Eva Gårding and Arthur S. Abramson

A STUDY OF THE PERCEPTION OF SOME AMERICAN ENGLISH INTONATION CONTOURS[1]

This is a report of some exploratory work on intonation where the experimenters used a machine called the Intonator. The Intonator, designed and built at Haskins Laboratories, is an electronic device, which makes it possible to impose artificial pitch contours on natural, recorded speech (Borst and Cooper, 1957).[2]

Outline and scope

In the earlier part of the paper we are concerned with the problem of finding some American English intonation contours which can be identified by native speakers on the basis of fundamental frequency alone. The later part describes our experiments with these contours.[3]

The experiments were designed with the purpose of determining to what extent it is possible to change the pitch-curve of a given contour before it starts sounding like a different one. The question of "same" and "different" was to be answered by means of controlled listening

[1] An abbreviated version of this paper was read before the Seventy-Fifth Annual Meeting of the Modern Language Association of America, Philadelphia, Pennsylvania, December 28, 1960. The experiments were carried out in two periods of three months in 1959 and 1960. At the Fifth International Congress of Phonetic Sciences work along similar lines was reported by Magdics, Isačenko and Hadding-Koch. In the hope that our results may be of interest to other workers we decided to publish them rather than await an opportunity to complete our investigation.

[2] As the speech is passed through an 18-channel Vocoder, it gets a somewhat machine-like quality. This and other limitations in the Intonator makes it perhaps better suited for an investigation of gross rather than fine features of intonation. Better instrumentation (Cooper, 1964) will now make it possible to check the results reported here and carry on the work with better control.

[3] A.S.A. is mainly responsible for the earlier part of the project, E.G. for the later.

tests. A comparison of possible changes in several contours was expected to give us—ultimately—some guidance in the choice of parameters for the description of intonation.

Procedure in choice of contours

To get a collection of naturally spoken intonation contours we made up a carrier sentence "Five thousand six hundred ten" and a description of five situations. A number was chosen because of its lack of emotional connotations. The sentence was uttered twice for each situation by five native speakers of American English with some but not striking dialect diversity. There were three women and two men.

Approximate labels for the situations described to our informants were the following.

1. Neutral statement
2. Yes or no question
3. Anger
4. Delighted surprise
5. Counting in a series

As could be expected the informants did not respond uniformly to the five situations. The highest agreement was found in situations 1 and 2. The greatest variation was found in the emotionally loaded situations 3 and 4. We eliminated those utterances that in our judgement were unique as to type and those that had been judged inappropriate by the speakers themselves. This left us with 27 renditions of the sentence "Five thousand six hundred ten". The magnetic tape segments containing these utterances were fixed to cards for the card-reader. A card-reader is a tape-playback device which permits a quick change from one tape-segment to another and therefore facilitates comparison between utterances.[1]

Test I

Five subjects were asked to sort the 27 utterances into groups of intonation types. All reference to the elicitory situations and semantic labels was excluded from the test directions.

[1] Ellamac Language Master, Ellamac Inc., Chicago Ill. (Cooper, 1964).

The subjects classified the utterances, setting up from eight to twelve classes. We then looked at the intersections of test responses to find classes on which at least four out of five subjects agreed. Out of the original 27 utterances only 13 were grouped into such classes, yielding five apparently distinct intonation types with more utterances representing some categories than others. These five types, labeled A, B, C, D, E, are, in the judgement of the experimenters, appropriate to the elicitory situations 1, 2, 3, 4, 5.

At this stage of our work we wanted to see if there would be general agreement on the "sameness" of the utterances that had been grouped together and to obtain information on degrees of similarity between them. The thirteen cards were reduced to ten with two representing each category.

Test 2

A subject was given one of the two versions of A as a model and was asked to rank the nine remaining utterances in order of similarity to it. This was repeated for each of the five situations with eleven subjects. The test was administered to one person at a time.

Using the data generated by this test we set up the notion of "distance" between utterances in the following way. The utterance thought most similar to a model was a distance of "one" from it; the utterance that was judged as least similar to a model was a distance of "nine" from it. Averaging over the subjects and the five test situations we obtained an overall distance for each pair of utterances.

Fig. 2, p. 65, shows the rank order of distances between contours. The average distance between two versions of the same situation is just over two. The next smallest distance is that between B-utterances and E-utterances, about four. These contours have a final rise as their common feature. The greatest distance is between A and D, objective statement and delighted surprise, about eight. Their most conspicuous contrastive feature is a fall in A from a mid level against a complex pitch movement at a high level in D.

The results of this test support our assumption that the two versions of each contour were the same, i.e. variants of the same contour.

On the usual assumption that pitch patterns provide the most important cues to intonation types, narrow band spectrograms were used to trace the fundamental frequency movements of the ten utterances used in Test 2. The average of the two curves for each category is presented in Fig. 1. For this purpose women's voices were reduced in fundamental frequency by half to bring them into the male range and the two utterances were aligned in time, syllable by syllable. Solid lines representing actual measurements are connected by dotted lines to show the continuity of speech.

Test 3

In order to determine whether the average pitch patterns we had arrived at contained enough cues to distinguish the five intonation contours we turned to speech synthesis. A sequence of English nonsense syllables was constructed /kal guləv ʃæt fobəl kʌn/ with the same number and patterning of stressed and unstressed syllables as in our original sentence. We also ended the final syllable in a nasal as in the original sentence. Our purpose in preparing this sequence was to have a carrier for the synthetic contours that was devoid of lexical content. A male speaker recorded this carrier sentence on a monotone to make it reasonably neutral as to concomitant features.

Each of the average pitch patterns (Fig. 1) was put on the carrier sentence, syllable by syllable, by means of the Intonator.[1] The five resulting utterances together with a sixth on a monotone of 130 cps were recorded on tape and put on cards appropriate for the card-reader.

