upon the features to be synthesized. While phonologists agree fairly well about the inventory of segmental phonemes of Southern English, they do not agree about the prosodic features: facts, theory and terminology are still very much in question. It therefore seemed wisest, in our early attempts at prosodic synthesis, to be selective rather than inclusive, to concentrate on a few features which appeared to play an indispensable grammatical role. Our preliminary prosodic model is thus a very crude one. It includes three types of features: pausal features, intonation features and prominence.

By a pausal feature is meant simply a sense-group boundary (Jones, 1962, p. 274). In natural speech there are probably several types of pause, and depending upon circumstances their acoustic correlates will include pre-pausal lengthening, and/or silence of varying duration and/or certain effects upon the beginning of the following sense-group. In our model there are two pausal features: final pause, occurring at the end of an utterance; and non-final pause, occurring between sense-groups.

The intonation features are the "tunes" of the sense-groups. The tunes of natural speech distinguish between different syntactic structures and relate successive sense-groups; they also give the listener information about the personality of the speaker, his emotional state, or his attitude toward his utterance. (These latter aspects of intonation have concerned us much less than the former.) The acoustic correlate of a tune is the modulation of the fundamental frequency of the voiced portions of the sense-groups. Armstrong and Ward (1931) recognize two tunes; O'Connor and Arnold (1961), taking into account significant variations in "pre-head", "head", "nuclear tone" and " tail ", recognize 16 tunes, which they group into ten phonemically distinct " tone groups ". Though we suspect that even this elaborate system is not complete, our model recognizes just three tunes, differing only in nuclear tone, that is, in the intonation on the stressed syllable of the last prominent word and on any following syllables. These three nuclear tones are the falling tone, the fall-rise tone and the rising tone. The falling tone typically occurs at the end of a final clause when the thought is complete, and silence or a change of subject might be expected to follow ; it may also occur at the end of the first of two independent co-ordinate clauses. The fall-rise tone typically occurs at the end of a non-final dependent clause, or at the end of a final clause when the thought is incomplete, and further discussion of the same subject might be expected. The rising tone typically occurs at the end of yes-or-no questions and requests for repetition. But these loose characterizations of tne tones must not be taken literally ; O'Connor and Arnold (1961, p. 32) quite properly assert that " any sentence type can be linked with any tone group."

Prominence is the marking of certain words as being of special significance, enabling the listener to perceive sentence structure, detect parallels and contrasts of meaning and resolve ambiguities (Lieberman, 1964). The acoustic correlates of prominence in natural speech are a duration and an intensity of the stressed syllable of the prominent word greater than the syllable would otherwise have ; and a change, normally an increase, in the fundamental frequency of this syllable. Thu^s^ an occurrence of prominence modifies the underlying tune. There is more than one degree of prominence in natural speech ; our model allows only one. Moreover, our model makes no use of intensity as an acoustical correlate of prominence. Fry (1955, 1958) and Bolinger (1958) give good ground for believing that the importance of intensity as a cue to prominence is much less than used to be thought. Our experience in prosodic synthesis has tended to support this view.

While our prosodic model includes prominence, it does not take any direct account of lexical stress. In a non-prominent word, a stressed syllable is not treated any differently from an unstressed syllable composed of the same segmental phonemes. However, a great deal of information about stress is implicit in the selection and sequence of segmental phonemes and the sequence of morphemes. Thus, the nucleus of the stressed syllable will frequently have greater *inherent* intensity and/or duration than neighbouring syllable nuclei ; the fact that the phoneme /ə/ or the morpheme /rɪ/ are normally unstressed is often sufficient to determine the stress ọf the word ; and the suffix /ɪk/ generally indicates that the stress must fall on the preceding syllable.

Nor does our model make any provision for rhythm. A more sophisticated model would no doubt embody a tendency toward isochronism, i.e., the equalization of successive intervals between points of prominence (Classe, 1939), but in view of " the extreme difficulty of describing or reducing to rules the innumerable rhythms heard in ordinary connected speech " (Jones, 1962, p. 242), consideration of rhythm was postponed for the time being.

