Intonation in text-to-speech synthesis: Evaluation of algorithms

Glenn Akers[a] and Matthew Lennig[b]
Bell-Northern Research, 3 Place du Commerce, Verdun, Quebec, Canada H3E 1H6

(Received 25 May 1984; accepted for publication 30 January 1985)

Two algorithms, termed schematic and naturalistic, for generating intonation contours in an English text-to-speech system are compared by eliciting preference judgments from a total of 21 subjects. The major problem for both algorithms, but especially for the schematic algorithm, has to do with accent assignment and with the determination of the intonation phrase rather than with the phonetic realization of accent through manipulation of $F0$. Due to parser errors, phrase boundaries are incorrectly identified in 30% of the sentences used in the three experiments. Moreover, the naturalistic algorithm uses a grammatical part-of-speech hierarchy which ranks nouns higher than verbs. Therefore, incorrect classification of verbs as nouns (the major classification error) results in an unintended accent. The results indicate that accent assignment and phrase determination are the primary areas requiring improvement in order to further increase the naturalness of synthetic speech intonation.

PACS numbers: 43.72.Ja

INTRODUCTION

Once considered a neglected topic, intonation has received considerable research attention in the past 10 years. Within the context of text-to-speech synthesis, intonation refers to the stream of fundamental frequency ($F0$) values used to model the timing of the glottal pulse source of voiced speech sounds. The fundamental function of intonation is to signify the prominence of the most important words in discourse. Prominence is signaled by either a significant rise or fall in pitch, the perceptual correlate of $F0$. Because intonation indicates the new information in an utterance, improvement in the modeling of intonation is one of the most important areas for increasing the intelligibility and enhancing the quality of text-to-speech synthesis systems (Olive and Nakatani, 1974; Nakatani and Schaffer, 1978). Another important function of intonation is to delimit and segment speech into higher-level syntactic units (Collier and ‘t Hart, 1975; Daneš, 1960).

Although several systems have been described for generating synthetic intonation contours (Clark, 1981; Mattingly, 1966; Maeda, 1976; ‘t Hart and Cohen, 1973; Witten, 1977), listeners still find prosody one of the markedly unnatural aspects of synthetic speech. In an attempt to understand how to create more natural-sounding intonation contours for text-to-speech synthesis, we have compared two intonation algorithms by means of forced-choice subjective preference tests.

The two algorithms we compared differ fundamentally in the way they model intonation. One of them, which we will refer to as the naturalistic algorithm, attempts to imitate the details of natural speech as accurately as possible. This includes rules for modeling segmental effects on fundamental frequency (micromelody). The other approach, which we term the schematic algorithm, is a broader-brush approach, which ignores phonetic detail, attempting to concentrate on global patterns. In contrast to the naturalistic approach, the schematic algorithm does not model segmental effects on $F0$. An important difference between the two models is that the naturalistic algorithm makes use of a detailed part-of-speech hierarchy to determine accent pitch levels, while the schematic algorithm employs only the dichotomy between function and content words. The schematic algorithm is due to Pierrehumbert (1980, 1981a, b).[1] The naturalistic algorithm, due to O’Shaughnessy (1976, 1979), is part of the MITalk-79 text-to-speech system.[2]

I. DESCRIPTION OF THE NATURALISTIC AND SCHEMATIC ALGORITHMS

The two algorithms compared represent implementations of different theoretical positions regarding the basic units of intonation. The naturalistic algorithm is based upon a model of intonation expressed in terms of pitch direction. This is essentially the position advocated by Bolinger (Abe and Kanekiyo, 1965; Bolinger, 1972a, b). In contrast, the schematic algorithm is based upon a model of intonation in terms of accent levels. It is within the American school tradition (Pike, 1945; Wells, 1945; Hockett, 1955; Trager and Smith, 1957; Liberman, 1975; Leben, 1976; and Goldsmith, 1976; see Bolinger, 1951 for an insightful comparison of the two approaches). A third approach to intonation, that of the British school (Halliday, 1967; Crystal, 1969, 1975) has been applied to speech synthesis by Mattingly (1966), Witten (1977), and Clark (1981), but is not addressed in this study.

A. Calculation of fundamental frequency

Both the schematic and the naturalistic algorithms use declination lines to determine $F0$ values for each frame of the synthetic utterance. The naturalistic algorithm calculates $F0$ peaks using a part-of-speech hierarchy to determine the degree of excursion from the declination line. It also employs a large number of rules to reproduce many details of $F0$ patterns observed in natural speech, such as consonantal effects. The schematic algorithm uses two declination lines, a topline and a baseline, which together determine the possible $F0$ range as a function of time. Both lines have negative slopes,

[a] Glenn Akers is currently affiliated with Language Research Foundation, 215 Washington Street, Belmont, MA 02178.

[b] Matthew Lennig is also with INRS-Telecommunications (University of Quebec).

TABLE I. Variable values for schematic algorithm: Experiments 1 and 2.

185.0 Hz	beginning topline value
135.0 Hz	beginning baseline value
95.0 Hz	ending topline value
65.0 Hz	ending baseline value
95.0 Hz	starting $F0$ value
110.0 Hz	continuation rise $F0$ value
65.0 Hz	terminal declarative $F0$ value
85.0 Hz	phrase accent $F0$ value

TABLE II. Variable values for the schematic algorithm: Experiment 3.

175.0 Hz	beginning topline value
125.0 Hz	beginning baseline value
105.0 Hz	ending topline value
75.0 Hz	ending baseline value
100.0 Hz	continuation rise $F0$ value
70.0 Hz	phrase accent $F0$ value

with the topline descending more quickly than the baseline. Topline and baseline are reset to their initial values at the beginning of each new intonation phrase. Starting and ending values for these lines used in our experiments are given in Tables I and II. $F0$ values for accented syllables are given by the formula

$$F0 = \text{baseline} + \text{accent target (topline} - \text{baseline)},$$

where accent target is a number between 0 and 1 assigned to each accented syllable as described in the next paragraph. Intermediate $F0$ values between targets are determined by parabolic interpolation (Pierrehumbert, 1981a).