Twelve subjects were asked, one at a time, to compare the synthetic patterns with the real ones. The subjects were to decide which of the synthetic patterns matched the standards. After each comparison the cards were shuffled to provide the possibility of assigning the same synthetic pattern to more than one standard. Finally any synthetic card left unassigned was to be matched against all five real-speech cards and assigned to the category of one of them.

There was 100 % agreement on the matching of the real and synthetic cards for Contours A, B, C, and D and 37 % for E. Since for the most part E was merged with B, we believe that the difference in fundamental frequency between our intonator versions of these two contours

[1] This method was used by Abramson (1962) in an investigation of tones in Thai. Each tone was imposed on each member of a set of five tonally differentiated words. Pitch contours were found to override all concomitant features. Eliciting responses from Thai subjects was much more straightforward, since the tones of their language have a clearly phonemic status.

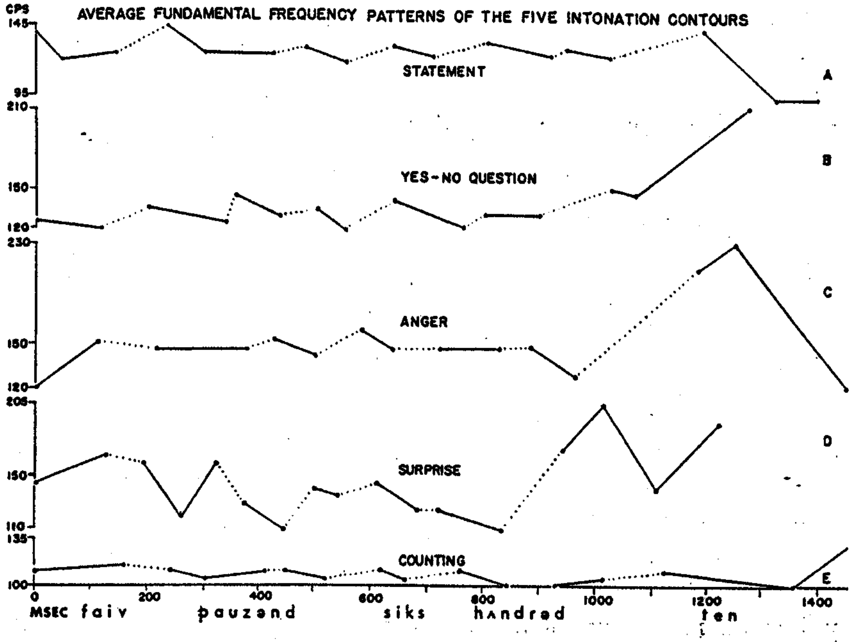

AVERAGE FUNDAMENTAL FREQUENCY PATTERNS OF THE FIVE INTONATION CONTOURS
CPS
145
95
STATEMENT
A
210
150
120
YES-NO QUESTION
B
230
150
120
ANGER
C
205
150
110
SURPRISE
D
135
100
COUNTING
E
0
200
400
600
800
1000
1200
1400
MSEC
faiv
pauzənd
siks
hʌndrəd
tɛn

Fig. 1

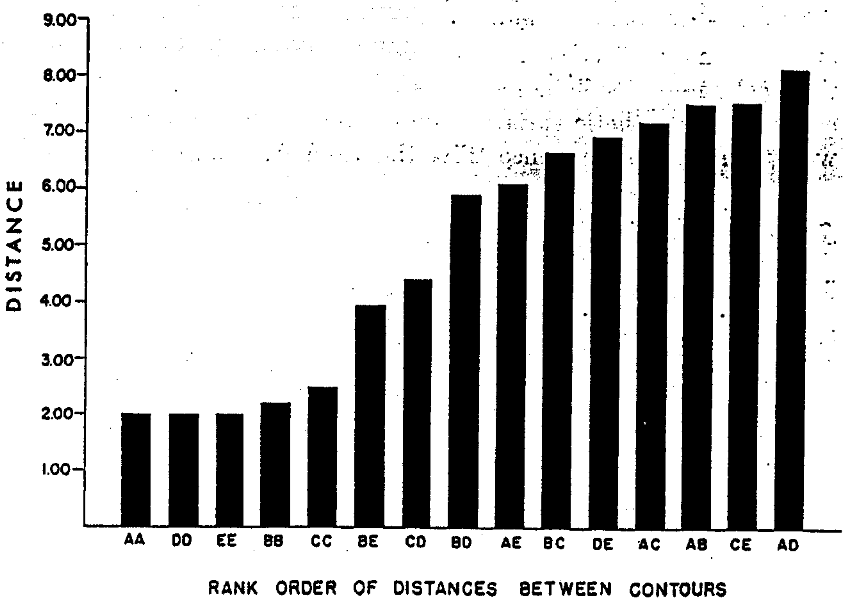

9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
DISTANCE
AA
DD
EE
BB
CC
BE
CD
BD
AE
BC
DE
AC
AB
CE
AD
RANK ORDER OF DISTANCES BETWEEN CONTOURS

Fig. 2

was not great enough to make them distinct. We suspect that the pitch distinction between B and E in the real speech utterances was supported by differences in intensity and duration. The synthetic monotone was assigned to E 90 % of the time.

Still, four out of five contours were easily distinguishable on the basis of pitch alone. The exploration of the pitch ranges of the curves could begin.

Reactions to changes in contours

Simplifications

In the production of the average synthetic contours for Test 3 the real speech contours had been aligned syllable by syllable. Judging from the results of this test minor temporal changes that had been brought about by this procedure did not have any effect on the identifications of A, B, C, and D.

Another simplification that had passed unnoticed in Test 3 was that points where measurements had been made for the average contours had been joined by straight lines instead of smoothly undulating lines, which are characteristic of pitch patterns in natural speech.[1,2]

In order to make a comparison between the essential features of our contours we decided to make the precontour (everything before the final, stressed syllable) a monotone of 130 cps and as a start concentrate our synthetic variations on the final syllables. For these tests we returned to the sequence "Five thousand six hundred ten".

