The perception of fundamental frequency declination

Janet Pierrehumbert

Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 1 November 1978; accepted for publication 7 April 1979)

A series of experiments was carried out to investigate how fundamental frequency declination is perceived by speakers of English. Using linear predictor coded speech, nonsense sentences were constructed in which fundamental frequency on the last stressed syllable had been systematically varied. Listeners were asked to judge which stressed syllable was higher in pitch. Their judgments were found to reflect normalization for expected declination; in general, when two stressed syllables sounded equal in pitch, the second was actually lower. The pattern of normalization reflected certain major features of production patterns: A greater correction for declination was made for wide pitch range stimuli than for narrow pitch range stimuli. The slope of expected declination was less for longer stimuli than for shorter ones. Lastly, amplitude was found to have a significant effect on judgments, suggesting that the amplitude downdrift which normally accompanies fundamental frequency declination may have an important role in the perception of phrasing.

PACS numbers: 43.70.Dn, 43.66.Hg

INTRODUCTION

The declination effect, or the tendency of pitch to drift downwards over the course of an intonation group, has been observed in many languages. For example, it is known to occur in French (Vaissiere, 1971), Finnish (Hirvonen, 1970), Japanese (Fujimura, personal communication), and in a large number of African languages (Silverstein, 1976; Welmers, 1973); it has been particularly well documented in work by Collier and t'Hart (1971), and t'Hart and Cohen (1973) on Dutch, and in Maeda's work on English (Maeda, 1976). The attention the phenomenon has received in these last works has focused on two questions: what physiological mechanism is responsible for the gradual descent in F_0, and how an intonation synthesis program may incorporate declination so as to produce a natural sounding result. With the notable exception of recent work by Cooper and Sorenson (1977), the possible role of declination in the linguistic system has been largely neglected.

Data from a number of investigations (Maeda, 1976; Sternberg *et al.*, 1979; Sorensen and Cooper, 1979 have shown that the slope of declination is less in a longer intonation group than in a shorter one. Experiments by Cooper and Sorensen (1979), have shown that the intonation group, as defined by a ramp of F_0 downdrift, is not strictly coextensive with the breath group; a new ramp is frequently started at a major syntactic boundary even without inhalation. Somewhat less frequently, inhalation takes place without declination ramp resetting. Results of this sort suggest that purely physiological explanations of declination are incomplete. While factors such as declining subglottal pressure (Collier, 1975), tracheal pull (Maeda, 1976), or aspects of laryngeal F_0 control provide a physiological underpinning for declination, it is clear that the speaker is able to control the process in such a way that each ramp coincides with a suitable syntactic unit. Ohala (1978) conjectures that the usefulness of resetting as a cue to syntactic organization has led to its being regularized in the linguistic system. This position is defended at length in Breckenridge (1977).

In view of this conclusion, it seems worthwhile to examine declination not only from the point of view of physiology, but also from the point of view of mental representation. For this reason, a series of experiments on how declination is perceived by speakers of English was carried out. The present paper describes the results of these experiments. In all experiments, subjects were asked to judge which of two given stressed syllables in a nonsense sentence was higher in pitch. It was found that in making this judgment, they corrected for the expected declination; the peak configuration for which they answered at random actually had a second peak which was lower in pitch than the first. Subsequent experiments examined the effects of pitch range, amplitude contour, and utterance length on the amount that subjects corrected for declination. These variables were selected for study because their relationship to the declination effect in production is of interest. Pitch range was chosen because informal examination of F_0 contours had shown that speakers tend to produce a steeper declination when using vivid, wide-pitch range intonation than when using more monotonous, narrow-pitch range intonation; in short an increase in pitch range affects the beginning of an intonation group more than the end. By examining the extent to which subjects' normalization for declination in perception can be affected by pitch range, we gain a clue to the extent to which the listener's mental representation of declination reflects the complexities of production patterns. Testing how the slope of expected declination is affected by utterance length provides a second clue, because slope is known to be affected by utterance length in production. Amplitude contour was used as an experimental variable because the available data show that the F_0 downdrift on an intonation group is accompanied by an amplitude downdrift, which typically totals 3–4 dB. A set of two experiments was carried out to test the hypothesis that the listener relies on the amplitude downdrift as well as the pitch downdrift in forming an impression of phrasing.