The features to be synthesized having been selected, the next step was to formulate programmable rules. The data on which these rules are based come from three sources: first, acoustic analyses of prosodic features such as those by Bolinger (1958), Uldall (1960, 1961, 1964), Hadding-Koch and Studdert-Kennedy (1964), Fry (1955, 1958), Lehiste and Peterson (1960, 1961) and others; this literature is small compared to what is available for the segmental phonemes. Second, handbooks intended for foreign students learning English, especially Armstrong and Ward (1931) and O'Connor and Arnold (1961). Such handbooks are most useful, though quite naturally they do not give extensive quantitative data. The two handbooks mentioned are supplemented by illustrative phonograph records (Armstrong and Ward, n.d.; O'Connor and Arnold, n.d.) and comparison of spectrograms of the recorded examples with the corresponding prosodic transcriptions in the text is most illuminating. Third, measurements of spectrograms of isolated sentences and technical prose, read aloud by one speaker in a deliberately perfunctory manner. It was hoped that such material would be reasonably free of personal, emotional or attitudinal colour, while exemplifying the prosodic features of grammatical significance. All of these sources were regarded as suggestive rather than prescriptive. As in the synthesis of segmental phonemes, the ultimate criterion of success is the acceptability of the utterance to a native speaker of Southern English.

THE SEGMENTAL SYNTHESIS SYSTEM

Before the prosodic rules and their operation in speech synthesis can be further explained, some account must be given of the basic segmental system, fully described by Holmes, Mattingly and Shearme (1964). This system consists of a nine-parameter, parallel-resonance electronic synthesizer, controlled by a punched paper tape; and a computer programme—let us refer to it as the segmental programme—which prepares this control tape. The set of parameter values for each of a series of 10 msec. units of time is represented on the paper tape as a sequence of five-bit characters.

There are two inputs to the segmental programme. The first is a "sentence", representing the utterance to be synthesized, and consisting of a series of literal symbols, an associated series of numbers, and certain "modifier" symbols. The numbers specify fundamental frequency. Each of the literal symbols denotes a "phonetic element" (the operational equivalent of a phoneme); these symbols approximate ordinary English spelling and are used because conventional phonetic symbols are impractical for computer display. The correspondence between phonemes and phonetic elements is given in Table 1, and a typical input sentence is shown in Fig. 1.

The second input is a set of tables for a particular dialect (in this case, Southern English); each table contains the sub-phonemic information required for the synthesis of one of the phonetic elements, such as the steady-state frequencies and amplitudes of the formants, the data needed to compute formant amplitude and frequency transitions, and as indication whether periodic excitation or random noise excitation is to be used. Also included in the input table is a standard duration tor the phonetic

TABLE 1

Phonetic elements and equivalent phonemes

Q	silence	F	f	AL	l̩	ER	ɜ
END	end	TH	θ	R	r (initial)	AR	ɑ
*P	p	S	s	RR	r (medial)	AW	ɔ
*T	t	SH	ʃ	W	w	UU	u
*K	k	H	h	Y	j	AI	eɪ
B	b	V	v	*I	ɪ	IE	aɪ
D	d	DH	ð	*E	ɛ	OI	ɔɪ
G	g	Z	z	*AA	æ	OU	aʊ
M	m	ZH	ʒ	*U	ʌ	OA	oʊ
AM	m̩	*CH	tʃ	*O	ɒ	IA	ɪə
N	n	J	dʒ	OO	ʊ	AIR	ɛə
AN	n̩	L	l	*A	ə	OOR	ʊə
NG	ŋ	LL	ɫ	EE	i	OR	ɔə

This table includes four elements [AN], [AM], [RR], [AL], not listed in Holmes (1964), and does not include supplementary elements, which are used to construct element sequences by the segmental programme but have no significance for the prosodic programme. Elements marked with * are unlengthenable in pre-pausal syllables.

element, expressed as a multiple of 10 msec. (The boundaries between elements implied by these standard durations are just convenient programming fictions which do not necessarily have any phonetic significance.) By the use of modifier symbols the standard duration of an element may be increased or decreased ; and the table for an element which may be a syllable nucleus includes not only the standard duration, appropriate for the nucleus of the stressed syllable of a prominent word, but also an alternative duration, appropriate for other syllable nuclei, which may be selected by the use of the [(] modifier. Fig. 1 illustrates modification of the standard duration by the [(] modifier, and also by the [>] modifier, each iteration of which increases the duration of an element by 20 msec.

The fundamental frequency F_0 is represented in the synthesizer by a pulse generator ranging in frequency from 50 to 250 c.p.s. For purposes of computation this range is quantized in 31 steps, spaced approximately logarithmically, the difference between successive steps being 1/13 of an octave or slightly less than a semitone. (The measured frequencies corresponding to each of the 31 steps are given in Table 2.)