Theoretically, accent assignment under the schematic algorithm refers to metrical tree structure as developed in Liberman (1975) and Liberman and Prince (1977). However, since the automatic generation of such tree structures for unlimited text is errorful and somewhat beyond the scope of present technological capabilities, the algorithm uses a heuristic for accent assignment suggested by Liberman: The last content word of the phrase receives an accent target of 1.0; earlier accents in the phrase alternately are assigned accent targets of 0.7 or 0.4. Liberman's heuristic requires a grammatical analysis which distinguishes only between content and function words, and therefore does not require a full parse of the input text. This heuristic aims to model the first-order approximation to intonation which states that in neutral declarative sentences the nuclear accent is the most prominent accent and that earlier accents in the phrase tend to alternate in prominence. The total number of accented words may differ between the schematic and naturalistic algorithms since Liberman's heuristic accents all content words while the naturalistic algorithm often does not accent the last content word in a noun phrase.

One of the most important differences between the schematic and naturalistic algorithms is in their approaches to accent assignment. The naturalistic algorithm ranks each word in the phrase in terms of a grammatical hierarchy. It therefore requires part-of-speech information from the parser. Only nouns, adjectives, adverbs, reflexive pronouns, some modals, quantifiers, interrogative adjectives, negatives, and sentential adverbs receive accent prominence unless a phrase does not contain any tokens from these categories. Therefore, incorrect classification of verbs as nouns by the MITalk parser (the major classification error) results in unintended accentuation.

The two algorithms also differ in terms of their placement in time of the $F0$ peaks. The schematic algorithm always places the peak in the vowel. The naturalistic algorithm usually places the peak in the vowel, except in accented syllables beginning with a consonant plus sonorant sequence, where the peak is placed in the sonorant (an exception is sequences of /dr/, for example "drop" and "dried" which have the peak in the vowel). The schematic algorithm places the peak later in non-nuclear accents (beginning at 60% of the total duration of the vowel) than in nuclear accents (beginning at 20% of the total duration of the vowel).

Examples of intonation curves produced by the algorithms for the sentence *The sink is the thing in which we pile dishes* are given in Figs. 1–4 and help to illustrate differences

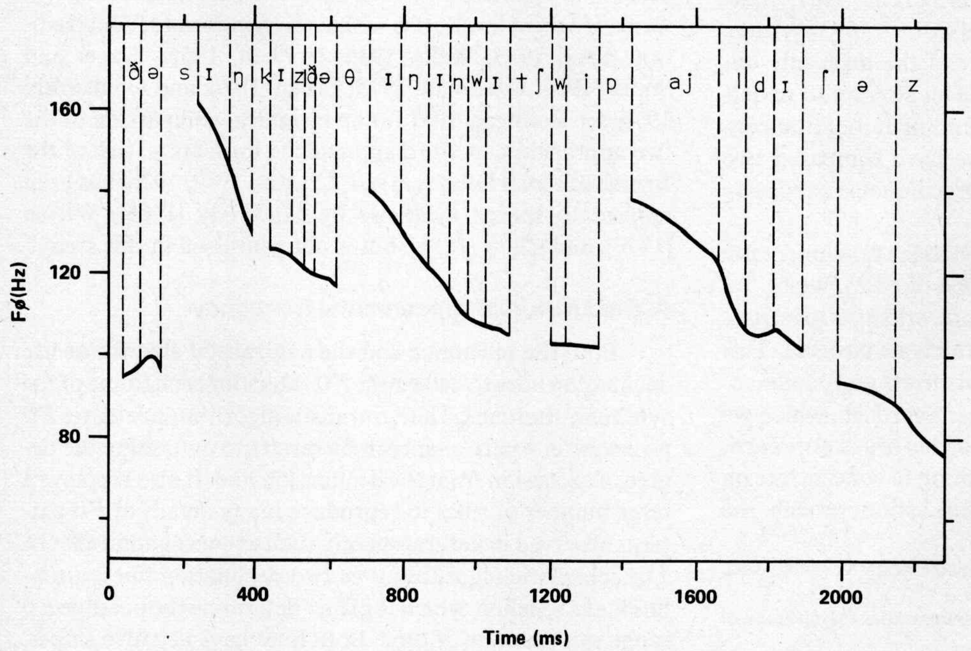

FIG. 1. Intonation pattern generated by the naturalistic algorithm for the sentence "The sink is the thing in which we pile dishes."

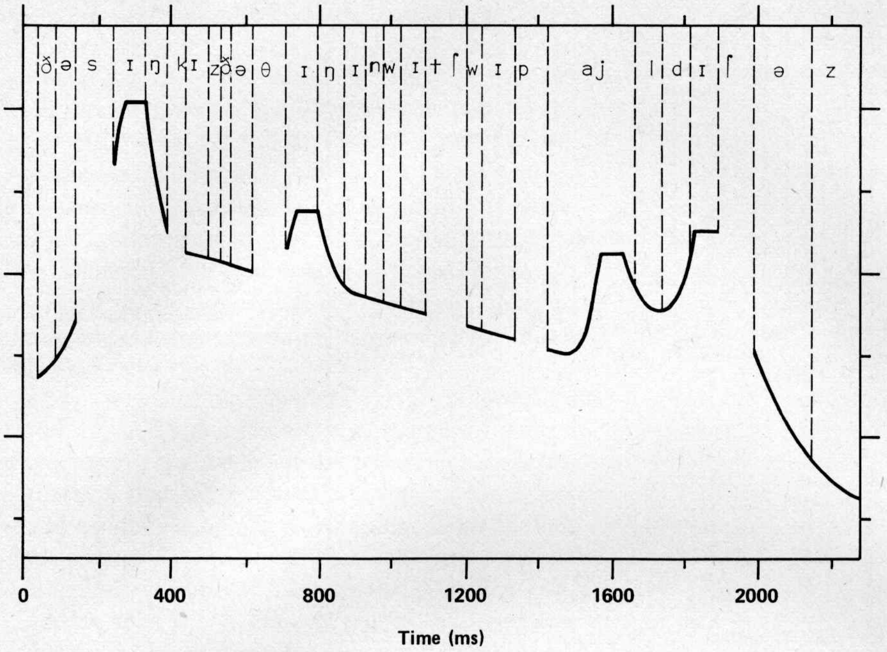

FIG. 2. Intonation pattern generated by the variant of the schematic algorithm which places high accent on the final stressed vowel. The sentence is "The sink is the thing in which we pile dishes."