I. EXPERIMENTAL METHODS

The stimuli for the experiments were constructed using a linear prediction analysis–resynthesis scheme. The version of linear prediction coding used here (Atal

0001-4966/79/080363-07$00.80

and Hanauer, 1971) represents the speech waveform in terms of two source functions, F_0 and amplitude, and 12 pseudoarea functions which characterize the filtering properties of the vocal tract. Using speech coded in this way, it is possible to manipulate the F_0 or amplitude of an utterance and then resynthesize it with essentially the original segmental phonetics. The sentence analyzed was a nonsense sentence, in which the speaker had replaced the syllable "ma" for each syllable of "The baker made bagels," while preserving the prosodic pattern (Liberman and Streeter, 1978). The F_0 contour on the last stressed syllable in this sentence was scaled up and down by small increments from its original position, so as to produce a set of contours like that schematized in Fig. 1. Each pitch contour was recombined with the area functions from the original utterance. Multiple tokens of each stimulus were then synthesized (the number of copies varied between nine and 15 in different experiments). All were randomized in blocks so as to insure that no more than two tokens of the same stimulus appeared in a row. Paid high school subjects heard each token once before recording on their answer sheets whether the first or second stressed syllable seemed higher in pitch.[1] Practice items were adminstered and scored before the start of each experiment to insure that the subjects understood the instructions.

Statistical methods described in Sec. II were used to estimate what value of the second peak would have elicited random responses. With certain reservations discussed below, this crossover point may be viewed as reflecting the second peak value which would make the second peak seem equal to the first. As mentioned above, pilot experiments established that subjects made a systematic, though not complete, correction for declination in judging which peak was higher; the second peak which was estimated to be subjectively equal to the first was in fact about 10 Hz lower.[2] The experiments reported here each involved several stimulus sets, which were constructed as in Fig. 1 and which differed from each other in values of the variable under investigation: pitch range, amplitude, or utterance length. Stimuli from all sets in a given experiment were randomized together, so that practice would affect the results for all sets equally. Comparing the crossover points for the stimulus sets in an experiment gave a measure of the importance of each experimental variable to the subjects' expectations about declination.

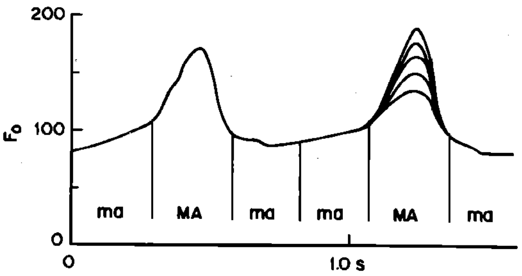

FIG. 1. The set of F_0 contours from one stimulus set superimposed so as to show how the F_0 contour on the second peak was varied. Alternate stimuli are omitted for the sake of legibility. Vertical lines represent syllable boundaries as determined from LPC parameters; "MA" is used here and below to indicate a syllable with pitch and duration cues for stress, while "ma" represents unstressed syllables.

The pitch range experiment used stimuli which preserved the amplitude contour of the original utterance. Since the original amplitude contour downdrifted, subsequent experiments used stimuli in which the possible effects of amplitude were better controlled; in the utterance length experiments, the amplitude was ramped up over the first and down over the last syllable and kept constant in between, while in the amplitude experiment, the artificial amplitude contours shown in Fig. 2 were used. The stimuli used in the amplitude and utterance length experiments also differed from those used in the pitch range experiment in two aspects of the pitch contour. First, the shape of the pitch contour on the second stressed syllable was copied from that on the first, and then scaled up and down by small increments as described above. This was done to control for any possible effects of peak flatness or location. Secondly, the F_0 contour on the unstressed syllables was adjusted so that the two syllables being compared would be flanked by syllables with the same F_0 contour. Without this control, any linguistic interpretation of the results would be uncertain because of known interference effects in the short term memory of tonal pitch (Deutsch, 1975). Even with these changes, the topline, baseline, and local F_0 changes in the stimuli fell within the range of variation represented in a corpus of utterances by the original speaker. In particular, it is very typical for the topline to have a steeper slope than the baseline (Breckenridge, 1977; Sorensen and Cooper, 1979). An attempt to replicate the original pitch range experiment using the better controlled stimuli was successful.[3]

II. STATISTICAL METHODS

The statistical analysis of the data had two aims: to estimate the crossover point for each set of stimuli, and to determine whether the crossover points for different stimulus sets in an experiment were significantly different.