In the input sentence F_0 is normally specified once for each element as a number in the range 1 to 31. This value is the value of F_0 at the boundary between the current element and the preceding element. The segmental programme determines the values of F_0 between successive boundary values by linear interpolation. Thus, any F_0

```
21   (          A

20              B
22              ER
                  3(  22   22   22   22   22   23   23   23  4(
21              D

18   (          I
17              N

15              DH
13   (          A

12              H
16              AA
15              N
14              D

13   (          I
11              Z

10              W
10   (          ER
10              TH

10              T
14              UU

12   (          I
11              N

10              DH
10   (          A

10              B
12    >>>>      OO
                  3(  13   13   13  4(
 5    >>>>      SH

 5              END
```

Fig. 1. An input sentence for the utterance

ə bɜd ɪn ðə hænd ɪz wɜθ tu ɪn ðə buʃ

TABLE 2

Fundamental frequency coding

PROGRAMME CODE	FREQUENCY	PROGRAMME CODE	FREQUENCY
1	52 c.p.s.	16	120 c.p.s.
2	55	17	126
3	59	18	132
4	63	19	138
5	67	20	146
6	70	21	152
7	74	22	157
8	77	23	164
9	80	24	169
10	84	25	177
11	90	26	188
12	97	27	200
13	103	28	215
14	109	29	226
15	116	30	238
		31	250

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 approximated by means of a programme option which permits the F_0 values for some or all time-units of an element to be explicitly specified in the input sentence. In Fig. 1 the column of figures at the left are the regular F_0 values ; the series of figures between the computer "triggers" [3(] and [4(], following the third element [ER], are special values of F_0 for the first eight time-units of that element ; the F_0 values for the remaining time-units of [ER] are computed by linear interpolation between the last of these special values, 23, and the next regular value, 21. Special F_0 values are also used for [OO] in the last word.

Since the segmental programme provides a means of varying duration and fundamental frequency according to specifications in the input sentence, it was convenient to embody the prosodic rules in a second programme—let us refer to it as the prosodic programme—the output from which is an acceptable input to the segmental programme. Thus the extended synthesis system consists of the two computer programmes and the electronic synthesizer in series. This arrangement is of course the equivalent in principle of a single device, the input to which is the input to the prosodic programme, and the output from which is synthetic speech.

The prosodic programme

Like the segmental programme, the prosodic programme has two inputs. The principal input is a string of phonetic element symbols, interspersed with various prosodic marks. The significance of these marks is as follows:

[']	precedes the stressed syllable of a prominent word.
[.]	falling tone, followed by final pause.
[,]	falling tone, followed by non-final pause.
[+]	fall-rise tone, followed by final pause.
[:]	fall-rise tone, followed by non-final pause.
[?]	rising tone, followed by final pause.

Here is a typical "input string":

A 'BERD IN DHA 'HAAND IZ WERTH
'TUU IN DHA 'BOOSH.

A string may include one or more sense-groups. A sense-group must include one or more prominent words, and must conclude with one of the five symbols specifying tone and pause. The symbols [+], [.], and [?] conclude a string. The phonetic elements of a sense-group up to the stressed syllable nucleus of the last prominent word constitute the head of the sense-group ; the remaining elements, its tail.

The second input to the programme is a set of prosodic tables, one for each segmental element, in which is stored certain information required for the synthesis of prosodic features and peculiar to that element. Other phonetic information is stored in the programme itself in the form of standard values assigned to a set of general programme parameters. (Hereafter, programme parameters are italicized.) The programme is written so that the value of any parameter can be changed to a desired value for a particular synthetic utterance by preceding the input string with a suitable declaration. The purpose of having prosodic tables external to the programme, and machinery for altering the general parameter values, is of course to facilitate the improvement of the prosodic rules ; neither changes in the tables nor declarations are made simply in order to touch up a particular synthesized utterance.

Given this input string, the prosodic tables and the general parameter values, the prosodic programme generates for the segmental programme an input sentence such as the one shown in Fig. 1, inserting silences, supplying durational modifiers and computing the F_0 boundary values and also special F_0 values as required.

