Measures of the sentence intonation of read and spontaneous speech in American English

Philip Lieberman, William Katz, Allard Jongman, Roger Zimmerman, and Mark Miller
Brown University, Providence, Rhode Island 02912

The visual abstraction procedure used in previous studies of declination was tested using 12 subjects who each fit the F_0 contours of 19 spoken short simple sentences with baselines. These baselines were found to be poorly replicated by the fitters. An objective all-points least-squares best-fit procedure was tested on this corpus and on a set of sentences that had been produced in both spontaneous and read speech by six speakers. The all-points linear regression line was a better descriptor of the F_0 contours than either baselines or toplines. Declination did not always occur in these simple declarative sentences; there was more variation present in the F_0 contours of sentences that had been uttered during spontaneous speech; 35% of the spontaneous sentences did not show declination; 45% of these sentences better fit the breath-group model. Their F_0 contours could be described by a level all-points linear regression line followed by a falling terminal segment.

INTRODUCTION

In recent years, a number of studies like those of Maeda (1976), Pierrehumbert (1979), and Cooper and Sorenson (1981) have proposed that the fundamental frequency contours of isolated sentences and sentences in connected speech can be described in terms of global declination lines. The total F_0 contour is characterized in terms of a line that extends across the total sentence or a major constituent. Global "baselines" or "toplines" are constructed to approximate the trend in time of either the valleys or the peaks of F_0 in a linear relationship that has a negative slope, i.e., $\Delta F_0/\Delta t < 0$, where F_0 is the change in F_0 over the duration Δt of a sentence. Furthermore, the falling declination lines are proposed as the salient acoustic cues that listeners use to segment the flow of speech into sentences or to mark out the utterances' constituent (grammatical) structure. Declination theories thus claim that baselines or toplines that gradually fall throughout the course of a sentence describe linguistically salient aspects of the F_0 contour. These claims differ from those of earlier studies like those of Armstrong and Ward (1926), Trager and Smith (1951), and Lieberman (1967), which claim that the F_0 contour of a sentence consists of two parts: a nonterminal part, and a terminal part, that have different slopes. Lieberman (1967) accounted for the two-part contours via a breath-group theory of intonation. Testing these competing theories involves, among other things, determining which metric (baselines, toplines, or lines which characterize the nonterminal portion of a breath group) best describes the fundamental frequency contours that actually occur in the production of speech.

The technique that has been used to determine the descriptive adequacy of many of these competing theories is essentially subjective, involving the visual inspection of an F_0 contour. For this reason, Maeda (1976) introduced a "visual abstraction" procedure in which baselines drawn through the valleys of a plot of F_0 were estimated by eye. Such a

procedure is present or implicit in subsequent theories involving declination (e.g., Vaissiere, 1974; Delgutte, 1978; Garding, 1979).

The present study has several related objectives. First, we will determine whether Maeda's visual abstraction technique yields consistent results when used by different observers. Second, we will develop and test an alternative, quantitative method. This quantitative method will then be employed to compare the intonation contours of samples of spontaneous and read speech to test both Maeda's (1976) declination theory and Lieberman's (1967) breath-group theory.

I. EXPERIMENT 1: TESTING THE "VISUAL ABSTRACTION PROCEDURE"

The visual abstraction procedure introduced by Maeda (1976) explicitly or implicitly structures much of the discussion concerning the descriptive adequacy of both baseline and topline declination theories. Maeda fitted straight lines to the F_0 valleys by visual abstraction. A necessary preliminary to determining whether such baselines are "better" descriptors of the F_0 contour than toplines or breath groups involves testing whether baseline slope data obtained by different researchers are replicable. If the visual abstraction procedure were to yield consistent results when used by different researchers examining the same sentences, one could then use this procedure to determine which theories make the best predictions.

A. Procedure

A total of 80 min of the spontaneous speech of seven adult, native speakers of American English (general American dialect) was recorded. Five of these speakers were students at Brown University who were recorded as a group during a dinner conversation. The other two speakers were members of the faculty who were recorded while they lec-

tured to their classes. The recordings were transcribed and complete sentences were extracted from the transcriptions. These complete sentences were then embedded in texts that were either identical, or contextually similar to, the original surrounding speech. The text that each speaker had produced was presented to each of the seven speakers in individual reading sessions that were held at least 3 days after the spontaneous speech was recorded. Each of the speakers thus read aloud the text containing his original spontaneous sentences. Each transcribed sentence and its surrounding text were read twice by each speaker. The speakers were simply asked to read the passages; there was no coaching or interaction on the part of the investigators, no provision of examples of how the sentences should be read.

All the recordings were made using a Nagra type 4.2 tape recorder and a Shure model SM81 condenser microphone. Short, unemphatic, simple declarative sentences were used because the declination theory predicts that these will have a declining baseline. Only 19 short, simple, declarative sentences were present in the corpus of 396 complete sentences extracted from the 80 min of spontaneous speech initially recorded. Each simple, declarative sentence contained only one independent clause, was nonemphatic, and was under 3.5 s in length. They are listed in Table I. The final corpus consisted of 57 sentences (19 spontaneous, 19 first reading, and 19 second reading).