between them. Figure 1 shows the output of the naturalistic algorithm. Figure 2 shows the output of the variant of the schematic algorithm which places high accent on the final stressed vowel. Figure 3 shows the output of the variant of the schematic algorithm which places low accent on the final stressed vowel. Figure 4 combines Figs. 1 and 2 to highlight differences between the naturalistic algorithm and the high accent variant of the schematic algorithm.

One obvious difference is that the accented vowels in the two schematic versions (Figs. 2 and 3) maintain their maximum $F0$ values for 60-ms intervals, giving rise to flat tops in the $F0$ curve, while the accented vowels in the naturalistic version (Fig. 1) attain their maximum $F0$ values at a single 5-ms frame. Furthermore, maximum $F0$ differences in the naturalistic version occur early in the vowel, while they occur later in the two schematic versions. Compare the peak $F0$ values for the vowel in *pile* in Fig. 1 versus Figs. 2 and 3.

The major difference between the schematic versions is that the stressed initial vowel of *dishes* in Fig. 2 receives a high $F0$ accent value, while it receives a low $F0$ accent value in Fig. 3. This low $F0$ accent value makes the intonation curve in Fig. 3 closer to that of Fig. 1 than that of Fig. 2.

II. EXPERIMENTAL PROCEDURE

In each of three experiments, ten native English-speaking listeners with no hearing or speech disabilities judged pairs of synthetic utterances which differed only in fundamental frequency. All utterances were generated using the MITalk-79 system with either the original (naturalistic) intonation module or with the schematic intonation module.

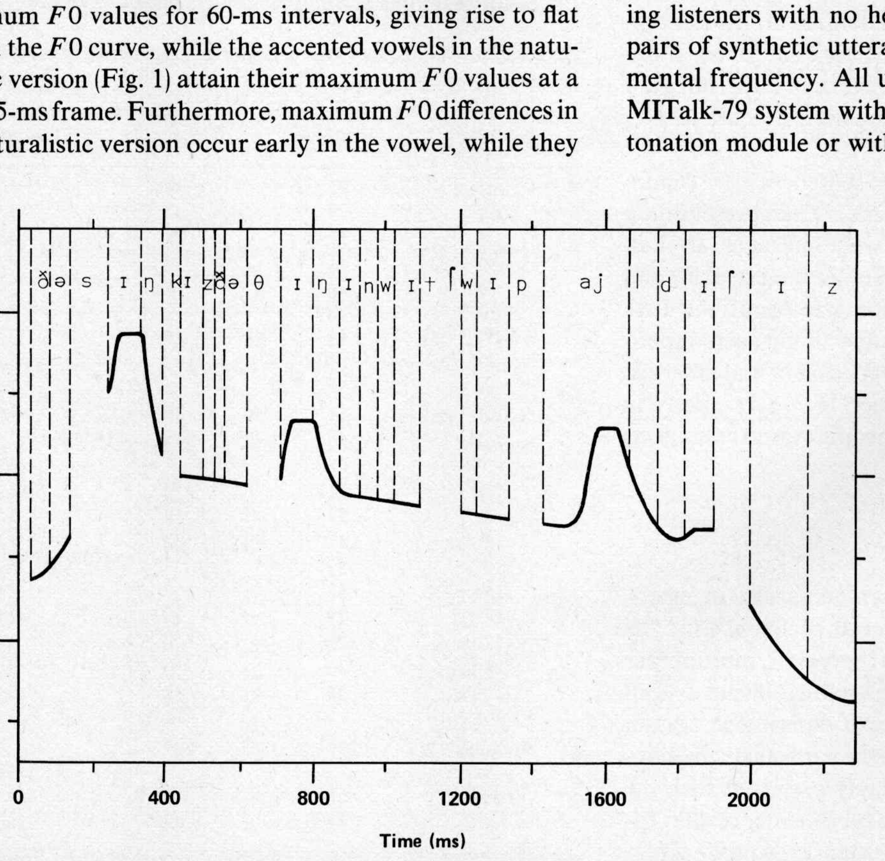

FIG. 3. Intonation pattern generated by the variant of the schematic algorithm which places low accent on the final stressed vowel. The sentence is "The sink is the thing in which we pile dishes."