Schema of variations

The intonation of the final syllables has at least one turning point and an endpoint. We decided to move these points up and down the pitch scale in steps of 10 cps. In this way we should be able to study the effect of a change in the pitch interval of the last pitch movement. This change would at the same time affect the rate of the pitch movement.

[1] Delattre (1963) has pointed out that the shape of a final curve is typically concave in English, German, and Spanish but convex in French.

[2] Isačenko and Schädlich (1963 and 1964) have demonstrated how they can approximate the syntactic function of German sentence-intonation using two levels of piecewise constant fundamental frequency.

The shape of the pitch movement was also varied by introducing new turning points. This was an attempt to examine the effect on perception of the "hooks" (as seen in spectrograms) which so often accompany major pitch movements.

Testing

In our testing we used the ABX technique in the following way. Two contours were set up as models. A number of intermediate curves were generated on the Intonator. These synthetic contours were then used as stimuli in a test. Listeners were asked to compare each stimulus to the two models. The stimulus could be judged either as the same or most like one of the models.

This testing technique was used with the following contours as models:

A and B. (Test 4. Fig. 4, p. 68, 11 generated contours. 14 subjects.[1])
B and D. (Test 5. Fig. 5, p. 75. 15 generated contours. 11 subjects.)
A and C. (Test 6. Fig. 6 a, b, c, d, pp. 76, 77. 29 generated contours. 10 subjects.)
A and E. (Test 7. Fig. 7 a, b, p. 78. 19 generated contours. 10 subjects.)

Test 4 will be described in detail.

Test 4

The opposition between A and B was the most obvious starting point for a study of cross-over points between contours. The difference between them seemed to reside mainly in the pitch movement of the last syllable (cf. Fig. 1). The pitch movement in this syllable was further simplified in that the level beginning of B and the level end of A were merged with the general pitch movement of the syllable. This change did not have any noticeable effect. Fig. 3, p. 68, shows these simplified versions of A and B, both with an identical precontour on a monotone of 130 cps. The end point of contour A was now moved up in steps of 10 cps until it reached 200 cps, thus making eleven stimuli which were recorded three times. The 33 test items were put in random order to form Test 4. The stimuli were presented to the listeners in groups of

[1] The subjects were employees of the Haskins Laboratories.

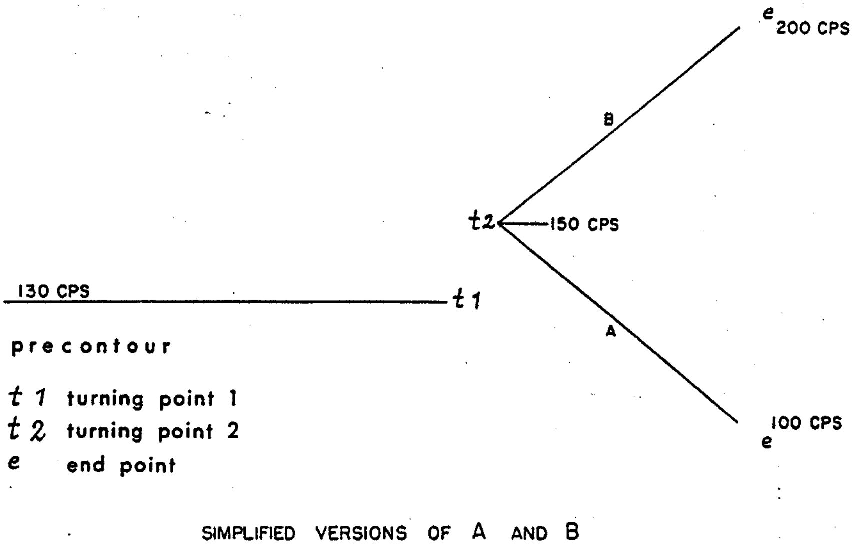

e 200 CPS
B
t2 150 CPS
130 CPS
t1
A
e 100 CPS
precontour
t 1 turning point 1
t 2 turning point 2
e end point
SIMPLIFIED VERSIONS OF A AND B

Fig. 3

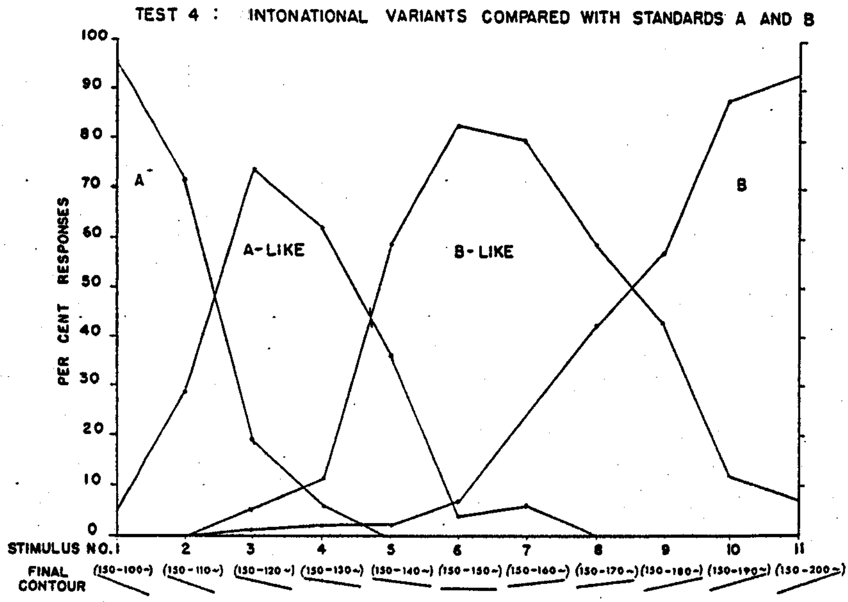

TEST 4 : INTONATIONAL VARIANTS COMPARED WITH STANDARDS A AND B
PER CENT RESPONSES
100
90
80
70
60
50
40
30
20
10
0
A
A-LIKE
B-LIKE
B
STIMULUS NO. 1 2 3 4 5 6 7 8 9 10 11
FINAL CONTOUR
(150-100~) (150-110~) (150-120~) (150-130~) (150-140~) (150-150~) (150-160~) (150-170~) (150-180~) (150-190~) (150-200~)

Fig. 4

three, each stimulus ultimately taking all three positions in the group. The two models A and B from Test 3 were heard before each group. (The placement of the stimulus in regard to the models turned out not to affect the results.)