F_0 contours were derived for the total corpus of 57 sentences using a computer system adapted for speech analysis and a pitch extraction algorithm that uses an autocorrelation procedure. The sentences were initially converted to digital signals at a rate of 10 000 samples per second using 10-bit quantization on a Digital Equipment Corporation PDP 11/34 computer. The computed F_0 contours were each visually "checked" by comparing them with the speech waveform. This was done to guard against the inclusion of erroneous data points at half or double the actual F_0. The F_0 contours were then converted to hard-copy output and were xeroxed. Nineteen of these plots (the first reading of each sentence) were presented to seven naive subjects who volunteered to

TABLE I. Short, simple, declarative sentences produced by five subjects.

| Speaker:		
PJ	1.	It'll copy the features of that.
	2.	It's another kind of control phenomenon.
	3.	It's the same basic phenomenon as subject-verb agreement.
	4.	Our mechanism doesn't account for that.
	5.	You can do a separate thing.
	6.	They're quite strong judgments.
PE	7.	I'm thinking of one of these items.
	8.	You could code the alphabet with five bits of information.
	9.	I care about the meaning.
	10.	You can lose information there.
	11.	I want the information there.
	12.	Your long-term store cannot be random.
	13.	You can't get any new information into long-term storage.
	14.	It's like a library.
	15.	This was the whole idea of subliminal perception.
	16.	You have all the time in the world to do it.
JD	17.	Tofu is politically correct.
JK	18.	We should get some cellophane.
MH	19.	He's going to be the new Freud.
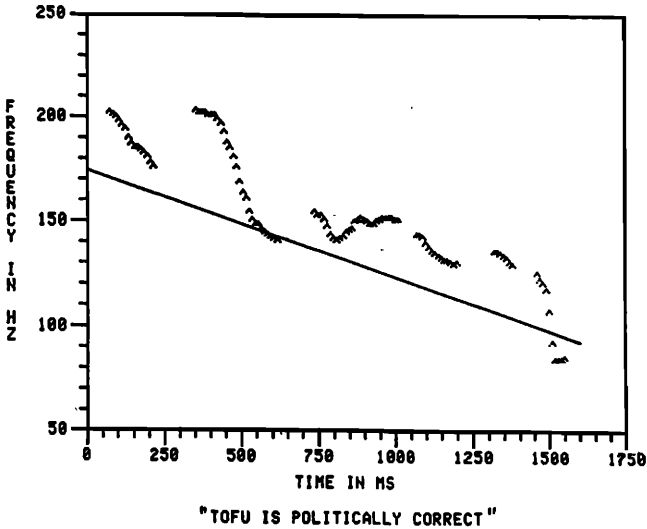
FIG. 1. Example of a visually determined baseline drawn by a naive subject.

participate in this study, as well as to the five authors of the study. A set of criteria was constructed based directly upon Maeda's visual abstraction technique. All of the subjects were asked to construct baselines in accordance with these criteria: (a) The baseline is the single straight line that best fits the valleys of the F_0 curve for the whole contour; (b) if an initial peak is present, the first valley following that peak should serve as a reference point to be included in the baseline; and (c) if at all plausible, a baseline of declining slope should be constructed.

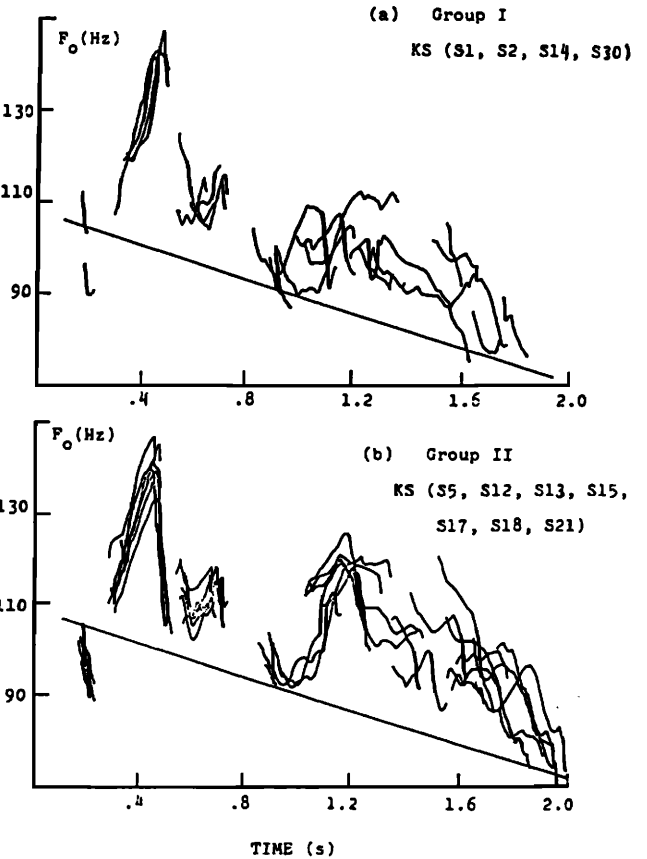Figure 1 shows a representative F_0 plot with a line fit to

FIG. 2. Examples of baselines from Maeda (1976) shown to subjects to facilitate the baseline drawing task.

TABLE II. Slopes of the lines fit by naive subjects and experimenters to the F_0 plots of 19 American English sentences.

| Sentence | Subject | | | | | | | Mean | Standard deviation |
	1	2	3	4	5	6	7		
1	−75.0	−85.0	−65.0	−47.0	−61.0	−47.0	−64.0	−63.43	13.83
2	42.0	29.0	−19.0	−4.0	−37.0	−11.0	61.0	8.71	35.73
3	−45.0	−35.0	−41.0	−28.0	−72.0	−12.0	−37.0	−38.57	18.25
4	−46.0	−40.0	−38.0	−29.0	−41.0	−11.0	−44.0	−35.57	12.12
5	−120.0	−72.0	−40.0	−70.0	−170.0	−57.0	−109.0	−91.14	44.66
6	−80.0	−69.0	−94.0	−39.0	−50.0	−78.0	−77.0	−69.57	18.93
7	−27.0	−33.0	−39.0	−19.0	−27.0	−24.0	−69.0	−34.00	16.70
8	−22.0	−25.0	−2.0	−4.0	−3.0	−16.0	−59.0	−18.71	20.10
9	−17.0	−18.0	33.0	−14.0	−16.0	−81.0	−44.0	−22.43	34.47
10	−21.0	−38.0	−23.0	−27.0	−43.0	−24.0	−68.0	−34.86	16.77
11	200.0	203.0	−5.0	52.0	−58.0	29.0	200.0	88.71	110.34
12	−19.0	−36.0	−18.0	−26.0	−30.0	−7.0	3.0	−19.00	13.47
13	−20.0	−28.0	−8.0	−18.0	−4.0	−2.0	−20.0	−14.29	9.69
14	45.0	419.0	−31.0	55.0	600.0	−648.0	37.0	68.14	393.77
15	−20.0	−6.0	−7.0	−9.0	−11.0	−7.0	−12.0	−10.29	4.82
16	−17.0	−22.0	−17.0	−29.0	−15.0	−18.0	−28.0	−20.86	5.64
17	−49.0	−56.0	−56.0	−54.0	−39.0	−56.0	−69.0	−54.14	9.01
18	−27.0	−36.0	−54.0	−36.0	−71.0	−39.0	−36.0	−42.71	14.85
19	−63.0	−70.0	−133.0	−77.0	−46.0	−57.0	−72.0	−74.00	28.00

it by one of the subjects. An example figure from Maeda (1976), Fig. 2, was also provided to guide the subjects.