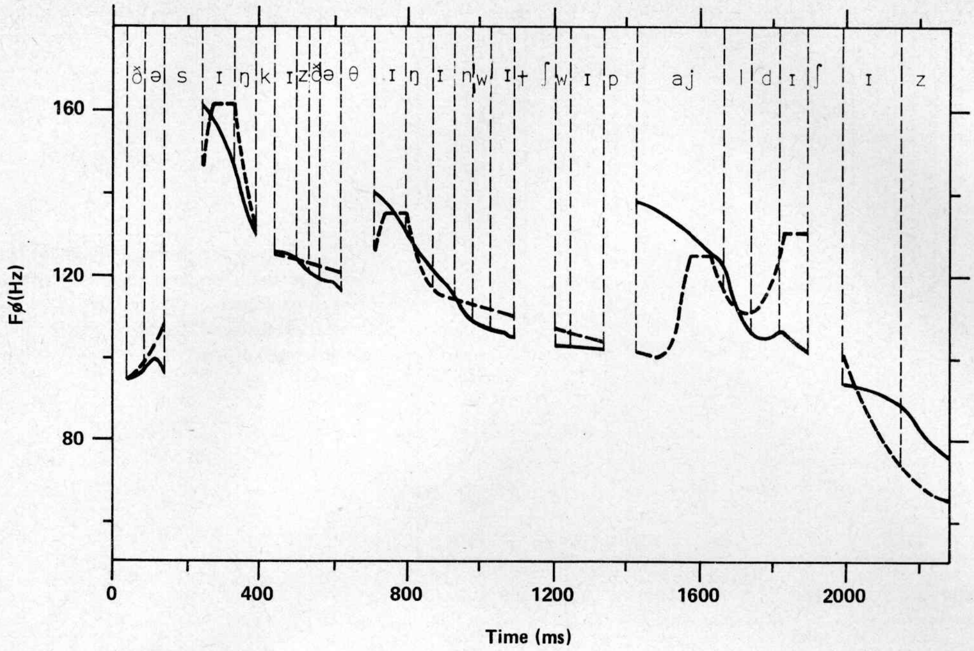

FIG. 4. Composite of Figs. 1 and 2. The solid line represents the naturalistic algorithm, while the dotted line represents the high accent version of the schematic algorithm.

While the naturalistic algorithm did not vary, three variants of the schematic algorithm were used, one in each of the three experiments.

In experiment 1, the schematic algorithm assigned a high accent target of 1.0 to the stressed vowel of the last content word in the phrase (as in Pierrehumbert, 1981a). In experiment 2, a low accent target of 0.1 was used instead. A blend of these two strategies was used in experiment 3: Successive phrases received alternating high and low accent targets. This seemed appropriate since the test sentences used in experiment 3 were longer, containing multiple intonation phrases. These ten complex sentences of longer duration (10–18 s) are reproduced in Appendix B. Experiments 1 and 2 employed 20 phonetically balanced sentences (IEEE, 1969), reproduced in Appendix A.

In all three experiments, the test sentences were randomized and presented twice in each order. Each presentation was in ABAB format, and subjects were instructed to indicate which version they preferred. Subjects were paid, and all are native English speakers residing in Montreal. Five males and five females served as subjects for each experiment. Eight of the ten subjects for experiment 1 also served as subjects for experiment 2. One subject (PD) participated in all three experiments. A total of 21 people served as subjects in the three experiments.

III. RESULTS

For the phonetically balanced sentences used in experiments 1 and 2, there is a slight overall preference for the naturalistic intonation algorithm. In the longer, multiphrase sentences used in experiment 3, there is no significant overall preference for either algorithm. In each experiment, certain sentences are strongly preferred with the naturalistic $F0$ contour, while other sentences are strongly preferred with the schematic $F0$ contour. In all three experiments, certain listeners had a strong preference for one algorithm or the other. In experiment 3 alone, a strong correlation was observed between listener's sex (significant at the 0.001 level) and utterance preference: Males prefer the schematic versions, while females prefer the naturalistic versions.

A. Experiment 1: High nuclear accent

The purpose of experiment 1 was to compare the naturalistic algorithm with a version of the schematic algorithm using high nuclear accent of 1.0. A summary of the preferences for each test sentence is given below in Table III. The

TABLE III. Cumulative preferences for the naturalistic versus schematic intonation (experiment 1).

Sentence	Naturalistic	Schematic	Duration (s)
(1)	21	19	3.5
(2)	7	33	3.1
(3)	11	29	3.4
(4)	31	9	3.0
(5)	33	7	3.3
(6)	38	2	3.4
(7)	11	29	3.1
(8)	7	33	3.4
(9)	32	8	2.7
(10)	11	29	3.0
Subtotal	202	198	
(11)	31	9	3.1
(12)	22	18	3.2
(13)	17	23	3.2
(14)	35	5	3.1
(15)	33	7	2.9
(16)	28	12	2.8
(17)	20	20	3.2
(18)	30	10	3.5
(19)	28	12	3.0
(20)	9	31	3.2
Subtotal	253	147	
Total	455 57%	345 43%	

duration of each sentence is given in the last column of the table.

A summary of the preferences, by subject, is given in Table IV. Although there is a small overall preference for the naturalistic versions, four of the ten subjects somewhat preferred the schematic versions.

For experiment 2, the schematic algorithm was modified to employ low pitch accent instead of high pitch accent. Phrase-final content words were assigned a low accent target of 0.1. As with the utterances used in experiment 1, earlier accents in the phrase were alternately assigned target values of 0.7 and 0.4.

A major problem for the schematic version of sentences 5, 6, 9, 11, 14, 15, 16, 18, and 19 used in experiment 1 seemed to be an unnatural prominence of the phrase-final accents. We hypothesized that more natural versions of these sentences would result by assigning a low accent target to their phrase-final accents. Therefore, if this analysis is correct then the schematic versions of these sentences (sentences 5, 6, 9, 11, 14, 15, 16, 18, and 19), and only these sentences, should be preferred more often in experiment 2 than in experiment 1.

Examination of Table V reveals that except for sentence 11, this prediction is supported. Moreover, subjects favored the deaccented schematic version over the naturalistic version of these nine sentences on 36 more occasions than they favored the accented schematic versions in experiment 1.

A summary of the preferences in experiment 2, by subject, is given in Table VI. Except for subject SD, all subjects preferred the naturalistic versions overall.

B. Experiment 3: Alternating accents

Experiment 3 was designed to compare the naturalistic algorithm to a modified version of the schematic algorithm on longer, more natural sentences than used for experiments 1 and 2. In order to introduce variation in the accentual pattern in experiment 3, the accent assignment algorithm used in generating the schematic versions alternately assigned an accent target of 1.0 or 0.1 to the stressed vowel of the last content word of each phrase. As with the stimuli used for experiments 1 and 2, earlier accents in the phrase were alternately assigned values of 0.7 and 0.4. A summary of the preferences for each test sentence used in experiment 3 is given in Table VII. The duration of each sentence is given in the last column.