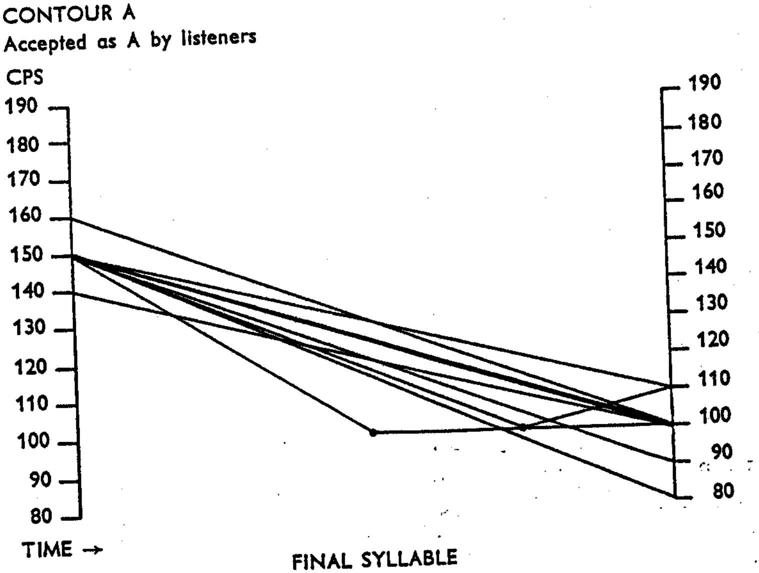
The 14 subjects who took the test were to compare each stimulus with the models and decide whether it was the same intonation contour as one of them or, if not, which of the two it was more like. The possibility of assigning a stimulus to neither A nor B by simply indicating greater similarity to one of them was allowed because of the chance of our having gone through the range of some third category.

Fig. 4 gives the percentage of votes for each of the four alternative responses to each stimulus, A, A-like, B, B-like. The following observations can be made.

The graph shows a striking skewness, the votes for A and A-like being outnumbered by the votes for B and B-like (B and B-like 290, A and A-like 172 votes). The cross-over point between A, A-like and B, B-like is between stimuli 4 and 5. No A-votes are to be found from 5 through 11. B-votes have a wider spread and are lacking in only two stimuli, 1 and 2. A final drop of more than 10 cps is needed to make the listeners liken a stimulus to a standard with falling pitch. Stimulus 5, for instance, has a final fall from 150 to 140 cps. Still it is judged as B-like, rising, 59 % of the time. Stimulus 6 with a level final contour is considered B-like in 83 % of the judgements. This recalls one of the results of Test 3. A monotone was assigned 90 % of the time to Category E, which has a rising final pitch. The predominance of votes for B and B-like suggests the following conclusion. The contrast between a question contour and a statement contour may very well be a contrast of level versus fall rather than a contrast of rise versus fall.[1]

The following considerations have to be kept in mind, however.
1. The placement of the precontour may have preconditioned the responses in favour of B. The turning point that is more significant to perception is perhaps 1 rather than 2 (see Fig. 3). Consequently we may need a drop below the comparatively long precontour on 130 cps to perceive the whole contour as falling.
2. The intensity curve of the carrier phrase (an utterance pronounced on a monotone) may have fitted a level or rising pitch contour better than a falling one, thus giving strong support to B and B-like responses.

[1] The same dichotomy has been pointed out by Uldall (1963).



• shows location of additional turning point. One is the maximum number for each contour. Thick line represents original contour.

Summary of results

Collecting all the information we gathered for each contour about various modifications accepted by subjects in Tests 4, 5, 6, and 7 we will now give a summary of our results. At this stage of our work we know more about the range of Contour A than about the others. We define the range of a contour to be all modifications (in a given schema) which are identified or accepted as that contour.

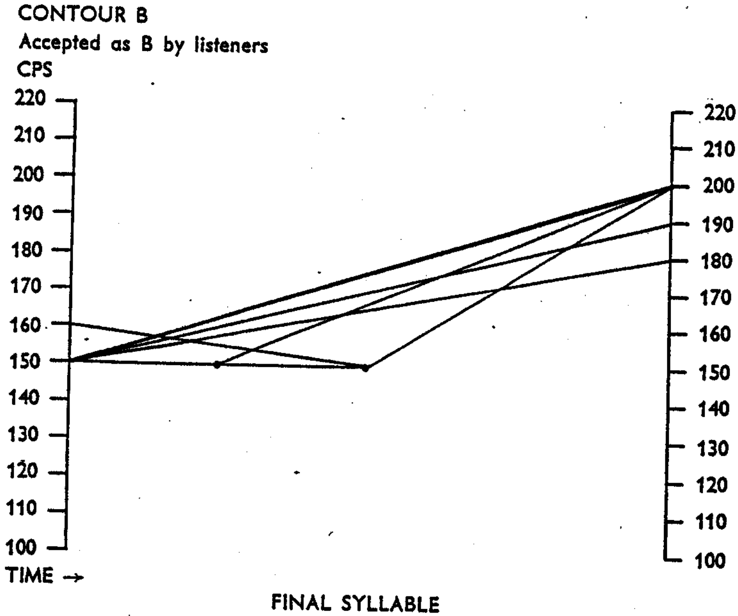
The above figure shows all the synthetic variants that were accepted as A. Their pertinent features will be described in terms of shape, size of interval[1], and level.

CONTOUR A

Shape: A falling pitch movement from a mid level.