B. Results and discussion

The slopes of the baselines that the subjects fit to the F_0 plots were calculated (19 sentences × 12 subjects = 228 baselines total). The uncertainty with regard to the slope of these calculations is in the order of 5 Hz/s. This uncertainty follows from the resolution of the frequency and time measurements on the graphs. The calculated slopes are presented in Table II as a matrix for each sentence and subject, and, in Table III, as a matrix for each sentence and experimenter. Standard deviations were calculated for the slopes of each sentence for the group of experimenters and naive subjects,

respectively. Note that roughly half of the naive subjects, despite instructions to the contrary, fit three of the sentences with baselines that did not show any declination. Since Maeda (1976) and other declination studies do not take into account sentences with nondeclining baselines, sentences 2, 12, and 15 were excluded from calculations of group standard deviation.

Both the subjects (who had no previous background in speech research or linguistic analysis) and the researchers showed great variation in the slopes of the baselines that they fit to particular F_0 contours. The standard deviation of the naive subjects (18 Hz/s) is larger than that of the experimenters (12 Hz/s). Neither group produces suitably replicable results using this procedure. Figure 3 shows the baselines

TABLE III. Slopes of the lines fit by naive subjects and experimenters to the F_0 plots of 19 American English sentences.

| Sentence | Experimenter | | | | | Mean | Standard deviation |
	1	2	3	4	5		
1	−65.0	−60.0	−60.0	−47.0	−30.0	−52.40	14.19
2	7.0	−65.0	−47.0	−23.0	−9.0	−27.40	28.89
3	−24.0	−23.0	−13.0	−32.0	−19.0	−22.20	6.98
4	−23.0	−6.0	−45.0	−30.0	25.0	−25.80	14.02
5	−72.0	−38.0	−94.0	−67.0	−67.0	−67.60	19.96
6	−51.0	−79.0	−6.0	−7.0	−14.0	−31.40	32.38
7	−15.0	−20.0	−23.0	−11.0	−19.0	−17.60	4.67
8	−5.0	10.0	−2.0	−1.0	−2.0	0.0	5.79
9	−3.0	−8.0	−16.0	−14.0	−24.0	−13.00	8.00
10	−35.0	−14.0	−23.0	−19.0	−52.0	−28.60	15.21
11	18.0	−19.0	25.0	8.0	72.0	20.80	33.15
12	−6.0	−1.0	−21.0	−21.0	−16.0	−13.00	9.08
13	−19.0	0.0	−18.0	−19.0	−15.0	−14.20	8.11
14	44.0	38.0	5.0	41.0	125.0	50.60	44.47
15	−5.0	−34.0	−8.0	−8.0	−2.0	−11.40	12.88
16	−7.0	−37.0	−37.0	−18.0	−11.0	−22.00	14.25
17	−16.0	−23.0	−44.0	−61.0	−62.0	−41.20	21.21
18	−4.0	−44.0	−36.0	−37.0	−40.0	−32.20	16.07
19	−41.0	−30.0	−46.0	−27.0	−52.0	−39.20	10.57

that were drawn by all the subjects and experimenters for two of the sentences. The variation in slope of the baselines drawn by different subjects illustrated in Fig. 3 is typical. It is thus difficult to evaluate the claims of studies like Cohen *et al.* (1982) or Maeda (1976) which are based on baselines derived by means of direct, visual inspection of the F_0 plot. Our data show that, while one person may fit a given F_0 plot with a baseline showing extreme declination, another person may fit the same F_0 plot with a baseline that has slight declination, or is nearly level. Our data likewise indicate it may be difficult to assess the claims of theories (e.g., Breckenridge and Liberman, 1977) which advocate "topline" models of F_0. Although we did not test the ability of subjects to fit F_0 plots with toplines, it is most unlikely that their performance of this task would be more consistent. The visual abstraction procedure, when applied to fitting baselines, does not appear to be replicable.

II. EXPERIMENT 2: QUANTITATIVE LEAST-SQUARES-FIT ALL-POINTS LINES

A quantitative method that does not depend on the subjective impressions of a human observer appears to be desirable given the lack of replication demonstrated in experi-
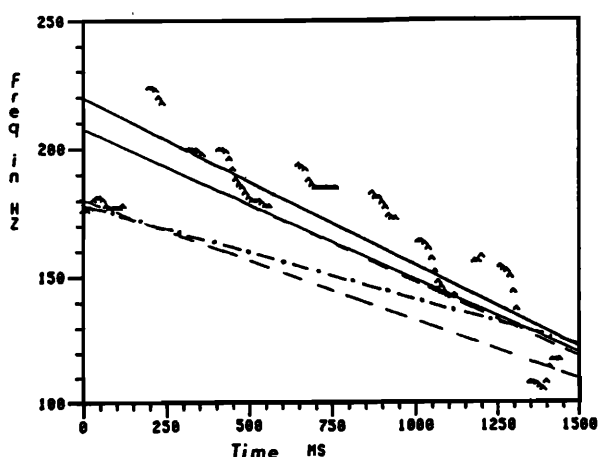
"It'll copy the features of that"
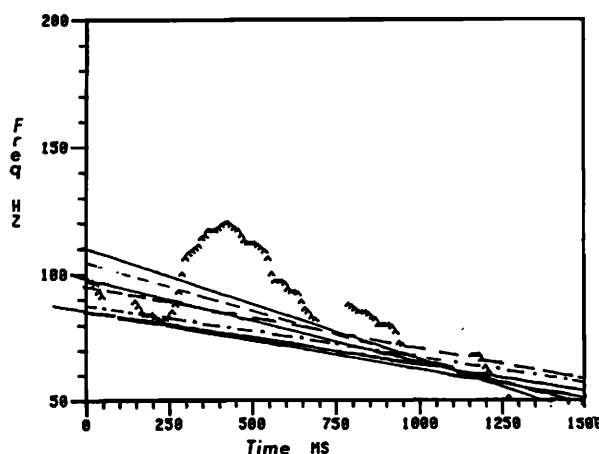
"You can lose information there"