To estimate the crossover point, the proportion of 2nd-peak-higher responses elicited by each stimulus in a set was plotted against the difference between the peak F_0 on the second stressed syllable and the peak F_0 on the first stressed syllable. An inverse cumulative normal transform was taken of the proportion 2nd-peak-higher responses, and a line was fit by least squares to the transformed data. The crossover point was taken to be the point at which the cumulative normal curve thus fitted had an ordinate of 0.5. (This means that in the figures and in the text, crossover points are not reported as values of the second peak, but rather as values of the second peak minus the first peak.) An example outcome of this procedure may be seen in Fig. 3, which shows the data points from the pitch range experiment and the curves fit to them. In this figure, the rightmost three data points from the wide pitch range stimulus set (solid dots) were left off in fitting the

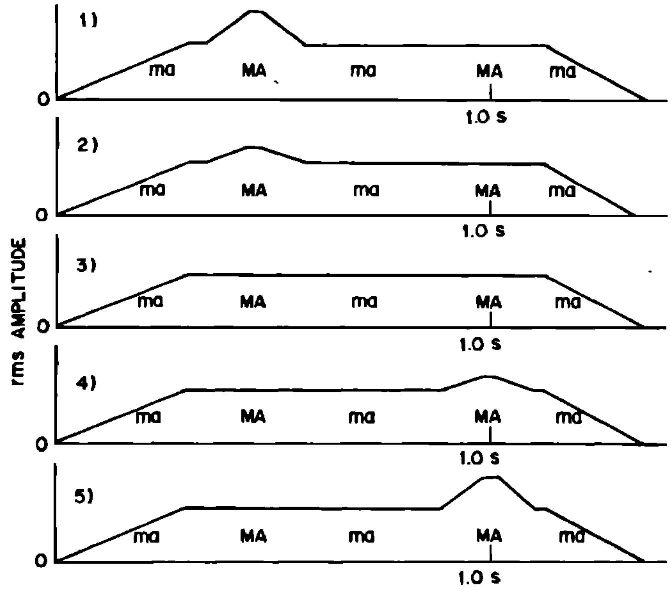

FIG. 2. The five amplitude contours used in the second experiment, which examined the effect of the relative amplitude of peaks on judgments of relative pitch. (Stimuli in this experiment were five rather than six syllables long.)

curve. This was done because these points provide little information about the location of the crossover point, and because the fitting procedure is extremely sensitive to values of the ordinate near 1 or 0. Had the points been left in, the estimate of the crossover point would have been heavily influenced by data points which do not in fact contain much information about it. In general, this problem was addressed by adopting the following rule for truncating data sets: If 0.95 of the responses to some stimulus were 2nd-peak-higher responses, all stimuli with an even higher second peak were left out in making the fit. Similarity, if no more than 0.05 of the responses to a stimulus were 2nd-peak-higher responses, all stimuli with an even lower second peak were left out. Approximately 10% of the data points from the experiments reported here were left out on these grounds.

Tukey's jackknife procedure (Mosteller and Tukey, 1977) was used in making the fit in order to assess whether crossover points were significantly different from one another. The jackknife is a useful distribution free procedure for setting approximate confidence limits on the results of a complex calculation. Ideally, if we wished to estimate a parameter F (in this case, the crossover point) by some calculation on the data from an experiment, we would do the experiment a large number of times with different subjects, and compute F_j for each repetition j of the experiment. We would take as our estimate of the parameter the mean of these results; since the mean of a sufficiently large number of independent measurements from the same distribution is normally distributed, it would also be possible to estimate a confidence interval for our final estimate $\overline{F_j}$ by computing the standard deviation of the F_j. The jackknife in some regards simulates the ideal, using the data from just one repetition of the experiment. Where k is the number of pieces of data (in this case, the responses of one subject to a set of stimuli), take