Size of interval: 40—70 cps. The end point can be moved up 10 cps (Stimulus 2, Test 4, p. 68) and moved down at least 20 cps (Contours 1 and 2, Test 7, p. 78). The beginning point of the final syllable can be

[1] *Size of interval* because *range* is used in a different meaning in this paper.



moved 10 cps upwards (Contour 6, Test 6, p. 76) and 10 cps downwards (Contour 2, Test 6).

Level: The level from which A starts is more important than the size of the interval. A stimulus with the interval of A starting from the level of C (Contour 11, Test 6) will be called C-like.

Changes of shape: If a turning point is introduced one fourth of the way from the beginning of the final syllable above A (Contours 24, 25, 26, Test 6) the curve is no longer identified with A. An additional turning point in the middle of the syllable makes listeners say that the curve is like C rather than like A (Contours 27, 28, 29, Test 6). This additional turning point then makes the contour more like the emotional model than the objective one. The explanation may be that a sustained mid level is perceived as a high level. An additional turning point one fourth of the way from the end or in the middle of the syllable below A does not disturb the identification of A (Contours 4 and 14, Test 7, p. 78).

CONTOUR B

Shape: A rising pitch movement.

Size of interval: At least 30 cps. The end point can be lowered 20 cps

(Stimuli 9 and 10, Test 4, p. 68). Contours with the end point lowered from 30 to 60 cps are judged as B-like as was reported under Test 4. An end point lowered 50 cps, i.e. a change producing a contour with a level pitch movement, is called B-like in more than 80 % of the judgements.

Level: In Test 3 Contour E was merged with B. One of the differences in pitch between them is that E starts from a low level. This difference in level then seems to be insignificant.

Changes of shape: A new turning point can be introduced one fourth of the way from the beginning of the syllable below B (Contour 2, Test 5, p. 75) and the curve will still be identified with B. A new turning point in the middle of the syllable (Contours 3 and 4, Test 5) reduces the B-votes by 30 %. The portion before the new turning point can be moved upwards 30 cps (Contour 6, Test 5) before listeners stop calling it B.

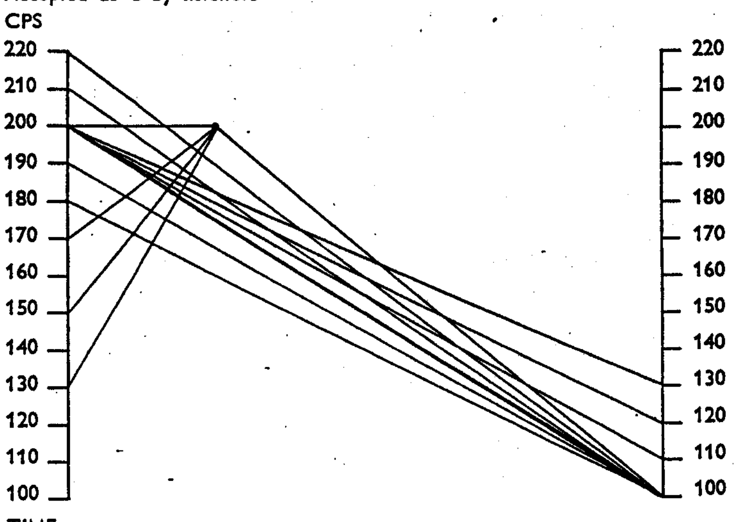
CONTOUR C

Shape: A falling pitch movement from a high level.

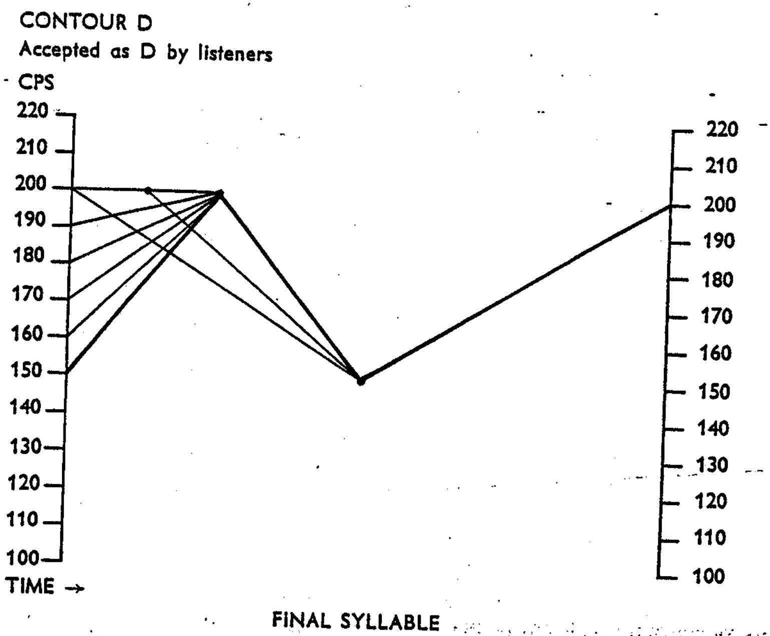
Size of interval: From 70 to at least 120 cps. The end point can be raised 30 cps (Contours 13, 14, 15, Test 6, p. 76). The beginning point can be moved up at least 20 cps (Contours 9, 10, Test 6).

CONTOUR C

Accepted as C by listeners



FINAL SYLLABLE



Level: The level from which C starts is more important than the interval (cf. Contour A).