FIG. 3. Representative series of baselines drawn for two F_0 plots by the five experimenters (top) and seven naive subjects (bottom).

ment 1. Qualitative inspection of the F_0 plots of our corpus suggested that the complete F_0 contour could, in many cases, be fit by a straight line that would be a "better" fit to the global F_0 contour than any lines connecting either the valleys or peaks of the F_0 contour. We, therefore, tested this hypothesis mathematically using an established linear regression technique for fitting theoretically linear sets of data points. The linear regression technique has several inherent advantages compared to the visual abstraction procedure. (a) It is easily calculated using a computer. (b) It inherently determines the slope of the "best fit" line. This parameter has been of general interest in recent studies of intonation (e.g., Cohen *et al.*, 1982). (c) a readily calculable adequacy measure, the Pearson product moment correlation coefficient, exists for this metric. This last point is quite important since it provides an objective procedure for the evaluation of the descriptive adequacy of different theories.

A. Procedure

We randomly selected a subset of 15 sentences from our corpus for use in a comparison of bottomline, topline, and all-points linearity. These sentences included five that were drawn from the corpus of spontaneous utterances, five from the first readings of these same sentences, and five from the second readings of these same sentences. For each sentence, we derived the "best fit" baselines for the valleys, "best fit" toplines for the local peaks, and "best fit" all-points lines for the entire F_0 contour. We also computed the Pearson correlation coefficients for each baseline, topline, and all-points line. The Pearson correlation coefficient provides a quantitative metric of whether a particular set of data points is actually in a linear relationship. Levels of significance were assigned to each line based on the value of the correlation coefficient and the number of degrees of freedom. The valley, peak, and all-points linear regression lines were calculated on a general-purpose digital minicomputer based upon the following analysis. Given n points $[(x_i,y_i)...(x_n,y_n)]$, the straight line $y = mx + b$, which minimizes the sum of squares error:

$$L = \sum_{i=1}^{n} [y_i - (mx_i + b)]^2$$

has slope (m):

$$m = n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \right)\left(\sum_{i=1}^{n} y_i \right)$$
$$\times \left[n\left(\sum_{i=1}^{n} x_i^2 \right) - \left(\sum_{i=1}^{n} x_i \right)^2 \right]^{-1}$$

and y-axis intercept (b):

$$b = \left(\sum_{i=1}^{n} y_i - m \sum_{i=1}^{n} x_i \right) n^{-1}.$$

The Pearson product moment coefficient for these points is given by the following formulas adopted from Larsen and Marx (1981):

$$r(x,y) = [E(xy) - E(x)E(y)]/\sqrt{\text{var}(x)}\sqrt{\text{var}(y)},$$

where

$$E(x,y) = \sum_{i=1}^{n} (x_i y_i/n), \quad (E = \text{expectation})$$

$$E(x) = \sum_{i=1}^{n} x_i/n, \quad E(y) = \sum_{i=1}^{n} y_i/n,$$

$$\mathrm{Var}(x) = \sum_{i=1}^{n} [x_i - E(x)]^2/n,$$

$$\mathrm{Var}(y) = \sum_{i=1}^{n} [y_i - E(y)]^2/n.$$

Substituting these definitions into the equation for $r(x,y)$ yields

$$r(x,y) = n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$$

$$\times \left[\sqrt{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\right.$$

$$\times \left.\sqrt{n \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}\right]^{-1}.$$

While usually applied to two random variables, this technique is well suited to fitting a line through a set of data points, where only one variable is random. In our analysis, x (the independent variable) represents time and y (the random variable) represents fundamental frequency. Thus (m), in the above analysis, is the slope $\Delta F_0/\Delta t$ of the least-squares fit line, and b is the F_0 value of the line at $t = 0$.

We used the following criteria to define the peaks and valleys of each F_0 contour. These points were then used to fit the peak and valley lines. (a) In general, the peaks/valleys of a contour are the local maxima/minima. (b) A given point may be labeled a peak/valley only when it occurs after at least a 10-Hz rise/fall from nonpeak/valley portions of the contour and is followed by at least a 10-Hz fall/rise. (c) If the initial/final points of the F_0 contour are local maxima or minima extending 10 Hz above/below the contour, they are labeled peaks/valleys.
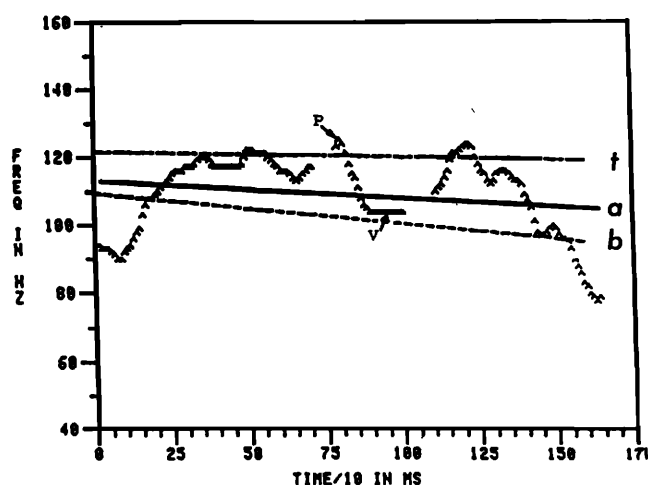

FIG. 4. The sentence "Your long-term store cannot be random" fitted using the linear regression procedure with an all-points line (A), a baseline (B), and a topline (T). A sample peak (P) and valley (V) are indicated.

