

# The role of fundamental frequency in signaling linguistic stress and affect: Evidence for a dissociation

GERALD W. McROBERTS, MICHAEL STUDDERT-KENNEDY, and DONALD P. SHANKWEILER  
*University of Connecticut, Storrs, Connecticut  
and Haskins Laboratories, New Haven, Connecticut*

The fundamental frequency ( $F_0$ ) of the voice is used to convey information about both linguistic and affective distinctions. However, no research has directly investigated how these two types of distinctions are simultaneously encoded in speech production. This study provides evidence that  $F_0$  prominences intended to convey linguistic or affective distinctions can be differentiated by their influence on the amount of final-syllable  $F_0$  rise used to signal a question. Specifically, a trading relation obtains when the  $F_0$  prominence is used to convey emphatic stress. That is, the amount of final-syllable  $F_0$  rise decreases as the  $F_0$  prominence increases. When the  $F_0$  prominence is used to convey affect, no trading relation is observed.

Speech communications often convey both linguistic content and the speaker's affective state. It is sometimes supposed that linguistic information and affect are independently encoded in the acoustic speech signal, since the same sentence may be spoken with varying affective tones of voice and, in consequence, take on different meanings (e.g., Scherer, Ladd, & Silverman, 1984). Indeed, studies showing that a speaker's intended affect can be identified in languages unfamiliar to the listener, or when the verbal content is removed by filtering, support this assumption (e.g., Davitz, 1964; Kramer, 1964; McCluskey, Albas, Niemi, Cuevas, & Ferrer, 1975; Starkweather, 1961). Further support comes from studies of hemispheric specialization, which suggest different degrees of involvement of the left and right hemispheres for linguistic and affective aspects of speech (e.g., Blumstein & Cooper, 1972; Heilman, Bowers, Speedie, & Coslett, 1984; Ley & Bryden, 1982; Shipley-Brown, Dingwall, Berlin, Yeni-Komshian, & Gordon-Salant, 1988; Tucker, Watson, & Heilman, 1977; Weintraub, Mesulam, & Kramer, 1981; Zurif, 1974). However, little research has directly addressed the issue of how linguistic and affective aspects of speech are encoded during speech production. The purpose of this study was to investigate the relation between linguistic and affective uses of voice fundamental frequency ( $F_0$ ).

Speakers use fundamental frequency to convey several linguistic distinctions, including both segmental features,

such as consonant voicing and vowel height, and prosodic, or suprasegmental, features, which typically extend over more than a single segment. Intonation belongs to the prosodic aspect of language; the term refers to variation of  $F_0$  for linguistic purposes. The acoustic manifestation of intonation is the fundamental frequency of the voice, which contributes to such linguistic distinctions as sentence type. For example, questions and statements are characterized by different fundamental frequency contours in English and in many other languages: Declaratives typically have a gradual decline in  $F_0$  from beginning to end, whereas questions (especially syntactically unmarked yes-no questions) tend to have an elevated or rising intonation contour (e.g., Bolinger, 1978; Ultan, 1978), either over the entire utterance or over its final syllable(s). A further linguistic use of  $F_0$  in production is contrastive stress, in which one or more words in a sentence may carry added stress to denote contrastive emphasis. Increases in degree of emphasis are associated with increases in  $F_0$  (Bolinger, 1958; Fry, 1955, 1958).

Fundamental frequency also conveys paralinguistic information, such as the affective state of the speaker. Few studies of spontaneously produced affective speech have been done. However, in those cases where spontaneous emotional utterances have been recorded and analyzed, both average  $F_0$  and  $F_0$  range are typically increased in comparison with less affectively marked speech (e.g., Williams & Stevens, 1969, 1972, 1981). Affective expressions simulated by actors are consistent with the results obtained for spontaneous speech in showing higher  $F_0$  for happiness, anger, and sometimes sadness (e.g., Fairbanks & Hoaglin, 1941; Fairbanks & Pronovost, 1939; Williams & Stevens, 1972; see Scherer, 1986, for a review).

Despite the fact that  $F_0$  is used to convey both linguistic and affective information, there appear to have been

This work is based on the doctoral dissertation of the first author and was supported by NICHD Grant HD-01994 and NIDCD Grant DC-00403. Correspondence should be addressed to G. W. McRoberts, who is now at the Department of Psychology, Stanford University, Stanford, CA 94305-2180.

—Accepted by previous editor, Charles W. Eriksen

no controlled experimental studies to investigate how linguistic uses of *F0* may be influenced by the simultaneous use of *F0* to convey affect. The present study addresses this question by investigating how linguistic and affective influences on *F0* are encoded in yes-no questions. A dissociation is demonstrated between the use of *F0* variation in the production of stress contrasts and its use to mark positive and negative affect.