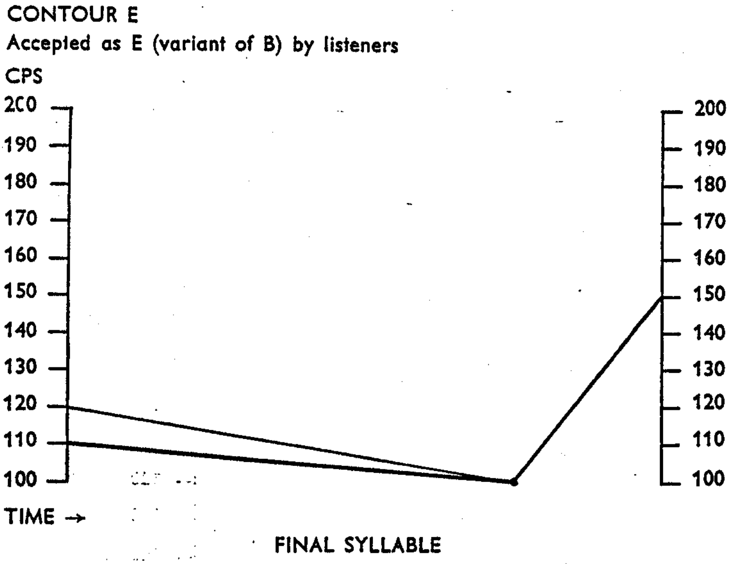
Changes in shape: A level or rising beginning which occupies one fourth of the duration of the final syllable above C can be omitted (Contours 18 and 19, Test 6). Even if this initial rise starts from a much lower level (e.g. at 130 cps as in Contour 16, Test 6, p. 76) the contour is still identified with C.[1] It should be noted that a similar change in A takes it out of the A-category. We can conclude that C has a larger range than A. The range of a contour is probably proportional to the amount of pitch movement that is involved in it.

CONTOUR D

Shape: A falling-rising pitch movement. No experiments have been made to explore the size of the interval and the level of D.

Changes of shape: The initial rise of the pitch movement in the final syllable can be omitted and the curve is still identified with D (Contour 8, Test 5, p. 75). In fact, the identification score improves by this change. Thus, the contour which physically has three movements can

[1] In synthesizing American intonation Delattre (1963) has found that a pre-slope rise can be omitted.



as far as perceptual relevance is concerned be described in terms of two movements, falling and rising.[1] Pitting this falling-rising variant of D against B (Contours 3, 4, 5, 6, 7, Test 5, p. 75) we find that the falling beginning of D can start from at least 170 cps before listeners say that it is like B rather than like D.

CONTOUR E

Already in Test 3 it was shown that the pitch difference between E and B was not great enough to make listeners keep them apart. This result indicates perhaps that the falling-rising pitch movement of the last syllable in E is not noticed, leaving the general rise between the beginning and the end as its characteristic feature. Test 7 gives some support to the hypothesis that E is perceived as a rising contour. The general reaction to the series of contours 9 through 12 is a case in point. In these stimuli the rising portion of the pitch-curve at the end is identical with that of E and kept constant; the falling former portion has been given various beginning points. As long as the end point is higher than the beginning point the contour is judged as identical with or (with decreasing beginning point) most like E, but listeners cross over to A when the beginning point is given the same pitch as the end point.

[1] If the duration of the final syllable had been longer, the result may of course have been different.

General conclusions

Our results indicate that each contour has a considerable margin within which changes can be made without any effect on perception, as long as these changes do not disturb a certain pattern. Using three pitch levels to mark the significant turning points and end points of the fundamental frequency, our original contours can be transcribed in the following way:

Tentative pattern for A: Five thousand six hundred ten. (2 ... 2 1)

" " " B: Five thousand six hundred ten. (2 ... 2 3)

" " " C: Five thousand six hundred ten. (2 ... 3 1)

" " " D: Five thousand six hundred ten. (2 ... 3 2 3)