Figure 4 shows a sentence that was fit using the linear regression procedure with an all-points line (A), a baseline (B), and a topline (T). A sample peak and valley, determined using the above criteria, are indicated.

B. Results and discussion

We randomly selected five spontaneous and the ten corresponding read sentences from our total corpus and compared the degree to which the local valleys of the F_0 plots could be fit by linear "bottomlines," peaks by linear "toplines," and the entire F_0 contour by an all-points linear regression line. Table IV lists this 15 sentence subset, and gives the level of significance and degrees of freedom for bottomline, topline, and all-points linear regression fits.

The results of the analysis are plotted in Fig. 5. The bar graphs pertain to the F_0 plots pooled across all speaking conditions, spontaneous and read. The conditions in which data points could not be fit with a straight line, i.e., were nonlinear, are grouped into one category at the bottom. The fit of
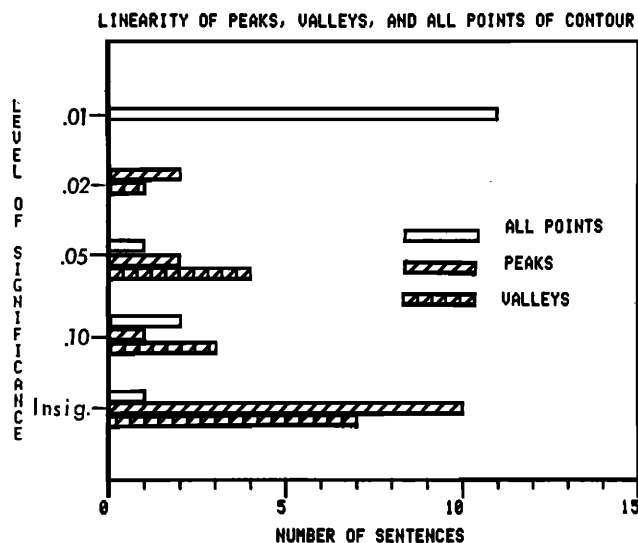
TABLE IV. Levels of significance and degrees of freedom for bottomline, topline, and all-points linear regression lines ($n = 15$ sentences; five read, ten spontaneous).

	Baseline				Topline				All points			
	m	r	df	Signif	m	r	df	Signif	m	r	df	Signif
Read sentence												
1	− 10.4	− .7556	2	0	− 14.8	− 0.2704	3	0	− 20.9	− 0.369	103	0.01
2	− 61.2	− 0.9537	3	0.02	− 59.3	− 0.9506	3	0.02	− 60.7	− 0.896	99	0.01
3	− 32.6	− 0.8416	2	0	+ 14.5	+ 0.9650	2	0.05	− 17.8	− 0.775	108	0.01
4	+ 28.3	+ 0.6695	3	0.10	+ 28.8	+ 0.2878	3	0	+ 29.4	+ 0.642	113	0.01
5	− 9.8	− 0.5715	1	0	+ 3.1	+ 0.1397	1	0	− 9.0	− 0.227	87	0.05
6	− 22.19	− 0.9948	1	0.10	− 5.3	− 0.1814	2	0	− 18.0	− 0.425	101	0.01
7	− 30.0	− 0.9549	2	0.05	− 28.2	− 0.9602	2	0.05	− 29.8	− 0.864	114	0.01
8	− 9.0	− 0.3684	3	0	− 1.8	− 0.2117	3	0	− 10.0	− 0.434	203	0.01
9	− 60.0	− 0.8023	4	0.05	− 80.9	− 0.8391	5	0.02	− 80.2	− 0.806	100	0.01
10	− 34.9	− 0.9273	2	0.05	− 29.3	− 0.7988	2	0	− 31.9	− 0.702	88	0.01
Spontaneous sentence												
1	− 20.0	− 0.7243	3	0	− 38.6	− 0.7338	3	0	− 38.3	− 0.765	124	0.01
2	− 56.2	− 0.8328	3	0.10	− 64.4	− 0.7426	4	0.10	− 51.3	− 0.811	116	0.01
3	− 24.1	− 0.6313	1	0	− 32.8	− 0.7559	1	0	− 21.6	− 0.590	179	0.10
4	+ 48.0	− 0.7935	4	0.05	+ 69.6	+ 0.7691	3	0	+ 43.1	+ 0.673	112	0.10
5	+ 5.1	0.5650	2	0	− 17.8	− 0.3641	2	0	− 2.7	− 0.077	92	0

LINEARITY OF PEAKS, VALLEYS, AND ALL POINTS OF CONTOUR

FIG. 5. Results of linearity assessment of all points, peaks, and valleys of 15 F_0 contours.

these data points in a linear relationship did not reach significance. The significantly linear plots are divided into four groups that show increasing levels of significance. It is clear that only the all-points lines achieved a significantly linear correlation at $p < 0.01$ level. Moreover, the linearity hypothesis for the all-points lines failed in only one of the 15 sentences of the pilot corpus.

In contrast, the data show that the apparent visual linearity which may be attributed to the local maxima and minima of F_0 contours is not a mathematical reality in more than 50% of the utterances examined. Furthermore, the linearity hypothesis for the all-points lines failed in only five of the 57 sentences of the entire corpus. Clearly, claims for the linearity of functions connecting the valleys (bottomlines) or the peaks (toplines) of F_0 contours are not consistent with these data. The all-points linear regression line appears to best describe the linear component of F_0 present in these F_0 contours. For this reason, we derived all-points "best-fit" lines for all 57 sentences in the corpus for use in subsequent analyses.