TABLE IV. Preferences by subject for the naturalistic versus schematic intonation algorithms (experiment 1).

	Naturalistic	Schematic
Females		
TD	38	42
LT	39	41
JS	41	39
MB	57	23
GW	61	19
Subtotal	236	164
Males		
PD	37	43
JL	38	42
RS	46	34
BM	47	33
KR	51	29
Subtotal	219	181
Total	455	345
	57%	43%

TABLE V. Cumulative preferences for naturalistic versus schematic intonation algorithms (experiments 1 and 2).

Sentence	Naturalistic			Schematic		
	exp 1	exp 2	change	exp 1	exp 2	change
(1)	21	29	+8	19	11	−8
(2)	7	25	+18	33	15	−18
(2)	11	19	+8	29	21	−8
(4)	31	31	0	9	9	0
(5)	33	24	−9	7	16	+9
(6)	38	30	−8	2	10	+8
(7)	11	20	+9	29	20	−9
(8)	7	17	+10	33	23	−10
(9)	32	30	−2	8	10	+2
(10)	11	20	+10	29	20	−10
Subtotal	202	245	+43	198	155	−43
(11)	31	32	+1	9	8	−1
(12)	22	24	+2	18	16	−2
(13)	17	18	+1	23	22	−1
(14)	35	30	−5	5	10	+5
(15)	33	32	−1	7	8	+1
(16)	28	24	−4	12	16	+4
(17)	20	29	+9	20	11	−9
(18)	30	26	−4	10	14	+4
(19)	28	24	−4	12	16	+4
(20)	9	20	+11	31	20	−11
Subtotal	253	259	+6	147	141	−6
Total	455	504	+49	345	296	−49
	57%	63%	+6%	43%	37%	−6%

For six of the ten sentences, the naturalistic version was preferred overall, although by an insignificant margin for sentence 25. Analysis of each pair of synthetic utterances indicates several factors which distinguish them. One impor-

TABLE VI. Preferences by subject for naturalistic versus schematic intonation algorithms (experiment 2).

	Naturalistic	Schematic
Females		
SD	35	45
CK	50	30
DH	54	26
PO	57	23
DM	62	18
Males		
MS	42	38
PO	42	38
PG	44	36
PD	52	28
RL	66	14
Total	504	296
	63%	37%

TABLE VII. Cumulative preferences for the naturalistic versus schematic intonation algorithms (experiment 3).

Sentence	Naturalistic	Schematic	Duration (s)
(21)	28	12	9.4
(22)	33	7	8.6
(23)	13	27	10.4
(24)	5	35	15.9
(25)	21	19	9.7
(26)	26	14	10.1
(27)	17	23	17.4
(28)	25	15	11.5
(29)	17	23	11.8
(30)	29	11	16.7
Total	214	186	
	53.5%	46.5%	

tant difference involves the generation of continuation rises. Since MITalk parser information was not used directly by the schematic algorithm, continuation rises were generated preceding sentence internal silence. Furthermore, the version of the schematic algorithm used to synthesize the utterances did not anticipate final voiceless segments, so that a phrase-final word ending in voiceless consonants contains no rise (in sentences 25 and 26, for example). Preference for the naturalistic version in sentences 21, 22, and 28 is also most likely due, in part, to differences in continuation rises.

Due to the arbitrary nature of the accent assignment procedure used by the schematic algorithm, several sentences (21, 22, 25, 26, 28, and 30) receive implausible accentuation. Furthermore, "telephone" in sentence 28 is specified with two primary stressed vowels since it is not listed in the lexicon and the compound stress rule in the SOUND1 module of MITalk fails to destress "-phone." However, two sentences (23 and 24) receive more appropriate accentuation by the accent assignment heuristic than by the naturalistic algorithm. One possible consequence of an accentual difference in a pair of utterances may be a difference in the perception of vowel length. For example, "cold" in sentence 3, "cause" in sentence 8, "wall" in sentence 10, "growth" in sentence 25, and "canals" in sentence 27 seem to be of more appropriate duration in the schematic version than in the naturalistic version since they are accented by the former algorithm and not by the latter.

Several pairs of stimuli differ in their $F0$ level. The naturalistic version of sentences 23, 24, 27, and 29 have an $F0$ level which is too high and/or too flat, while the schematic version of sentences 25, 26, and 30 have an $F0$ level which is too low at the end of the phrase.

A summary of the preferences in experiment 3, by subject, is given in Table VIII. There is no significant overall preference for either set of test sentences, but there is a significant correlation between the subject's sex and preference for sentence version: Males prefer the schematic versions and females prefer the naturalistic versions.

IV. CONCLUSION

The major problems for both algorithms, but especially for our implementation of the schematic algorithm, have to do with accent assignment and with the determination of the intonation phrase rather than with the phonetic realization of accent through manipulation of $F0$. Due to parser errors, phrase boundaries are incorrectly identified in 30% of the sentences used in the three experiments. Moreover, since the naturalistic algorithm utilizes a grammatical part-of-speech hierarchy which ranks nouns higher than verbs, incorrect classification of verbs as nouns (the major classification error) results in an unintended accent. Informal evaluation of synthesized utterances indicates that with appropriate phrasing (manually determined), utterances generated by the two algorithms are more similar. This is due, at least in part, to their identical phrasing.

TABLE VIII. Preferences by subject for the naturalistic versus schematic intonation algorithms (experiment 3).