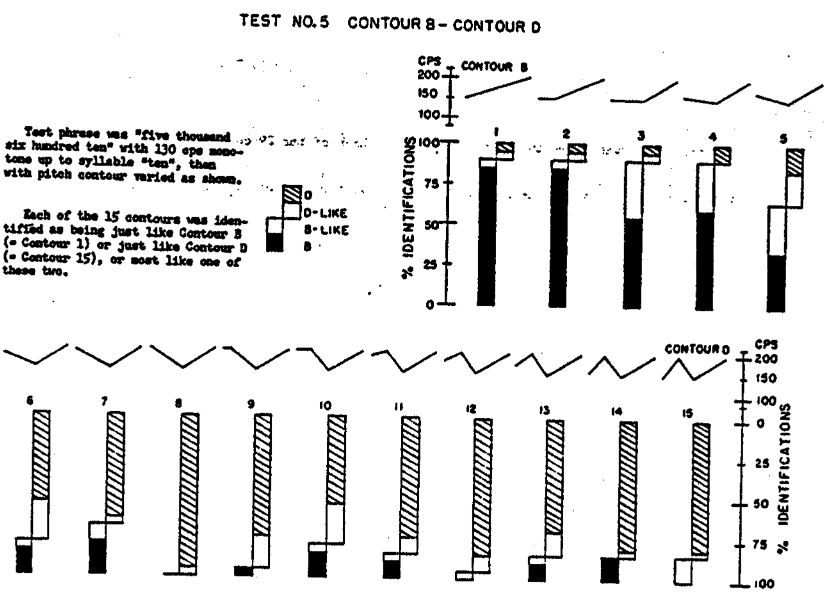


Fig. 5

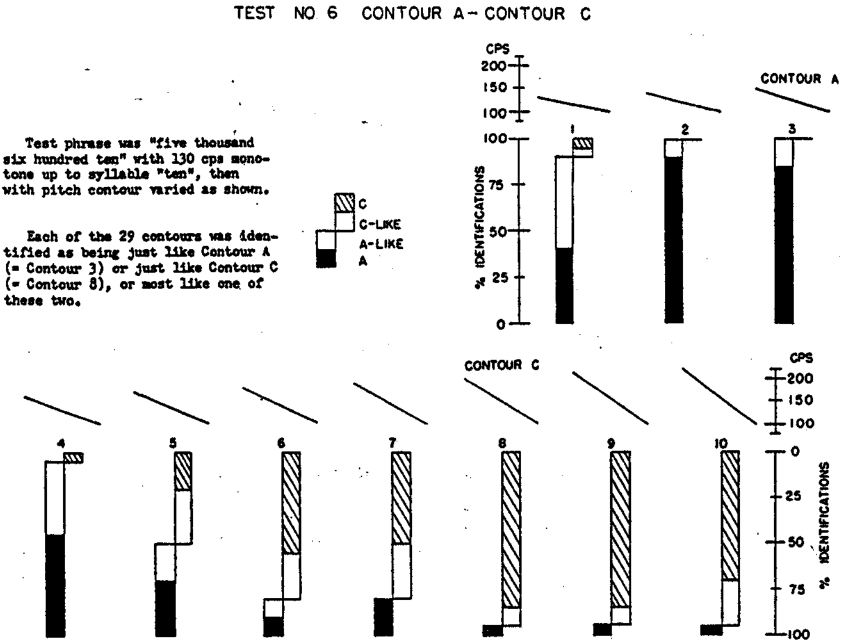

TEST NO. 6 CONTOUR A – CONTOUR C
Test phrase was "five thousand six hundred ten" with 130 cps monotone up to syllable "ten", then with pitch contour varied as shown.
Each of the 29 contours was identified as being just like Contour A (= Contour 3) or just like Contour C (= Contour 8), or most like one of these two.
C
C-LIKE
A-LIKE
A
CPS
200
150
100
CONTOUR A
CONTOUR C
% IDENTIFICATIONS
100
75
50
25
0
1
2
3
4
5
6
7
8
9
10

Fig. 6 a

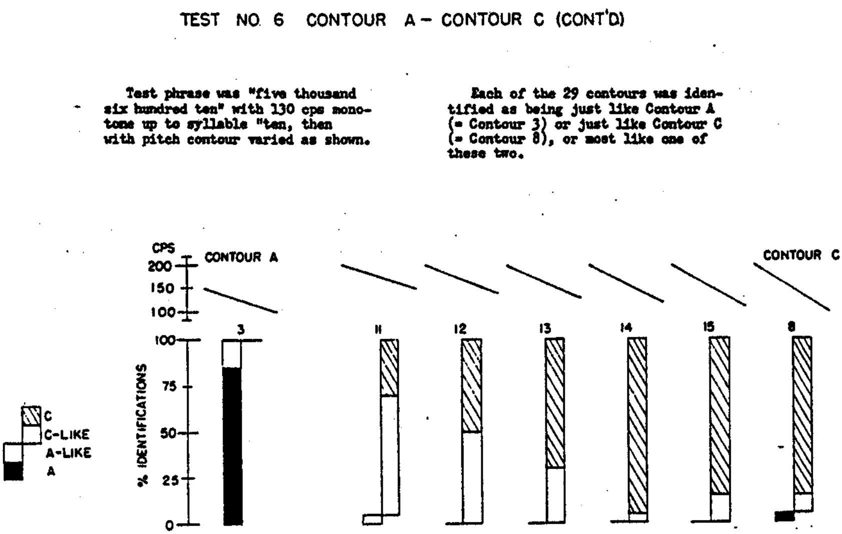

TEST NO. 6 CONTOUR A – CONTOUR C (CONT'D)
Test phrase was "five thousand six hundred ten" with 130 cps monotone up to syllable "ten, then with pitch contour varied as shown.
Each of the 29 contours was identified as being just like Contour A (= Contour 3) or just like Contour C (= Contour 8), or most like one of these two.
CPS
200
150
100
CONTOUR A
CONTOUR C
% IDENTIFICATIONS
100
75
50
25
0
3
11
12
13
14
15
8
C
C-LIKE
A-LIKE
A

Fig. 6 b

TEST NO. 6 CONTOUR A - CONTOUR C (CONT'D.)

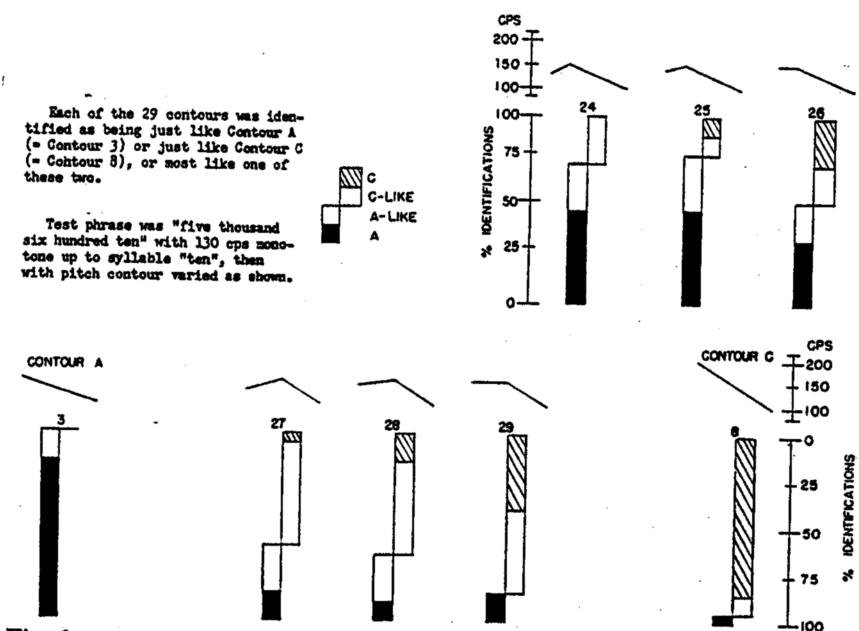


Fig. 6 c

TEST NO. 6 CONTOUR A - CONTOUR C (CONT'D.)