The phonetic element at the beginning of the head of the first sense-group is assigned an arbitrary F_0 value in the upper middle of the available range ; we have been using *21* (152 c.p.s.). During the head, F_0 is allowed to vary between a prescribed maximum and minimum, say *25* and *10* (177 c.p.s. and 84 c.p.s.). As long as there is no interruption of voicing, and no prominent word occurs, the rules require that F_0 fall gradually at a slope dependent on the current syllable nucleus. The calculation of the F_0 boundary value associated with the next phonetic element requires the F_0 value associated with the current element, the duration stored in the prosodic table for that element and expressed in 10 msec. time-units, and the negative slope assigned by the programme and expressed in F_0 steps per time-unit. The slope used for a syllable nucleus is stored in its prosodic table, and this same slope is assigned to all voiced consonants between this syllable nucleus and the following syllable nucleus or voiceless consonant or pause, whichever occurs first (durations and slopes of syllable nuclei are given in Table 3). Voiced consonants preceding the first syllable nucleus of a sense-group are assigned an arbitrary slope of *−1/7*. The relationship between the value of the slope and the identity of the syllable nucleus, perhaps somewhat surprising, is suggested by our admittedly limited data, but has not so far as we know, been mentioned elsewhere. We do not know how general the phenomenon is, or how important it is for speech synthesis; perhaps we would do as well or better with a fixed slope.

If a sequence of one or more voiceless elements occurs, F_0 is of course not actually used by the synthesizer, but must nevertheless be represented in the control tape, and therefore specified in the input sentence. The same value of F_0 which has been calculated for the first voiceless element is associated with each of the following voiceless elements. At the beginning of the next voiced element, F_0 increases *1* step; its fall is then resumed at the slope of the following syllable nucleus. This slight increase of F_0 after a voiceless sequence corresponds to a phenomenon observed in our data, and the same reservations apply as in the case of phoneme-dependent slope.

The occurrence of a prominence mark results in a *3*-step increase in F_0. If the voiced portion of the stressed syllable of the prominent word follows a voiceless sequence, a total increase of four steps occurs at the beginning of the first voiced element. But if there has been no voiceless sequence since the previous syllable nucleus, the three-step rise is distributed over the interval between the previous syllable nucleus and the midpoint of the stressed syllable nucleus. No use is made of the slopes previously assigned to the elements in this interval, and the programme must associate a sequence of special F_0 values with the stressed syllable nucleus in the input sentence (see Fig. 1). The differential placement of the prominence peak dependent on preceding voicing or voicelessness was clearly observable in our data and corresponds with perceptual results reported by Classe (1939).

Once the prominence peak has been reached, whether at the onset of voicing or in the middle of the syllable nucleus, the F_0 fall resumes. However, a special set of slope values is used for the remainder of the stressed syllable nucleus and the following voiced consonants (see Table 3). These slope values are either less negative than or equal to the corresponding non-prominent values, and in the case of [EE] and [AR] are chosen so as to be effectively zero. Thus, in our model, prominence produces not only an F_0 peak on the stressed syllable nucleus, but also a tendency for F_0 to fall more gradually during the stressed syllable than during other syllables.

The rules for the head also require that the stressed syllable nucleus of a prominent

TABLE 3

Prosodic parameter values of syllable nuclei

	DURATIONS		SLOPES	
	(10 msec. time-units)		(in F_0 steps/time-unit)	
	Non-prominent	Prominent	Non-prominent	Prominent
AM	6	—	−1/7	—
AN	6	—	−1/7	—
AL	12	—	−1/7	—
I	6	6	−1/4	−1/4
E	4	8	−1/4	−1/7
AA	5	10	−1/7	−1/7
U	6	9	−1/7	−1/7
O	6	9	−1/4	−1/7
OO	4	6	−1/4	−1/4
A	4	—	−1/4	—
EE	7	11	−1/7	−1/31
ER	16	16	−1/4	−1/4
AR	15	15	−1/7	−1/31
AW	10	16	−1/7	−1/7
UU	9	14	−1/7	−1/7
AI	12	17	−1/7	−1/7
IE	12	17	−1/7	−1/7
OI	12	17	−1/7	−1/7
OU	12	17	−1/7	−1/7
IA	12	17	−1/7	−1/7
AIR	12	17	−1/7	−1/7
OOR	15	20	−1/7	−1/7
OR	12	17	−1/7	−1/7

word have a duration greater than or equal to the duration of the same element in a non-prominent word (see Table 3). The two possible durations for each syllable nucleus are stored in its prosodic table, since they are required for the calculation of the F_0 values; and the prosodic programme indicates the selection of a non-prominent duration to the segmental programme by inserting the durational modifier [(] in the input sentence. The standard durations of elements other than syllable nuclei are not changed during the head.

In this way, successive F_0 values are calculated and durations adjusted until the prominence peak of the stressed syllable of the last prominent word is reached. The specification of F_0 during the rest of the sense-group, the tail, depends upon which tone is indicated by the prosodic mark which concludes the sense-group.