	Naturalistic	Schematic
Females		
LT	14	26
TD	26	14
MB	27	13
SL	29	11
GW	31	9
Subtotal	127	73
Males		
BM	10	30
PD	13	27
TP	20	20
KR	19	21
JL	25	15
Subtotal	87	113
Total	214	186
	53.5%	46.5%

The main basis for distinguishing utterances produced by the two algorithms is their accentual pattern. For the shorter, phonetically balanced utterances used in experiment 1, naturalistic versions are slightly preferred overall. In most utterances for which naturalistic versions were strongly favored, the final noun is inappropriately accented by the schematic algorithm and correctly deaccented by the naturalistic algorithm. However, the results for experiment 2 indicate that deaccenting phrase-final nouns in the schematic versions results in more natural accentuation for some sentences. Therefore, a modified version of the schematic algorithm which simply deaccents the final noun of each phrase should result in greater preference for that algorithm than was found in experiment 1. Liberman and Pierrehumbert (1984) have also addressed this problem of final peak lowering; a solution is incorporated in the synthesizer described by Anderson *et al.* (1984).

V. RECOMMENDATIONS FOR FUTURE WORK

The two most important areas for improving intonation synthesis are in the determination of the intonation phrase and in accent assignment. The determination of the intonation phrase, which is the domain for the specification of accent values, is logically prior to the assignment of accent targets. Although little work has been done on the location of pauses and phonological phrase boundaries in fluent speech (but see Harris *et al.*, 1981 and Kreiman, 1982 for perceptual investigations, and Downing, 1970 for a linguistic perspective), it is reasonable to place pauses and intonation phrase boundaries at the boundaries of major clauses. For those interested in the practical goal of maximizing quality improvement and minimizing research time, we recommend that rather than investing effort in obtaining a consistently correct syntactic analysis, which is a difficult and currently unsolved problem, work should be concentrated on elaborating the heuristic used to determine intonational phrases. In addition, work needs to be done to develop a better accent assignment heuristic, possibly based on a simple discourse model.

The heuristic used in the schematic algorithm for accent assignment may be improved simply by deaccenting repeated nouns, and by deaccenting verbs. One possibility for an improved heuristic is to assign prominence to two syllables, one at the beginning of the sentence, and one at the end. Stressed vowels of content words between the two accents alternately get low- or mid-range accent targets. If the last constituent in the phrase consists of a noun phrase, accent the first prenominal modifier, if possible. However, reference to discourse and semantic information is a prerequisite for significantly more natural acccentuation.

There are additional enhancements which require a discourse component. Some examples of discourse-related modifications which would increase the naturalness of the intonation contour are as follows:

(1) Pitch range should be widened for emphasis.
(2) A pulse register, i.e., lower $F0$ values, should be used for paragraph and discourse endings.
(3) Accented syllables should be lengthened.
(4) Longer pauses should be used between paragraphs than between sentences.
(5) Speech rate should be reduced at pauses and at constituent boundaries.
(6) Final $F0$ should be higher for initial conditional clauses than for initial temporal clauses.
(7) The pitch range of quoted text should be raised.
(8) The pitch range should be widened for main sentences but narrowed and lowered (with lower intensity) for parenthesized text (Bolinger, 1978) and perhaps for subordinate clauses (Lea, 1980).
(9) Longer utterances should have higher initial $F0$ than shorter utterances (Sorensen and Cooper, 1980).
(10) Higher $F0$ should be used at the paragraph beginnings (Lehiste, 1975).
(11) A separate intonation phrase should be used for nonrestrictive relative clauses but not for restrictive relative clauses (Garro and Parker, 1982).
(12) Contrast to preceding text should be indicated by pitch level (Barry, 1981).
(13) Repeated referents having a common reference should be deaccented (Terken, 1981).

Further improvements to our implementation of the schematic algorithm may be based upon optimization of the program variable values, such as the accent target values, the rise and fall times around the accent target (currently 40 cs), the location of the accent peak in time, the topline and baseline values, the starting $F0$ value, continuation rise $F0$ value, terminal $F0$ value, and phrase accent value. Topline and baseline values, for example, do not currently utilize the observation made by 't Hart and Cohen (1973) for Dutch, that the declination line slope should decrease 3% every 100 ms at the beginning of sentences, but should gradually decrease to a fixed value of 0.5% per 100 ms after 5 s, although similar results are observed for English (Pierrehumbert, 1979). Similarly, specification of accent target values may best be specified on a logarithmic scale (Fujisaki, 1981).

Phonetic conditioning effects on $F0$ values, which are modeled well by the naturalistic algorithm, should be tested in the context of the schematic algorithm. Our experiments show only a small naturalness advantage to be gained by including phonetic conditioning effects; that advantage could be measured more precisely via a controlled experiment in which phonetic conditioning effects are added one at a time to the schematic algorithm. Potentially useful phonetic conditioning effects include segmental conditioning factors, for example, that $F0$ values following voiceless consonants are relatively higher than those following voiced consonants (Lehiste and Peterson, 1961; Lea, 1973; Mattingly, 1966; and Haggard *et al.*, 1970). While this difference in $F0$ may not be important for perception of the accent pattern, it probably is important for perception of the voicing distinction. However, intrinsic effects for vowels, for example, higher $F0$ values associated with phonetically higher vowels, are not observed in fluent readings (Umeda, 1981). The determination of $F0$ values at phrase boundaries needs to be modified in our implementation of the schematic algorithm to account for voiceless segments, so that realized initial and final values are equal to appropriate target values. Furthermore, Pierrehumbert (1979) cites experiments which suggest that amplitude downdrift is important for the perception of phrasing. Finally, introducing a small amount of irregularity ("jitter") in the $F0$ values is reported to result in increased naturalness (Lehiste, 1970; Rozsypal and Millar, 1979), although one experiment did not confirm this result (O'Shaughnessy, personal communication, 1982).

Although we feel that detailed phonetic conditioning effects on $F0$ such as those described in the preceding paragraph are likely to yield some improvement in naturalness, our results indicate that such improvements would be of a secondary nature. The schematic algorithm, which is simpler than the naturalistic one because it does not attempt to model phonetic conditioning effects, performs almost as well as the latter. Assuming that the small overall preference for the naturalistic algorithm is due to such phonetic conditioning effects, we would be better off devoting our efforts toward a more effective determination of intonation phrase boundary locations and accent assignment than toward a more precise model of detailed phonetic conditioning effects on $F0$.