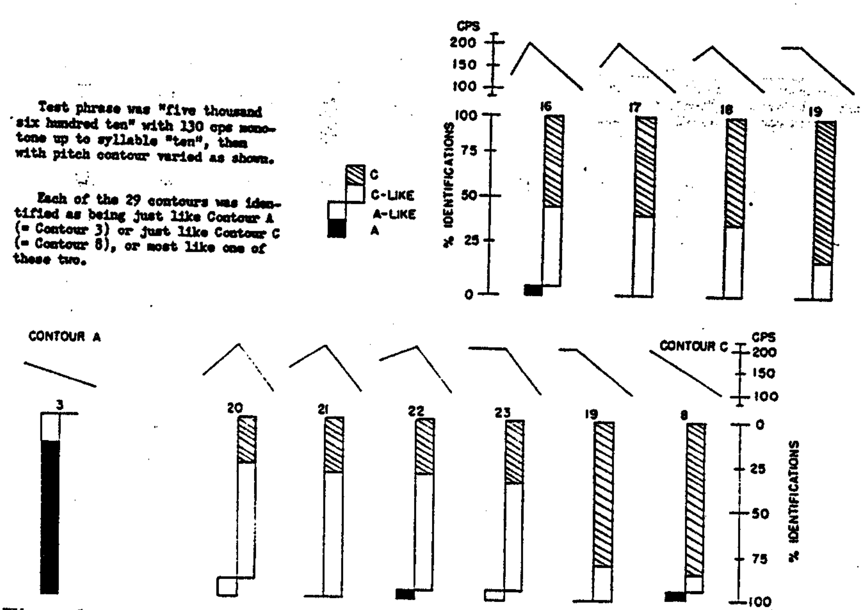


Fig. 6 d

TEST NO. 7 CONTOUR A — CONTOUR E (CONT'D)

Test phrase was "five thousand six hundred ten" with 130 cps monotone up to syllable "ten", then with pitch contour varied as shown.

Each of the 19 contours was identified as being just like Contour A (= Contour 3) or just like Contour E (= Contour 13), or most like one of these two.

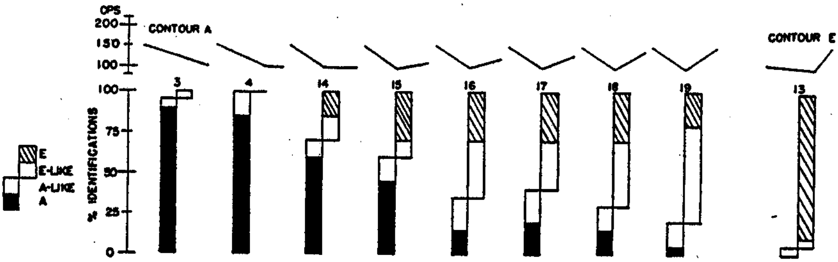
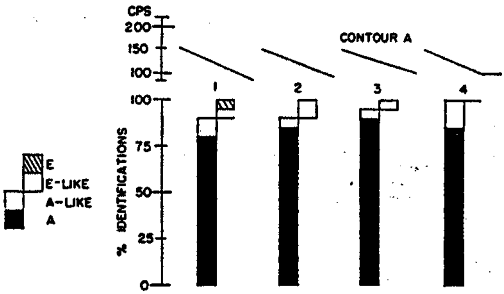


Fig. 7 a

TEST NO. 7 CONTOUR A — CONTOUR E

Test phrase was "five thousand six hundred ten" with 130 cps monotone up to syllable "ten", then with pitch contour varied as shown.

Each of the 19 contours was identified as being just like Contour A (= Contour 3) or just like Contour E (= Contour 13), or most like one of these two.



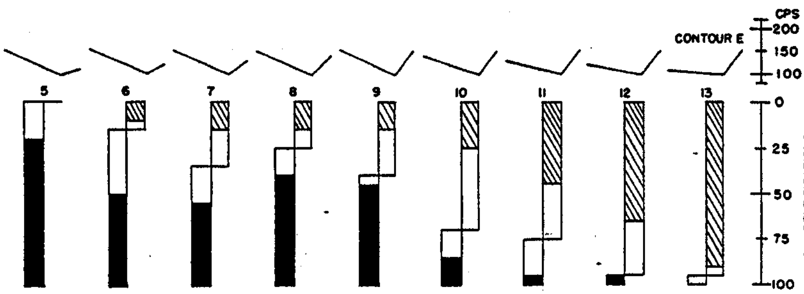


Fig. 7 b

Acknowledgement

Grateful acknowledgement is made to the Haskins Laboratories for research facilities and to the Carnegie Corporation of New York for funds made available for the support of this work.

Summary

The fundamental frequency movements of five American English intonation contours were varied in a systematic fashion and the variations were submitted to identification tests. The results of these tests have shown that a contour can stand considerable change before it loses its identity. The important features of the synthetic contours used in this investigation can be described in terms of rectilinear movements between three pitch levels.

References

ABRAMSON, A. S., (1962) *The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments*. Bloomington: Indiana U. Res. Center in Anthro., Folklore and Ling., 1962.

BORST, J. M. and COOPER, F. S., (1957) Speech Research Devices Based on a Channel Vocoder. JASA, 29, 777 (Abstract).

COOPER, F. S., (1964) Instrumental Methods for Research in Phonetics. Paper given at the Fifth International Congress of Phonetic Sciences, Münster.

DELATTRE, P., (1963) Comparing the prosodic features of English, German, Spanish and French. IRAL, 1, 193.

HADDING-KOCH, K. and STUDDERT-KENNEDY, M., (1964) Intonation contours, evaluated by American and Swedish test subjects. Paper given at the Fifth International Congress of Phonetic Sciences, Münster.

ISAČENKO, A. V. und SCHÄDLICH, H.-J., (1963) Erzeugung künstlicher deutscher Satzintonationen mit zwei kontrastierenden Tonstufen. Monatsberichte der Deutschen Akademie der Wissenschaften zu Berlin, 5, Heft 6, 365.

— (1964) Untersuchungen über die deutsche Satzintonation. (Berlin. Vorabdruck)

MAGDICS, K., (1964) Acoustic correlates of some Hungarian emotive intonation patterns. Paper given at the Fifth International Congress of Phonetic Sciences, Münster.

ULDALL, E., (1962) Ambiguity: Question or Statement? or "Are you asking me or telling me?" Proc. of the Fourth International Congress of Phonetic Sciences. (The Hague)

Haskins Laboratories, New York, and Institute of Phonetics, Lund