$$F_*(j) = kF(\text{all}) - (k-1)F(j) \ ,$$

where F(all) represents the result of doing the calculation using all the data, and $F(j)$ is the result of doing the calculation including all but the jth piece of data. In this case, F (all) is the crossover point of a curve fit to the pooled responses of all k subjects, while $F(j)$ is the crossover point of a curve fit to the pooled responses of $k-1$ out of k subjects. $F_*(j)$ provides a measure of how much difference the responses of the jth subject make by computing how much the crossover point changes when they are left out. Tukey has found that the mean of the k $F_*(j)$, $\overline{F_*(j)}$, is a better estimator of the true result of the calculation than is F(all), though the two values are usually very close. $\overline{F_*(j)}$ is the estimate of the crossover point used here. Its standard deviation can be estimated by $(s^2/k)^{1/2}$, where s^2 is the sample variance of the $F_*(j)$. A confidence interval for $\overline{F_*(j)}$ can then be estimated using Student's t distribution for $k-1$ degrees of freedom. The method does not assume that the data are drawn from a normal distribution; the t test is applicable because the mean of k independent measurements from any distribution approaches a normal distribution for large k, just as the t distribution does.

III. PITCH RANGE EXPERIMENT

The stimuli for the first experiment consisted of two sets. In the first set, the first peak was at 151 Hz, or

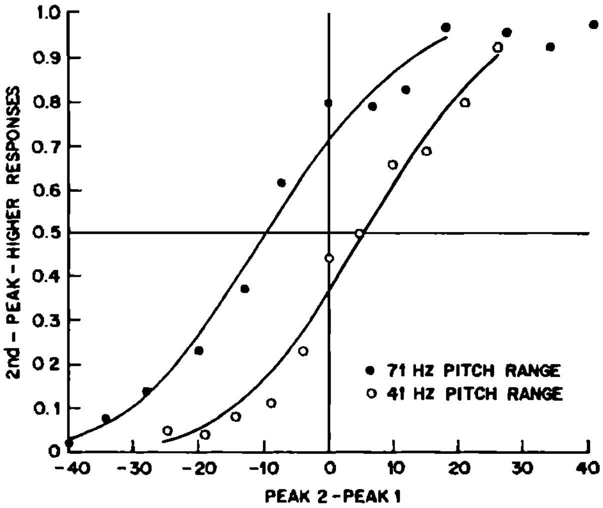

FIG. 3. The results of the pitch range experiment. Solid dots represent pooled 2nd-peak-higher responses to the wide pitch range stimuli. Open dots represent responses to the narrow pitch range stimuli. Solid curves are cumulative normal curves fit by jackknifing. The ordinate at the intersection of a curve with the vertical axis is the proportion of 2nd-peak-higher responses elicited when the two peaks were actually equal. The abscissa at the intersection with the horizontal axis indicates the relation between the two peaks which would cause subjects to answer at random. This crossover point may be viewed as the point at which the two peaks would seem equal.

TABLE I. Peak values on the second stressed syllable in the pitch range experiment.

	Wide pitch range stimuli	Narrow pitch range stimuli
	111 Hz	
	117	96 Hz
	123	102
	131	107
	138	112
	144	117
2 peaks equal	151	121
	157	126
	163	131
	169	136
	178	142
	185	147
	192	

71 Hz higher than the baseline point at which the intonation contour started. In the second set, the first peak was at 121 Hz, so the initial rise spanned 41 Hz. In each set, the second peak was varied by small increments as described above. The peak values actually used are displayed in Table I.[4]

There were an equal number of stimuli in each set with a second peak higher than the first as lower than the first. Ten tokens of each stimulus were made and all were randomized together. There were nine subjects, so that altogether 90 responses to each stimulus were obtained.

The results for the pitch range experiment are summarized in Fig. 3. The crossover point for the wide pitch range stimuli was −9.2 Hz, while the crossover point for the narrow pitch range stimuli was 5.6 Hz. This difference was found to be significant at the 99% level. There was likewise 99% certainty that the crossover point for the wide pitch range stimuli was less than 0; this means that we can feel confident that the subjects made a correction for declination in judging the wide pitch range stimuli, rather than reacting on the basis of their acoustical features alone.