For a falling tone, the F_0 value associated with the element immediately following the stressed syllable nucleus is *8* steps below the peak value or step *6* (70 c.p.s.), whichever is greater. Thereafter, F_0 falls more gradually. The remaining voiced elements are assigned a slope of *−1/12*, the voiceless elements a slope of zero, and the F_0 values are calculated accordingly. F_0 is not permitted to fall below *1* (52 c.p.s.).

For a fall-rise tone, a fall of *5* steps and a following rise of *5* steps are distributed over the final voiced elements of the sense-groups, beginning with the last syllable nucleus. If the last syllable nucleus is also the stressed syllable nucleus of the last prominent word, the fall-rise begins at the prominence peak. Otherwise, the portion of the stressed syllable nucleus after the prominence peak, and all voiced elements between the stressed syllable nucleus and the final syllable nucleus, are assigned a slope of *−1/7*, voiceless elements are assigned a slope of zero, and the F_0 values are calculated accordingly. The prosodic programme must associate special F_0 values with the element during which the minimum value of the fall-rise occurs. F_0 is not permitted to fall below *1* (52 c.p.s.).

For a rising tone, the F_0 value associated with the element immediately following the stressed syllable nucleus is *4* steps above the peak value. Thereafter F_0 continues to rise. The remaining voiced elements are assigned a slope of *1/12* and the F_0 values calculated accordingly. A sequence of voiceless elements increases F_0 by *1* step, as during the head. F_0 is not permitted to rise above *25* (177 c.p.s.).

Both non-final and final pauses require pre-pausal lengthening. The rule followed is: proceeding backwards from the end of the sense-group, increase the duration of each element by 8 time-units until either a syllable nucleus has been lengthened, or an element marked in the prosodic tables as "unlengthenable" for this purpose is encountered. The unlengthenable elements are indicated by an asterisk in Table 1 and include the voiceless stops and affricates and most of the short vowels.

This pre-pausal lengthening is taken into account in the calculation of the F_0 values of the three tones, and the prosodic programme indicates the desired lengthening to the segmental programme by inserting the modifiers [>>>>] in the input sentence. We are aware that this lengthening rule is crude and even incorrect, since in natural speech, syllables before the ultima in a polysyllabic tail should also be slightly lengthened.

In the case of a non-final pause, the prosodic programme inserts the element [Q] "silence", in the input sentence, with a [>] modifier. Since the standard duration of [Q] is ten time-units, this results in a silence 12 time-units long. This initial value of F_0 for the next sense group is set at *17* (126 c.p.s.), and the programme then proceeds exactly as before.

In the case of a final pause, the programme inserts the dummy element [END], completing the preparation of the input sentence.

Discussion

Since the preparation of an input string requires only a few minutes, and the remaining steps in the preparation of the control tape are carried out by the computer, it is quite easy to synthesize a large amount of speech. In practice, however, we limited our corpus to some 15 utterances, each of which was synthesized several times in the course of testing progressively more refined versions of the prosodic rules. The utterances selected for synthesis—many of them taken from Armstrong (1931)—include

(1) simple statements, consisting of only one sense-group:
IT 'IZNT IGZAAKTLI WOT IE 'WONT.
'NOU IEM AZ FIT AZ A 'FIDAL AGAIN.

(2) questions of similar prosodic structure:
'WIE DOANT YUU 'MIEND YOOR 'OAN 'BIZNIS.

(3) yes-or-no questions:
DUZ 'DHIS TRAIN STOP AT KLAAPAM
'JUNGKSHAN?
DID YUU 'KUM BIE 'MOATAKAR?

(4) simple sentences with more than one sense-group:
'SUMWUN: 'SUMWAIR: WONTS A 'LETA
FRAM 'YUU.

(5) compound and complex sentences:
IEM GOAING 'HOAM NOU, AAND IEM 'NEVA
KUMING 'BAAK.
IF YUU AR 'KUMING: 'PLEEZ LET MI 'NOA.

(6) longer complex sentences:
LIEK 'MOAST 'OALLD PEEPAL: HEE WAZ 'FOND
AV TAWKING ABOUT 'OALLD 'DAIZ, AND 'AAZ
HEE HAD NOAN 'HOASTS AV INTRASTING AND
IMPAWTANT 'MEN: HAAD A TA'NAISHAS
MEMARRI: AND SPOAK DHA MOAST 'FINISHT
'INGGLISH: IT WAZ A 'PLEZHA TA LISAN TA
HIZ REMIN'ISANSIZ.