III. EXPERIMENT 3: SPONTANEOUS VERSUS READ SPEECH; BREATH-GROUP VERSUS DECLINATION

Previous studies (Lieberman and Tseng, 1981; Umeda, 1982) have noted greater variability and less declination in spontaneous speech than in read speech. Lieberman and Tseng (1981), Lieberman et al. (1982), and Tseng (1981) also note that spontaneous speech F_0 contours more often fit the breath-group model of intonation than declination theories.

At the descriptive level, the differing claims of the breath-group and declination theories hinge upon the way in which F_0 falls in the course of a simple declarative sentence. Declination theories claim that there is a gradual descent over the course of an intonation group (Pierrehumbert, 1979; Cooper and Sorenson, 1981). The breath-group model (Lieberman, 1967) differs from declination models insofar as it claims that the F_0 contour consists of two elements: a nonterminal segment or phase and a terminal segment or phase.

The terminal is basically defined as a final rapid rise or fall in F_0; terminals appear to function as linguistic cues to mark out the utterances' phrase structure (Trager and Smith, 1951; Lieberman, 1967; Vanderslice and Ladefoged, 1972; Atkinson, 1973). The nonterminal portion of the breath-group, in its most basic form, tends to be level. The nonterminal F_0 contour may be perturbed by F_0 variations that mark out phonetic prominence, emphasis, or phonemic tones in languages like Chinese. It can, but does not necessarily, have a downwards slope (as declination theories claim). This distinction is important.

Declination theories claim that the downwards slope of the F_0 contour throughout the sentence is a psychologically "real," universally present, acoustic cue that listeners track in order to segment the flow of speech into syntactically meaningful segments (e.g., Maeda, 1976) or to assign F_0 peaks to phenomena such as linguistic stress (e.g., Pierrehumbert, 1979). The breath-group theory, like earlier theories (e.g., Armstrong and Ward, 1926; Jones, 1932; Trager and Smith, 1951), instead claims that the terminals are used by listeners to segment the flow of speech into syntactically meaningful segments.

The first level of descriptive adequacy that a theory must meet is whether the phenomena that it proposes are physically present. We, therefore, used the all-points least-squares fit procedure to see how well F_0 contours could be described when they were divided into a terminal and nonterminal phase following the breath-group theory, versus how well they could be described using overall declination.

The basic prediction of the breath-group theory (Lieberman, 1967) for unemphatic, simple declarative sentences is that these sentences will usually have intonation contours that consist of a nonterminal contour that does not have to show any declination, followed by a terminal that shows strong declination. The terminal is, in other words, independent of the rest of the F_0 contour with respect to slope and always falls. For the breath-group theory to hold for spontaneous speech, there should be a diminution of declination, i.e., a noticeable increase in the mean slope and slope standard deviation for sentences in which a terminal portion has been removed.

A. Procedure

The linear fit of the all-points lines to the F_0 contours provides the basis for objective comparison of the difference in declination between spontaneous and read speech. We, therefore, grouped and compared the slopes of the all-points linear regression lines according to whether they were derived from spontaneous or read speech samples. Mean slope and slope standard deviations were calculated for these two groupings.

Our test of whether the terminal is independent of the nonterminal portion of the contour was fairly simple; we derived the nonterminal F_0 contour of each utterance in our corpus by simply removing the last 150 ms of the F_0 data points. A more precise test would involve determining the duration of the terminal, which varies from speaker to speaker (Atkinson, 1973). Mean slope and slope standard deviations were calculated for nonterminal contours in both

read and spontaneous conditions. Thus a four-way comparison of read/spontaneous and terminal/nonterminal contours was undertaken.

B. Results

The results of these analyses are as follows:

(a) There is a greater degree of variation in slope present in spontaneous than in read speech. The standard deviation for the spontaneous sentences was approximately 50 Hz/s; that of the read sentences was 24 Hz/s. The histogram shown in Fig. 6 shows the percent of sentences having slopes within 10 Hz/s of the center of each of the bars. It may be observed that the samples of read speech have slopes that tend to cluster about the − 20 Hz/s value, whereas the slopes derived from spontaneous speech are more evenly distributed. Further, while 86% of the read sentences showed declination, only 66% of the spontaneous sentences showed declination.

(b) Figure 7 shows the distribution of slope for the nonterminal F_0 contours of the sentences that had been uttered spontaneously versus the nonterminal contours of the sentences that had been read. The distribution of slopes for the read condition is very similar to that shown in Fig. 6 in which

FIG. 7. Frequency distribution of slopes of all-points linear regression lines for spontaneous ($n = 19$) and read ($n = 38$) sentences. The 150-ms-long terminal portion was removed from each contour. Slopes of sentences: terminals removed.

the terminals were included. The distribution function for the sentences that were uttered in spontaneous speech, however, differs from that of Fig. 6. There is even greater variation than for spontaneous sentences with terminal portions removed. The standard deviation in spontaneous speech slopes is 21 Hz/s greater with the terminal removed than when the terminal is present (representing a 42% increase). In addition, the percent of spontaneous sentences that had declining slopes decreases from 66%–55% of the sample.