It seems perplexing at first that the crossover point for the narrow pitch range stimuli was actually higher than 0. Does this mean that the subjects expected a pitch updrift in the narrow pitch range stimuli, and corrected accordingly? This explanation seems unlikely, in view of the fact that a pitch updrift is found only in marked intonation patterns in English, such as yes/no questions. It seems more probable that a response bias shifted both curves to the right of their true locations. Since there were the same number of stimuli with the second peak physically higher than the first as lower, the subjects by correcting for declination would perceive a greater number of stimuli with the second peak higher. On the assumption that the "true" correction for declination would be 16 Hz for the wide pitch range stimuli (16 Hz was the actual F_0 drop in the recording from which the stimuli were constructed) and 0 Hz in the narrow pitch range case (0 Hz is the most conservative hypothesis which does not posit an updrift), we would expect that about 65% of the responses in the experiment would be 2nd-peak-higher responses. In fact, the pitch range experiment elicited 51% 2nd-peak-higher responses. All subsequent experiments, which used a set of second peak values designed to be balanced around a perceptual rather than a physical point of equality with the first peak, elicited between 48% and 53% 2nd-peak-higher responses. The fact that the percentage was close to 50% regardless of the range of variation in the second peak suggests strongly that subjects balanced their answer sheets. The conclusion to be drawn is that the location of a crossover point for a stimulus set is a less interpretable experimental result than are differences between crossover points for stimulus sets in the same experiment.

IV. AMPLITUDE EXPERIMENTS

In the stimuli for these two experiments, the relative amplitude and relative pitch of the two peaks were varied orthogonally. The pitch contours used differed from those used in the pitch range experiment by controlling for peak location and flatness and for F_0 on unstressed syllables as described above. The first peak was at 169 Hz, with the nine second peak values evenly spread in the range from 135 to 192 Hz. As mentioned above, this range of variation was chosen so that roughly the same number of stimuli would be perceived as having the first peak higher as having the second peak higher. The amplitude contours used are shown in Fig. 2. In the first, the amplitude of the first peak is 4 dB greater than that of the second; in the second, the first is 2 dB greater; in the third, the two stressed syllables have the same amplitude. In the fourth set, the second peak is 2 dB greater than the first, and in the fifth, it is 4 dB greater. All combinations of F_0 and amplitude contour were synthesized.

In the first experiment that was done using these stimuli, the subjects were asked to judge which stressed syllable was higher in pitch. There were ten subjects, who heard nine tokens of each stimulus. Figure 4 shows one standard error on each side of the estimated crossover point in pitch judgments for each of the five amplitude contours. The amplitude contour had a striking effect on the location of the crossover point; increasing the amplitude on a peak increased the number of impressions that its pitch was higher. In order to quantify this effect and determine whether it is statistically significant, a jackknife was used to fit a regression line through the estimated crossover points. The slope of the fitted line was 1.5 Hz/dB, which was found to be different from 0 at greater than 99% significance level. This means that an amplitude downdrift of 4 dB, which can commonly be found in the course of an intonation group, would shift the crossover point by 6 Hz. While an effect of this magnitude obviously does not rival the importance of F_0, it suggests that amplitude may have an important role in the perception of phrasing.

In at least one sense, however, it is not clear how the results of this experiment are to be interpreted. Did the amplitude contour affect the subjects' perception of pitch, or was it the case that they gave judg-

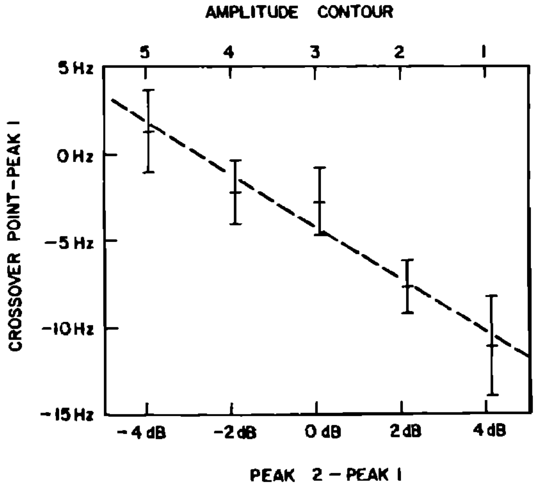

FIG. 4. The results of the amplitude experiment. Vertical bars indicate the crossover point in pitch judgments as estimated for each amplitude contour, with one standard error of measurement on either side. The dashed line is the result of a jackknifed regression of crossover point on relative peak amplitude.

ments of relative prominence, weighing in their perceptions of both pitch and loudness, when they had been asked to judge pitch alone? Research on the perception of pure steady tones would suggest that the first hypothesis is wrong. An amplitude difference of 4 dB would have a negligible effect on the perception of the pitch of any pure tone up to 5000 Hz, which is the cutoff frequency for the digitized speech used here; the slight effect which exists would furthermore be in opposite directions for frequency components above and below 2500 Hz (Gulick, 1971). However, the perception of complex moving tones is badly enough understood that the applicability of these results to the present experiments is somewhat uncertain. So, a second experiment was conducted to determine whether subjects could separate the relative pitch of the peaks from their relative loudness when asked to do so.