### Question Intonation in English

The intuition of traditional phonetics that the distinction between declaratives and questions in English is signaled by a difference in terminal *F0* glide (e.g., Pike, 1945; Uldall, 1960)—specifically, a final-syllable *F0* rise for questions and *F0* fall for declaratives—is only partially supported by data from production. Whereas declarative utterances reliably show a decrease in *F0* over the duration of the utterance (Cohen, Collier, & t'Hart, 1982; Cooper & Sorenson, 1981; Pierrehumbert, 1979), the final-syllable *F0* rise for questions is not obligatory (Cohen, 1972; Fries, 1964). Attempts to relate variation in question contour to syntactic categories (e.g., "wh" questions, tag questions, yes-no questions) have not proved successful (Cohen, 1972). For example, in investigations of yes-no questions produced during radio and television shows (e.g., Fries, 1964; Lee, 1980), only 40%–55% of the questions had a rising intonation pattern.

Various researchers have suggested that differences in the amount of final rise for questions may be associated with paralinguistic factors, such as the attitude or emotion of the speaker (Crystal, 1969; Jassem, 1972; Lee, 1956, 1960). Thus, Lee (1956, 1960) finds that questions with falling endings tend to have a firm and insistent quality. Crystal (1969) suggests that rising endings are more friendly and interested, and Jassem (1972) finds fall-rise endings to be more friendly, familiar, informal or intimate than rising endings.

These observations suggest a relation between the perceived shape of question contour and judgments of the speaker's attitude or emotion.<sup>1</sup> Unfortunately, these suggestions are not based on controlled experiments. Rather, in most cases, judgment of the speaker's attitude was based solely on the investigator's impressions, rather than on judgments by a group of listeners. In addition, since listener judgments often differ from measured *F0* contours (Hadding-Koch & Studdert-Kennedy, 1964; Jassem, 1972), the relation of the perceived intonation contours to the actual *F0* contour for a specific attitude or emotion is unclear.

While careful investigation of the use of final rise in various affective contexts might shed light on this issue, it appears that no systematic acoustic and perceptual study of question intonation has been carried out in which speakers used *F0* to convey a linguistic distinction and simultaneously to convey contrasting attitudes or emotions. The present investigation is intended to clarify the effect of positive and negative affect on question intonation in English.

### Trading Relations in Intonation for Yes-No Questions

A series of perceptual experiments by Hadding-Koch and Studdert-Kennedy (1964, 1965a, 1965b; Studdert-Kennedy & Hadding, 1973) broke new ground by investigating the whole intonation contour of an utterance, rather than just its terminal glide, and by calling both for linguistic judgments of the whole utterance and for psychophysical judgments of the glide. Thus, in their second experiment, Studdert-Kennedy and Hadding (1973) used a speech vocoder to impose parametric variations in *F0* on the naturally spoken utterance *November* [no 'vɛm bæ]. They constructed a set of 72 contours that included patterns typical of a male speaker's intonation in both Swedish and American English yes-no questions. The principal variations were in the height of the peak *F0* on the stressed second syllable and in the extent of the rise or fall in *F0* on the final syllable. They also prepared sets of pulse-train analogs and sine-wave analogs with *F0* contours identical to those of the speech stimuli.

These stimuli were presented in random test orders to groups of Swedish and American listeners, who made both linguistic judgments of the whole utterance (question/statement) and psychophysical judgments of the terminal glide (rise/fall) on the full speech stimuli and for psychophysical judgments of the terminal glide on the sine-wave and pulse-train analogs. The results on the linguistic judgments confirmed the findings of their previous study (with a different utterance) in two respects. First, for both groups of listeners, terminal glide was the single most powerful determinant of question/ statement judgments: all contours judged as *question* over 90% of the time had a rising terminal glide. Second, for both groups of listeners, the height of the peak *F0* on the medial stressed syllable also affected linguistic judgments: a peak of 200 Hz, starting from an initial *F0* of 130 Hz, required less terminal rise than did a peak of 160 Hz for listeners to judge the utterance a question. Studdert-Kennedy and Hadding dubbed this effect a reciprocal *trading relation* between stress peak and terminal rise.

The psychophysical judgments of the full speech stimuli were less consistent than were the question/statement judgments, for both groups of listeners. Overall, contours heard as terminally rising were judged to be questions; however, not all contours judged to be questions were heard as terminally rising. This asymmetry arose because the effect of the peak *F0* was less consistent both within and between subjects for the psychophysical judgments than for the linguistic judgments. Moreover, the effect of the peak *F0* was entirely absent from judgments of the terminal glide on the sine-wave and pulse-train analogs. The authors therefore concluded that the peak *F0*-final rise trading relation was specifically linguistic in its origin.

If this is so, we might suspect that listeners' perceptual systems are attuned to properties that normally obtain in speech, and a parallel trading relation might be found in speech production—that is, as peak *F0* is raised

to convey stress or emphasis in a question, the amount of final rise should be reduced. In fact, an account of the trading relation based on physiological constraints on speech production was proposed by Lieberman in 1967, but the matter has not been pursued further.