C. Discussion

One of the more striking aspects of the pattern of variation of the slope of utterances is the extreme variability for both read and spontaneous speech. Given this variability, it is difficult to see how listeners could track a sentence level declination slope for an English sentence to derive putative linguistic distinctions signaled by subtle variations in F_0. The small 5-to 15-Hz variations in F_0 that are often associated with distinctions in stress in linguistically oriented studies (e.g., Pierrehumbert, 1979) would be completely obscured by the overall variability in F_0. The arguments given by Pierrehumbert (1979) for the psychological reality of declination,
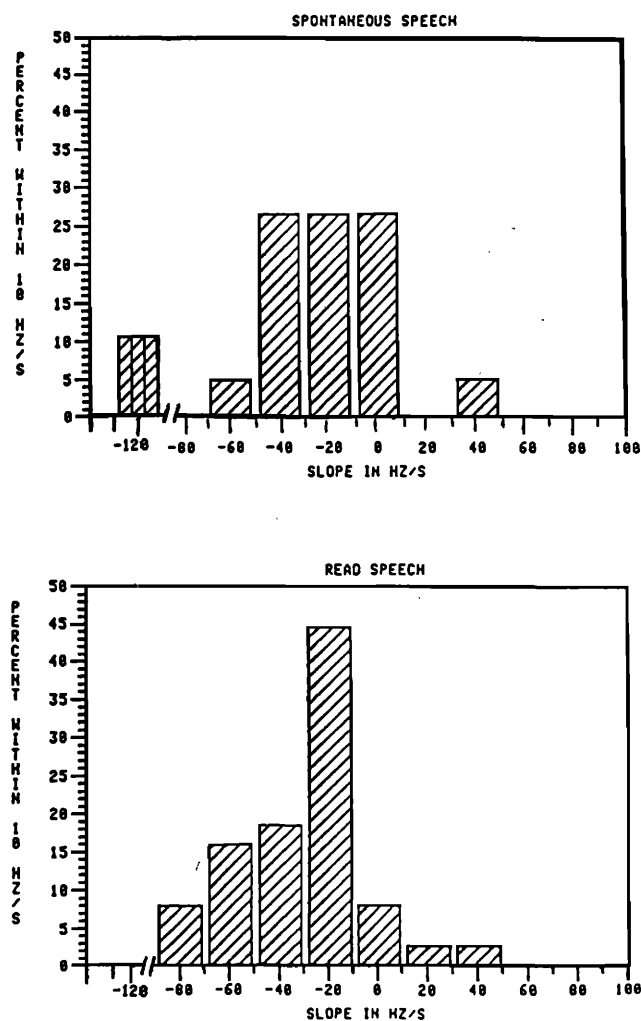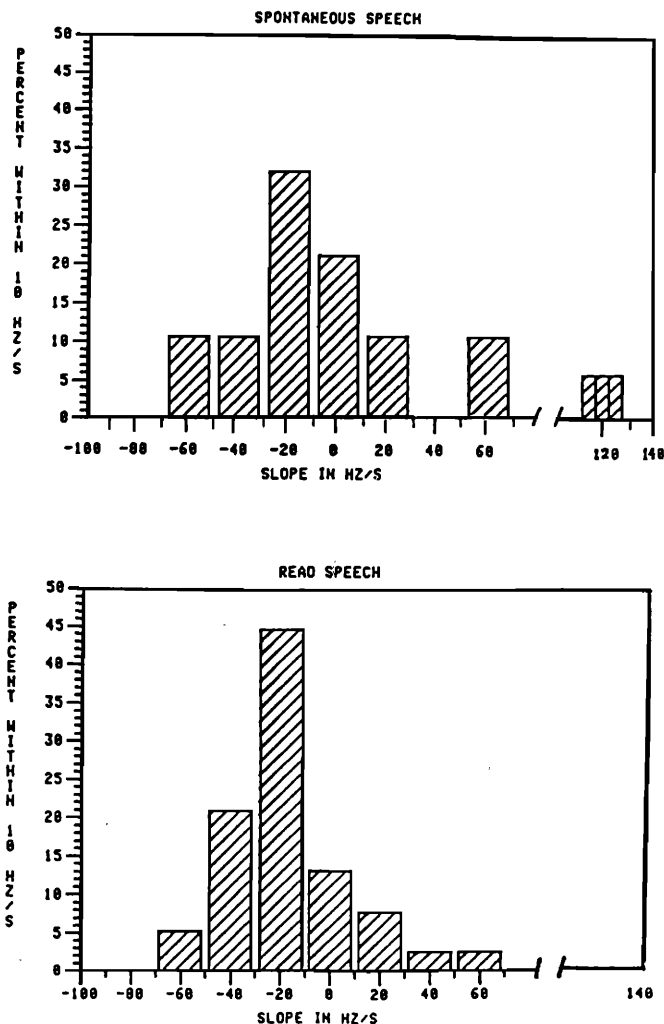

FIG. 6. Frequency distribution of slopes of all-points linear regression lines for spontaneous ($n = 19$) and read ($n = 38$) sentences. Slopes of sentences: terminals included.

moreover, depend on the listeners having, for any given language, a clear notion of a putative "standard" average declination.

The variability present in the slopes of the sentences that we measured also bears on another issue. Linguists tend to think of variation as a "performance" effect that masks a speaker's "competence" (Chomsky, 1965). The procedures used to collect the corpora of speech for linguistic studies of intonation often implicitly assume that the "competence" of a speaker will be manifested by the reduction of variation. A number of techniques have been used which appear to have as their goal the reduction of variation by means of explicit control of the speakers' productions. In a number of experiments, subjects have thus been "prepared" for their productions or have been asked to reread sentences until their productions match the expectations of the experimenters (Cooper and Sorensen, 1977, 1981; Danly and Shapiro, 1983). Similarly, a number of researchers have served as their own subjects (O'Shaugnessy, 1979; O'Shaughnessy and Allen, 1983) or have relied on a single subject's utterances (Maeda, 1976; Poser, 1983). Although such techniques do allow for a greater degree of control over variation, they run the risk of being circular, first assuming that a type of intonation "competence" exists and then adducing F_0 data as proof for it. The data that we have discussed show that a great deal of variability exists; however, one characterizes the overall F_0 contour. Techniques that limit this variability are inherently suspect and may lead to inappropriate correlations between linguistic constructs and putative F_0 "cues" that would be lost in the pattern of F_0 variation typical of normal read or spontaneous speech.

The data are also not consistent with the strong claims of theories like Pierrehumbert (1979) and Maeda (1976) which state that declination is a linguistic universal. The data further suggest that read speech may not be the most appropriate material on which to base or test theories for the linguistic functions of intonation. Both breath-group and declination theories have as their focus the role of F_0 contours in segmenting the flow of speech into syntactic groupings. Both theories claim that the F_0 contour contains cues that most often delimit stretches of speech that approximate well-formed sentences. It would appear to be appropriate to test these theories on samples of spontaneous speech, since these theories attempt to account for how human beings actually derive the syntactic structure of utterances. Reading is a comparatively recent invention in human history; moreover, most human communication is spontaneous, not read from text.