Subjects were asked to judge first the relative loudness of the two peaks, and then the relative pitch, for each of the tokens presented. The token was repeated twice in order to facilitate this judgment. Testing was carried out on two successive days in order to elicit ten paired responses per stimulus without fatiguing the subjects. There were eight subjects.

Multiple regression was used to provide an indication of the subjects' success in separating pitch and loudness. The functions fitted were:

$$R_L = 0.085A + 0.005F + 0.50 \quad (r^2 = 0.918)\ ,$$

$$R_P = -0.012A + 0.018F + 0.623 \quad (r^2 = 0.881)\ ,$$

where R_L is the proportion of 2nd-peak-louder responses, R_P is the proportion of 2nd-peak-higher responses, A is the amplitude difference between the two peaks (peak 2−peak 1) and F is the difference in peak frequency. For comparison, the same analysis of the data from the first experiment gives:

$$R_P = 0.025\,A + 0.019\,F + 0.584 \quad (r^2 = 0.907)\ .$$

TABLE II. Stimulus sets in utterance length experiment A.

Set	Sentence	Time between peaks
1	ma MA ma MA ma	0.56 s
2	ma MA ma ma MA ma	0.80 s
3	ma MA ma ma ma MA ma	1.24 s

Thus, while in the first experiment, increasing the relative amplitude of the second peak increased the probability of a 2nd-peak-higher response, in the second experiment, subjects hypercorrected; increasing the amplitude of the second peak decreased the probability of a 2nd-peak-higher response. The magnitude of the effect of amplitude in this experiment was, however, about half the magnitude of the effect found in the first experiment. Thus subjects with only minimal practice had some success in separating pitch and loudness; we infer that the results of the first experiment were not due to a psychophysical effect of amplitude on pitch perception, but rather to the subjects' providing judgments of relative prominence rather than relative pitch.

V. UTTERANCE LENGTH EXPERIMENTS

Two experiments examined the affect of utterance length on the expected declination. In the first experiment, the total length of the utterance was varied by manipulating the number of unstressed syllables in the middle portion of the utterance. Stimuli were constructed with one, two, or three unstressed syllables between the two stressed syllables. To make the third set, which had one more syllable than the original utterance, the LPC parameters for one of the original medial unstressed syllables were reduplicated. The F_0 contour between the two stressed syllables had the same shape in all three sets of stimuli, having been stretched or shrunk to the correct duration to span one, two, or three unstressed syllables. Table II summarizes the differences among the three sets of stimuli. The F_0 contour on the first two syllables in all sets was the same as in the amplitude experiment; the second peak values were likewise varied in the same way. The experiment involved 13 subjects who responded to 15 tokens of each stimulus.

The crossover points estimated for sets 1, 2, and 3 were −6.9, −9.2, and −8.4 Hz, respectively. The probability that the difference between two sets was due to chance was for all three pairs greater than 30%; in particular, there was greater than 35% probability that the difference between sets 1 and 3 was due to chance. In view of the fact that manipulating pitch range and amplitude had produced highly significant differences in the correction subjects made for declination, it was concluded that the total declination expected in a short to medium length utterance is unaffected by the length of the utterance, as measured in syllables or in actual time.

This result does not, however, rule out the possibility that the expected declination varies with the number of stressed syllables. Thus a second experiment was carried out which manipulated this variable. In this experiment, all the stimuli were seven syllables long. In set 1, three unstressed syllables intervened between the stressed second and sixth syllables. In the remaining stimuli, the fourth, or middle, syllable was stressed. This meant that there were three stressed syllables in the utterance; subjects were asked to compare pitch on the first and last. There was some concern that comparing the first and third of three stressed syllables would be a difficult task for the subjects, and that they would adopt a strategy of comparing the third to the second, or to a value expected from extrapolating a line connecting the first and the second. In this case the choice of second peak value would determine the results of the experiment. As a check on whether such a strategy was adopted, two different values for the second peak were used. Set 2 had a second peak of 149 Hz, and set 3 had a second peak of 163 Hz. In each set, the first peak was at 169 Hz and the last peak was varied as in the first experiment. The experiment involved 18 subjects who responded to 12 tokens of each stimulus.