Given the crudity of the basic scheme, the quality of the resulting synthetic speech is not unsatisfactory for a first attempt. The rules which control variations in fundamental frequency seem fairly successful ; and the intonation features, for which F_0 is the only correlate, are almost always correctly identified by listeners without previous knowledge of the utterance, and are considered reasonably natural. However, these features have a low information content ; the targets are large and easy to hit. Moreover, the tune and the verbal content of the sentence reinforce one another, and we have not yet made any systematic attempt in the manner of Uldall (1961) to study the listeners' identification of the tunes when used with otherwise ambiguous utterances.

The features which involve adjustments of duration are not quite as successful. When listeners are asked to locate the peaks of prominence, they sometimes select other syllables of relatively long inherent duration as well as, or instead of, those specified as prominent in the input string. As previously mentioned, the rule for pre-pausal lengthening needs revision. And finally, the lack of rules to control rhythm is quite apparent in the longer utterances.

Summary and conclusions

The basic J.S.R.U. system for segmental synthesis by rule has been extended to permit synthesis by rule of certain prosodic features. The prosodic computer programme, the principal input to which is a string of phonetic symbols interspersed with prosodic markers, prepare the input sentence for the segmental programme, which in turn prepares the control tape for the electronic synthesizer. The limited set of features synthesized so far includes pausal features, intonation features and prominence. The preliminary results thus far are encouraging, but the rules, particularly those controlling variations in duration, can certainly be further refined.

I wish to express my gratitude to J. D. O'Connor and Elizabeth Uldall for many valuable suggestions, to John Gill for serving as an experimental subject ; and to John Holmes for his constant encouragement and constructive criticism.

References

Armstrong, L. E. and Ward, I. C. (1931). A Handbook of English Intonation (Cambridge).

Armstrong, L. E. and Ward, I. C. (n.d.). A Handbook of English Intonation. Linguaphone Institute records 76788-93.

Bolinger, D. (1958). A theory of pitch accent in English. *Word*, 14, 109.

Classe, A. (1939). The Rhythm of English Prose (Oxford).

Coleman, H. O. (1914). Intonation and emphasis. *Miscellanea Phonetica I*, 6.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *J. acoust. Soc. Amer.*, 27, 765.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1, 126.

Fry, D. B. (1964). The dependence of stress judgments on vowel formant structure. *Proc. Fifth Int. Congr. Phonetic Sciences* (Basel), 306.

Hadding-Koch, K. (1961). Acoustico-phonetic Studies in the Intonation of Southern Swedish (Lund).

Hadding-Koch, K. and Studdert-Kennedy, M. (1964). An experimental study of some intonation contours. *Phonetica*, 11, 175.

Holmes, J. N., Mattingly, I. G. and Shearme, J. N. (1964). Speech synthesis by rule. *Language and Speech*, 7, 127.

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

Jones, D. (1962). An Outline of English Phonetics (Cambridge).

KINGDON, R. (1958). The Groundwork of English Intonation (London).

LEHISTE, I. and PETERSON, G. (1960). Duration of syllable nuclei in English. *J. acoust. Soc. Amer.*, 32, 693.

LEHISTE, I. and PETERSON, G. (1961). Some basic considerations in the analysis of intonation. *J. acoust. Soc. Amer.*, 33, 419.

LIEBERMAN, P. (1964). Intonation and syntactic processing of speech. Preprint of paper read at Symposium on Models for the Perception of Speech and Visual Form, Boston.

LIEBERMAN, P. (1965). On the acoustic basis of the perception of intonation by linguists. *Word*, 21, 40.

MEYER, E. A. (1903). Englische Lautdauer (Uppsala).

O'CONNOR, J. D. and ARNOLD, G. F. (1961). Intonation of Colloquial English (London).

O'CONNOR, J. D. and ARNOLD, G. F. (n.d.) Intonation of Colloquial English. Longmans records LGO. EP. 200-203.

ULDALL, E. (1960). Attitudinal meanings conveyed by intonation contours. *Language and Speech*, 3, 223.

ULDALL, E. (1961). Ambiguity: question or statement? or "Are you asking me or telling me?" *Proc. Fourth Int. Congr. Phonetic Sciences* (The Hague), 779.

ULDALL, E. (1964). Dimensions of meaning in intonation. In Abercrombie, D., et al. (eds.), *In Honour of Daniel Jones* (London).