ACKNOWLEDGMENTS

The authors would like to acknowledge the helpful interactions they have had with the following people: Janet Pierrehumbert and Mark Liberman (both of A. T. & T. Bell

Laboratories), Douglas O'Shaughnessy (INRS-Telecommunications), and Paul Mermelstein (Bell-Northern Research and INRS-Telecommunications). They also thank Peter Doehring (McGill University) for his assistance in testing and statistical analysis.

APPENDIX A: SENTENCES SYNTHESIZED IN EXPERIMENTS 1 AND 2

(1) The goose was brought straight from the old market.
(2) The sink is the thing in which we pile dishes.
(3) A whiff of it will cure the most stubborn cold.
(4) The facts don't always show who is right.
(5) She flaps her cape as she parades the street.
(6) The loss of the cruiser was a blow to the fleet.
(7) Loop the braid to the left and then over.
(8) Plead with the lawyer to drop the lost cause.
(9) Calves thrive on tender spring grass.
(10) Post no bills on this office wall.

(11) A yacht slid around the point into the bay.
(12) The two met while playing on the sand.
(13) The ink stain dried on the finished page.
(14) The walled town was seized without a fight.
(15) The lease ran out in sixteen weeks.
(16) A tame squirrel makes a nice pet.
(17) The horn of the car woke the sleeping cop.
(18) The heart beat strongly and with firm strokes.
(19) The pearl was worn in a thin silver ring.
(20) The fruit peel was cut in thick slices.

APPENDIX B: SENTENCES SYNTHESIZED IN EXPERIMENT 3

(21) Sentences with more than one accent invoke relationships not only between each accent and the sentence as a whole, but among the accents themselves.

(22) Many returning tourists have complained about minimal hotel services where their rooms went uncleaned and the beds unmade for days.

(23) Just as present technology had to await the explanations of physics, so one might expect that social invention will follow growing sociological understanding.

(24) The fundamental aim in the linguistic analysis of a language is to separate the grammatical sequences which are sentences of the language from the ungrammatical sequences which are not sentences of the language and to study the structure of the grammatical sequences.

(25) This reflects the company's commitment to the vital role of research and development, as the engine that drives our growth and creates the future for all of us.

(26) This led in 1973 to the publication of a Revised Report and a definition of a language representation in terms of the ISO character set.

(27) Travelers entering from the desert were confounded by what must have seemed an illusion: a great garden filled with nightingales and roses, cut by canals and terraced promenades, studded with water tanks of turquoise tile in which were reflected the glistening blue curves of a hundred domes.

(28) If you are having difficulty in resolving a problem, please call our business office where a service representative has a record of your telephone service and will be pleased to assist you.

(29) The speech which you have just heard was produced during June, 1982 at the BNR/INRS computer laboratory on Nuns' Island by a text-to-speech synthesis system.

(30) While it is significant that the common stock of Northern Telecom has outperformed the markets in both Canada and the United States over the last year, it is even more significant that it has rebuffed downward trends in price movements of many other high-tech stocks in the US.

[1] Anderson *et al.* (1984) and Liberman and Pierrehumbert (1984) report on a more recent version of Pierrehumbert's earlier algorithm. The more recent algorithm discussed in these two works has corrected certain problems noted in our conclusions, in particular regarding final peak lowering, micromelody, and timing of the final boundary tone. The following discussion will be of the earlier algorithm. While the more recent version may produce better intonation curves, the earlier algorithm also made strong and interesting claims and will be allowed to stand on its own merit.

[2] The MITalk-79 system was used with permission of MIT.

Abe, I., and Kanekiyo, T., Eds. (**1965**). *Forms of English: Accent, Morpheme, Order* (Harvard U. P., Cambridge, MA).

Anderson, M.D., Pierrehumbert, J., and Liberman, M.Y. (**1984**). "Synthesis by Rule of English Intonation Patterns," Proc. 1984 IEEE Int. Conf. Acoust. Speech Signal Process., 2.8.1–2.8.4.

Barry, W. J. (**1981**). "Prosodic Functions Revisited Again!," Phonetica **38**, 320–340.

Bolinger, D. (**1951**). "Intonation: Levels Versus Configurations," Word **7**, 199–210.

Bolinger, D. (**1972a**). "Accent is Predictable (If You're a Mind-Reader)," Language **48**, 833–844.

Bolinger, D. (**1972b**). "Relative Height," in *Intonation*, edited by D. Bolinger (Penguin, Harmondsworth).

Bolinger, D. (**1978**). "Intonation Across Languages," in *Universals of Human Language 2: Phonology*, edited by J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (Stanford U. P., Stanford, CA), pp. 471–524.

Clark, J. E. (**1981**). "A Low-Level Speech Synthesis System," J. Phon. **9**, 451–476.

Collier, R., and 't Hart, J. (**1975**). "The Role of Intonation in Speech Perception ," in *Structure and Process in Speech Perception*, edited by A. Cohen and S.G. Nooteboom (Springer-Verlag, Berlin).

Crystal, D. (**1969**). *Prosodic Systems and Intonation in English* (Cambridge U. P., Cambridge, England).

Crystal, D. (**1975**). *The English Tone of Voice* (Arnold, London).

Daneš, F. (**1960**). "Sentence Intonation from a Functional Point of View," Word **16**, 34–54.

Downing, B.T. (**1970**). "Syntactic Structure and Phonological Phrasing in English," Ph.D. thesis, University of Texas (unpublished).

Fujisaki, H. (**1981**). "Dynamic Characteristics of Voice Fundamental Frequency in Speech and Singing—Acoustical Analysis and Physiological Interpretations," Fourth F.A.S.E. Symposium, Venice, Italy.