The crossover points for sets 1 and 2 were −11.1 and −10.9, respectively. It was estimated that there was greater than 45% probability that the difference between these two values was due to chance. The crossover point for set 3, with the higher medial peak, was −6.8 Hz; this result was significantly different from the results for the other two sets. From these values we can conclude that increasing the utterance length as measured in stressed syllable count does not increase the expected declination; adding the stressed syllable in set 2 had no significant effect, while the medial peak in set 3 produced an effect in the wrong direction to support such a hypothesis. However, the contrast between the result for set 3 and the results for sets 1 and 2 would seem to call for further investigation.

How are these results related to what is known about production? The mental representation of declination suggested by these results would be an accurate reflection of production patterns if slope were inversely proportional to length in production. Maeda claims that this is the case, and his claim has been partially corroborated by Sternberg *et al.*, who found the slope was inversely proportional to length in the utterance of lists of two to five words. However, Sorensen and Cooper's data on sentences of eight and sixteen syllables show two effects. They found that the total topline declination was greater in the long sentences than in the short ones; the subjects accomplished this mainly by starting higher, and to some extent by ending lower. In addition, their data show that the slope of declination is less in longer sentences than in shorter ones; by cumulating the data in their Table V, we see that the mean slope of declination in their short sentences is 6.7 Hz/syllable, and 4.0 Hz/syllable in the long sentences. Taken together, the results of Sorensen and Cooper and of Sternberg *et al.* suggest a hypothesis: the speaker preferentially adjusts declination for the length of the intonation group by adjusting the slope, but when the length of the utterance makes it necessary, he begins higher in order to maintain a perceptually salient slope. Two important ingredients in this hypothesis are the assumption that declination plays an important role in signaling phrasing, so that information is lost if declination is not clearly marked, and the assumption that the bottom of the speaker's range is fairly fixed, so that required adjustments must be made at the top. Under this hypothesis, the present results of the present perception experiments accurately mirror production patterns; it might be expected, however, that a similar experiment using longer stimuli would pick up a perceptual counterpart to Sorensen and Cooper's first result. Clearly additional work on both production and perception is indicated to determine whether this picture is an accurate one.

VI. CONCLUSIONS

The first conclusion to be drawn from the experiments reported here is that speakers normalize for declination in judging the relative height of peaks in the intonation contour. Normalization has been reported in other aspects of speech: for example, Ladefoged and Broadbent (1957) and Verbrugge and Strange (1976) have studied normalization for the formant space of the speaker, Summerfield and Haggard (1972) report normalization for tempo, and Klatt and Cooper (1975) have found normalization for durational effects. Like earlier findings of normalization, the present results show that the relationship between the physical properties of the speech waveform and linguistic units is not invariant. Normalization for the speaker's formant space means that the mental representation of the vowels cannot be equated with actual formant values; normalization for declination implies that the relative salience of stressed syllables in neutral intonation is not directly mirrored in F_0 values.[5] An important feature of the present finding, however, is that the feature of the speech waveform which is normalized for is itself a carrier of linguistic information: while local excursions in pitch carry information about the relative salience of different syllables, declination carries information about the syntax of the sentence. This means that the effects of declination must be set aside at one level of linguistic processing while they contribute at another.

The mental representation of declination which is involved in normalization was found to reflect some major features of the pattern of declination found in production. It was found that speakers expect more declination in wide pitch range utterances than in narrow pitch range utterances, and that the expected slope of declination is less for a longer utterance than for a shorter one. The amplitude downdrift which typically accompanies F_0 downdrift was also found to have a part in the mental representation of declination. The effect of amplitude was perhaps surprisingly large, given that Streeter (1978) found only a small effect of amplitude on the perception of phrasing. One notes, however, that her work was concerned with the implementation of phrasal boundaries within an intonation group; the contrast of her results to those reported here may mean that amplitude becomes important only at a broader level of phrasing than she investigated.

ACKNOWLEDGMENTS

The work reported here was carried out at Bell Telephone Laboratories, Murray Hill, N. J. The help and support of Osamu Fujimura and Mark Liberman are gratefully acknowledged.

---

[1] The experimental procedure obviously measures expectations about the declination of the topline (a line connecting the peaks in the F_0 contour) rather than the baseline. This is appropriate, because topline declination is both more perceptually salient than baseline declination and more reliably found in production.