The data of experiment 3 show that sentences derived from spontaneous speech do not show as consistent a pattern of declination as do similar sentences read by the same speakers. It is evident from the data for spontaneous speech that terminals are generally falling for simple declarative sentences, and that the nonterminal part of the F_0 contours have less declination than do those derived from read speech. This is consistent with the claims of the breath-group theory. It is also germane to note that the data demonstrate that neither declination nor breath-group theories of intonation adequately describe all of the sentences of the corpus. This is consistent with the results of previous studies (Armstrong and Ward, 1926; Lieberman and Tseng, 1981; Tseng, 1981) that took account of sentences derived from spontaneous speech.

IV. CONCLUDING SUMMARY

The experiments show that data on the slopes of F_0 contours derived by means of the visual abstraction procedure, previously used in studies of declination, have too little reliability. It is further demonstrated that an "all-points" line derived by means of a quantitative best-fit least-squares error criterion is a better descriptor of sentence F_0 contours in speech than either bottom-or top-line declination models. Analyses of similar sentences drawn from spontaneous and read speech uttered by the same speaker, using the all-points model, show differences between the read and spontaneous conditions. There is more variation present in spontaneous speech; the slopes of the F_0 contours derived from spontaneous speech tend to depart from the declination model more often than is the case with F_0 contours derived from read sentences. The F_0 contours derived from spontaneous speech fit the breath-group model of intonation better than those derived from read material.

The strong claims of declination theories, i.e., that are downwards slope of the F_0 contour is a necessary and universal acoustic cue that segments the flow of speech into sentences, are not consistent with the data, particularly the material derived from spontaneous speech. The F_0 plots, in particular those derived from spontaneous speech, show that neither falling F_0 terminals of the breath-group theory nor global, sentence-long declinations are always present. Human listeners can apparently determine the syntactic structure of speech in the absence of specific intonational cues, using their general knowledge of the lexicon, morphology, and syntax of English. Similar results have been noted for Mandarin Chinese (Tseng, 1981), and probably are characteristic for all human languages and speakers. The data show that F_0 cues are often present that are in accord with the predictions of various theories of intonation. The data, however, argue for further studies that analyze F_0 contours derived from spontaneous, as well as read speech, using objective, replicable metrics like all-points linear regression lines.

Atkinson, J. R. (1973). "Aspects of intonation in speech: Implications from an experimental study of fundamental frequency," Ph.D. dissertation, University of Connecticut.

Armstrong, L. E., and Ward, I. C. (1926). Handbook of English Intonation (Teubner, Berlin).

Breckenridge, J., and Liberman, M. Y. (1977). "The declination effect in perception," J. Acoust. Soc. Am. 61, 51.

Chomsky, N. A. (1965). Aspects of the Theory of Syntax (MIT, Cambridge, MA).

Cohen, A., Collier, R., and 't. Hart, J. (1982). "Declination: Construct or intrinsic feature of speech?," Phonetica 39, (4–5), 254–273.

Cooper, W. E., and Sorensen, J. M. (1977). "Fundamental frequency contours at syntactic boundaries," J. Acoust. Soc. Am. 62, 682–692.

Cooper, W. E., and Sorensen, J. M. (1981). Fundamental Frequency in Sentence Production (Springer, New York).

Danly, M., and Shapiro B. (1983). "Speech prosody in Broca's aphasia," Brain Lang. 16, 171–190.

Delgutte, B. (1978). "Technique for the perceptual investigation of F_0 contours with application to French," J. Acoust. Soc. Am. **64**, 1319–1332.

Garding, E. (1979). "Sentence intonation in Swedish," Phonetica **36**, 207–215.

Jones, D. (1932). *An Outline of English Phonetics* (Dutton, New York).

Larsen, R. J., and Marx, M. L. (1981). *An Introduction to Mathematical Statistics and Its Applications* (Prentice-Hall, Englewood Cliffs, NJ).

Lieberman, P. (1967). *Intonation, Perception, and Language* (MIT, Cambridge, MA).

Lieberman, P., Landahl, K., and Ryalls, J. H. (1982). "Sentence intonation in British and American English," J. Acoust. Soc. Am. Suppl. 1 **71**, S112.

Lieberman, P., and Tseng, C. Y. (1981). "On the fall of the declination theory: Breath group versus 'declination' as the base from for intonation," J. Acoust. Soc. Am. Suppl. 1 **67**, S63.

Maeda, S. (1976). "A characterization of American English intonation," Ph.D. thesis, MIT (unpublished).

O'Shaughnessy, D. (1976). "Linguistic features in fundamental frequency patterns," J. Phonet. **7**, 119–145.

O'Shaughnessy, D., and Allen, J. (1983). "Linguistic modality effects on fundamental frequency in speech," J. Acoust. Soc. Am. **74**, 1155–1171.

Poser, W. J. (1983). "On the mechanism of F0 downdrift in Japanese," J. Acoust. Soc. Am. Suppl. 1 **71**, S69.

Pierrehumbert, J. (1979). "The perception of fundamental frequency declination," J. Acoust. Soc. Am. **66**, 363–369.

Trager, G. L, and Smith, H. L. (1951). *Outline of English Structure, Studies in Linguistics, No. 3* (Battenberg, Norman, OK).

Tseng, C.Y. (1981). "An acoustic study of tones in Mandarin," Ph.D. dissertation, Brown University (unpublished).

Umeda, N. (1982). " 'F0 declination' is situation dependent," J. Phonet. **10**, 279–290.

Vanderslice, R., and Ladefoged, P. (1972). "Binary suprasegmental features and transformational word-accentuation rules," Language **48**, 819–838.

Vaissiere, J. (1974). "On French prosody," Res. Lab. Electron. Q. Prog. Rep. No. 114, MIT, 212–213.