Garro, L., and Parker, F. (**1982**). "Some Suprasegmental Characteristics of Relative Clauses in English," J. Phonet. **10**, 149–161.

Goldsmith, J. (**1976**). "An Overview of Autosegmental Phonology," Linguist. Anal. **2**, 23–68.

Haggard, M., Ambler, S., and Callow, M. (**1970**). "Pitch as a Voicing Cue," J. Acoust. Soc. Am. **47**, 613–617.

Halliday, M. A. K. (**1967**). *Intonation and Grammar in British English* (Mouton, The Hague).

Harris, M.O., Umeda, N., and Bourne, J. (**1981**). "Boundary Perception in Fluent Speech," J. Phonet. **9**, 1–18.

Hockett, C.F. (**1955**). *A Manual of Phonology* (Indiana Univ. Publ. Anthropol. and Linguist. **11**, Baltimore).

IEEE. (**1969**). "IEEE Recommended Practice for Speech Quality Measurements," IEEE Trans. Audio Electroacoust. **AU-17** (3), 225–246.

Kreiman, J. (**1982**). "Perception of Sentence and Paragraph Boundaries in

Natural Conversation," J. Phonet. **10**, 163–175.

Lea, W. A. (**1973**). "Segmental and Suprasegmental Influences on Fundamental Frequency Contours," in *Consonant Types and Tones*, edited by Larry M. Hyman [Southern California Occasional Papers in Linguistics (1), University of Southern California, Los Angeles], pp. 17–70.

Lea, W. A. (**1980**). "Prosodic Aids to Speech Recognition," in *Trends in Speech Recognition*, edited by Wayne A. Lea (Prentice Hall, Englewood Cliffs, NJ), pp. 166–204.

Leben, W. R. (**1976**). "The Tones in English Intonation," Linguist. Anal. **2**, 69–107.

Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA).

Lehiste, I. (**1975**). "The Phonetic Structure of Paragraphs," in *Structure and Process in Speech Perception*, edited by A. Cohen and S.G. Nooteboom (Springer-Verlag, Berlin), pp. 195–203.

Lehiste, I., and Peterson, G. E. (**1961**). "Some Basic Considerations in the Analysis of Intonation." J. Acoust. Soc. Am. **33**, 419–425.

Liberman, M. Y. (**1975**). "The Intonational System of English," Ph.D. thesis, MIT (Indiana University Linguistics Club, Bloomington, IN).

Liberman, M. Y., and Pierrehumbert, J. (**1984**). "Intonational Invariance Under Changes in Pitch Range and Length," in *Language Sound Structure*, edited by M. Aronoff and R.T. Oehrle (MIT, Cambridge, MA).

Liberman, M. Y., and Prince, A. (**1977**). "On Stress and Linguistic Rhythm," Linguist. Inquiry **8**, 249–336.

Maeda, S. (**1976**). "A Characterization of American English Intonation," Ph.D. thesis, MIT (unpublished).

Mattingly, I. G. (**1966**). "Synthesis by Rule of Prosodic Features," Lang. Speech **9**, 1–13.

Nakatani, L. H., and Schaffer, J. A. (**1978**). "Hearing 'Words' Without Words: Prosodic Cues for Word Perception," J. Acoust. Soc. Am. **63**, 234–245.

Olive, J. P., and Nakatani, L.H. (**1974**). "Rule-Synthesis of Speech by Word Concatenation: A First Step," J. Acoust. Soc. Am. **55,** 660–666.

O'Shaughnessy, D. (**1982**). Personal communication.

O'Shaughnessy, D. (**1976**). "Modelling Fundamental Frequency and its Relationship to Syntax, Semantics, and Phonetics," Ph.D. thesis, MIT (unpublished).

O'Shaughnessy, D. (**1979**). "Linguistic Features in Fundamental Frequency Patterns," J. Phonet. 7, 119–145.

Pierrehumbert, J. (**1979**). "The Perception of Fundamental Frequency Declination," J. Acoust. Soc. Am. **66**, 363–369.

Pierrehumbert, J. (**1980**). "The Phonology and Phonetics of English Intonation," Ph.D. thesis, MIT (unpublished).

Pierrehumbert, J. (**1981a**). "Synthesizing Intonation," J. Acoust. Soc. Am. **70**, 985–995.

Pierrehumbert, J. (**1981b**). "Synthesizing English Intonation," Proc. GALF Prosody Symp. 29–30 May 1981, University of Toronto, pp. 179–189.

Pike, K. (**1945**). *The Intonation of American English* (Univ. Michigan P., Ann Arbor, MI).

Rozsypal, A. J., and Millar, B. F. (**1979**). "Perception of Jitter and Shimmer in Synthetic Vowels," J. Phonet. **7**, 343–355.

Sorensen, J. M., and Cooper, W. E. (**1980**). "Syntactic Coding of Fundamental Frequency in Speech Production," in *Perception and Production of Fluent Speech*, edited by R. A. Cole (Erlbaum, Hillsdale, NJ).

Terken, J. M. B. (**1981**). "The Distribution of Pitch Accents as a Function of Informational Variables II," IPO Annu. Prog. Rep. **16**, 39–43.

't Hart, J., and Cohen, A. (**1973**). "Intonation by Rule: a Perceptual Quest," J. Phonet. **1**, 309–327.

Trager, G. L., and Smith, H. L. (**1957**). "An Outline of English Structure," American Council of Learned Societies, Washington, DC.

Umeda, N. (**1981**). "Influence of Segmental Factors on Fundamental Frequency in Fluent Speech," J. Acoust. Soc. Am. **70**, 350–355.

Wells, R. (**1945**). "The Pitch Phonemes of English," Language **21**, 27–39.

Witten, I. H. (**1977**). "A Flexible Scheme for Assigning Timing and Pitch to Synthetic Speech," Lang. Speech **20**, 240–260.