[2] By way of comparison, the original recording from which the stimuli were generated had a topline declination totalling 16 Hz.

[3] A difference of 12 Hz was found between the crossover point for the wide pitch range stimuli (which had the first peak at 169 Hz) and the narrower pitch range stimuli (which had the first peak at 130 Hz); the response bias effect in the original pitch range experiment, which is discussed below, was also found.

[4] Because the computer program which was used to generate the stimuli operates on the basis of rounded pitch period values rather than fundamental frequency values, it was impossible to generate a stimulus set in which all increments were exactly the same.

[5] In marked intonation contours, the relationship between F_0 and salience is further complicated by the fact that stressed syllables may have a low tone. See Libermen, M. (1975). "The Intonational System of English," doctoral dissertation, MIT for discussion of such contours and for further references.

Atal, B., and Hanauer, S. (1971). "Speech analysis and synthesis by linear prediction of the speech wave," J. Acoust. Soc. Am. 50, 637–655.

Breckenridge, J. (1977). "Declination as a Phonological Process," Bell Lab. Techn. Memo., Murray Hill, NJ.

Collier, R. (1975). "Physiological Correlates of Intonation Patterns," J. Acoust. Soc. Am. 58, 249–255.

Collier, R., and 't Hart, J. (1971). "A Grammar of Pitch Movements in Dutch Intonation," IPO Annual Progress Report 6, 17–21.

Cooper, W., and Sorensen, J. (1977). "Fundamental Frequency Contours at Syntactic Boundaries," J. Acoust. Soc. Am. 62, 683–692.

Cooper, W., and Sorensen, J. (1979). *Fundamental Frequency in Sentence Production* (in press).

Deutsch, D. (1975). "The Organization of Short Term Memory for a Single Acoustic Attribute," in *Short Term Memory*, edited by D. Deutsch and J. A. Deutsch (Academic, New York), pp. 108–148.

Fujimura, O. (personal communication).

Gulick, W. L. (1971). *Hearing: Physiology and Psychophysics* (Oxford U. P., New York).

Hirvonen, P. (1970). *Finnish and English Communicative Intonation* (University of Turku, Turku).

Klatt, D., and Cooper, W. (1975). "Perception of vowel duration in sentence contexts," J. Acoust. Soc. Am. 57, 47.

Ladefoged, P., and Broadbent, D. (1957). "Information Conveyed by Vowels," J. Acoust. Soc. Am. 29, 98–104.

Liberman, M., and Streeter, L. (1978). "Use of nonsense-syllable mimicry in the study of prosodic phenomena," J. Acoust. Soc. Am. 63, 231–233.

Maeda, S. (1976). "A Characterization of American English Intonation," doctoral dissertation, MIT, Cambridge.

Mosteller, F., and Tukey, J. (1977). *Data Analysis and Regression* (Addison–Wesley, Reading, MA).

Ohala, J. (1978). "Production of Tone," in *Tone: A Linguistic Survey*, edited by V. Fromkin (Academic, New York).

Silverstein, R. (1976). "A Strategy for Utterance Production in Hausa," Stud. Afr. Linguist. S6.

Sorensen, J., and Cooper, W. (1979). "Syntactic Coding of Fundamental Frequency in Speech Production," in *Perception and Production of Fluent Speech*, edited by R. A. Cole (Erlbaum, Hillsdale, NJ) (in press).

Sternberg, S., Wright, C., Knoll, R., and Monsell, S. (1979). "Motor programs in rapid speech: additional evidence," in *Perception and production of fluent speech*, edited by R. A. Cole (Erlbaum, Hillsdale, NJ) (in press).

Streeter, L. (1978). "Acoustic determinants of phrase boundary perception," J. Acoust. Soc. Am. 64, 1582–1592.

Summerfield, A. Q., and Haggard, M. (1972). "Articulatory Rate versus Acoustical Invariants in Speech," J. Acoust. Soc. Am. 52, 113.

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

Vaissère, J. (1971). "Contribution â la synthêse par regles du francais," doctoral dissertation, Université des Langues et Lettres de Grenoble.

Verbrugge, K., and Strange, W. (1976). "What information enables a listener to map a talker's vowel space," J. Acoust. Soc. Am. 60, 198–212.

Welmers, W. E. (1973). *African Language Structures* (University of California, Berkeley).