The importance of specifying the perception-production link should not be underestimated. The perceptual studies of Studdert-Kennedy and Hadding (1973; Hadding-Koch & Studdert-Kennedy, 1964) used hybrid stimuli formed by imposing various synthetic intonation contours on a single utterance. In making these stimuli, the investigators manipulated *F0* without regard to the constraints of natural speech production. For example, although the range of *F0* values used in constructing the experimental intonation contours was based on spectrographic data from natural productions (Hadding-Koch, 1961), the combinations of *F0* values were arbitrary. Moreover, all the stimuli had the same overall duration. As a result, the stimulus set was artificially constrained and much of the variability in frequency spectrum, amplitude envelope, duration and *F0* characteristic of naturally produced speech was absent in these hybrid stimuli.

The first goal of the present study was to determine whether the relation noted in the Studdert-Kennedy and Hadding perceptual studies normally obtains in naturally spoken utterances by testing for a peak *F0*-final rise trading relation in speech production. To pursue this goal, Experiment 1 was designed to determine whether a trading relation occurred when the peak *F0* on the syllable preceding a question final rise was elevated to convey contrastive stress. The expected effect was observed; therefore, in Experiment 2, we explored whether the trading relation occurred when *F0* was raised for a reason other than stress placement (i.e., in conjunction with variation in affective tone). We predicted that it would not, and our prediction was confirmed.

A more stringent test of the relation between peak *F0* and final rise among contrastively stressed and affectively toned productions would examine only that subset of tokens for which listener judgments coincided with the speakers' intent. Therefore, in Experiment 3, we elicited both question-statement and emotional polarity judgments for the utterances produced in Experiment 2 and then reanalyzed the data of that experiment for those tokens that were perceived by naive listeners in agreement with speaker intent.

## EXPERIMENT 1

The purpose of Experiment 1 was to explore the influence of peak *F0* on the amount of final rise in the production of questions. Specifically, this study attempted to confirm, by analysis of male speakers' productions, the conclusions of Studdert-Kennedy and Hadding that a trading relation exists between peak *F0* and final rise, such that *F0* peaks at 200 Hz require less final rise than do tokens with peak *F0* of 160 Hz or below. Linguisti-

cally relevant variations in peak *F0* were induced by having speakers produce questions with varying degrees of contrastive stress and without contrastive stress. We hypothesized that variations in peak *F0* resulting from differences in contrastive stress would influence the amount of final rise used by speakers when producing questions—in other words, we predicted a trading relation between the peak *F0* and the amount of final rise. To test this hypothesis, acoustic analysis was performed to determine the relation between second-syllable peak *F0* and final-syllable *F0* rise.

## Method

**Subjects.** Four adult males, native speakers of American English and ranging in age from 23 to 35 years, were the talkers. All were from the northeastern or midwestern regions of the United States and without marked regional accents.

**Test utterances.** Two utterances were used in this study: *November* and *I did it*. *November* was included to enable a direct comparison with the findings of Studdert-Kennedy and Hadding (1973). The sentence *I did it* was included for the following reasons: (1) it is meaningful as both question and statement, both with and without contrastive stress, (2) it is voiced throughout, (3) it serves as a phonetic control on *November* by eliminating the difference in vowel quality between the second and third syllables and the nasalization and frication, which might interact with *F0*, and (4) its three constituent syllables are independent morphemes, as opposed to the three syllables in *November*, which are not. Only the questions are considered in the following analyses.

**Procedure.** The experimenter and each talker were seated in a sound booth, with the talker positioned approximately 24 in. from a microphone. The microphone provided input to a remotely controlled tape recorder in an adjacent sound booth. Brief descriptive scenarios provided contextual support for the subject in each condition. For example, in one scenario the speaker responds to a young child who says that Christmas is in November. The response is a contrastively stressed question, "November? ... Don't you mean December?" The *November* scenarios were described by the experimenter, and each talker produced his responses to each scenario. The experimenter, without resorting to models or examples, encouraged the talkers to vary emphasis over a range of values in the contrastive stress condition. The procedure was repeated for the *I did it* scenarios.

Each speaker produced the two sentences in each of three conditions: (1) declarative (statement) with neutral stress, (2) question with neutral stress, and (3) question with contrastive stress. Each speaker produced 4-5 repetitions of each sentence in each condition for a total of approximately 30 tokens per subject and 120 total tokens, of which one third were statements and the two thirds subjected to analysis were questions.

**Acoustic analysis.** Each talker's recorded tokens were digitized at a 10-KHz sampling rate using the Haskins Laboratories PCM (Pulse Code Modulation) system. Input levels were held constant across each talker's entire session. The digitized tokens were stored on disk. Individual tokens were prepared for analysis using the Wave Form Editing and Display (WENDY) software at Haskins Laboratories. Fundamental-frequency analysis was performed using Interactive Laboratory System (ILS) software (Signal Technology, Inc., 1978). The parameter settings for initial analysis of all tokens included a 10-msec sampling window with a 50% overlap between adjacent windows, minimum and maximum *F0* values of 75 Hz and 400 Hz, respectively, and a voicing threshold value of -400. (The sampling window and overlap parameters provide for some smoothing of the cycle-to-cycle variation in *F0*, especially

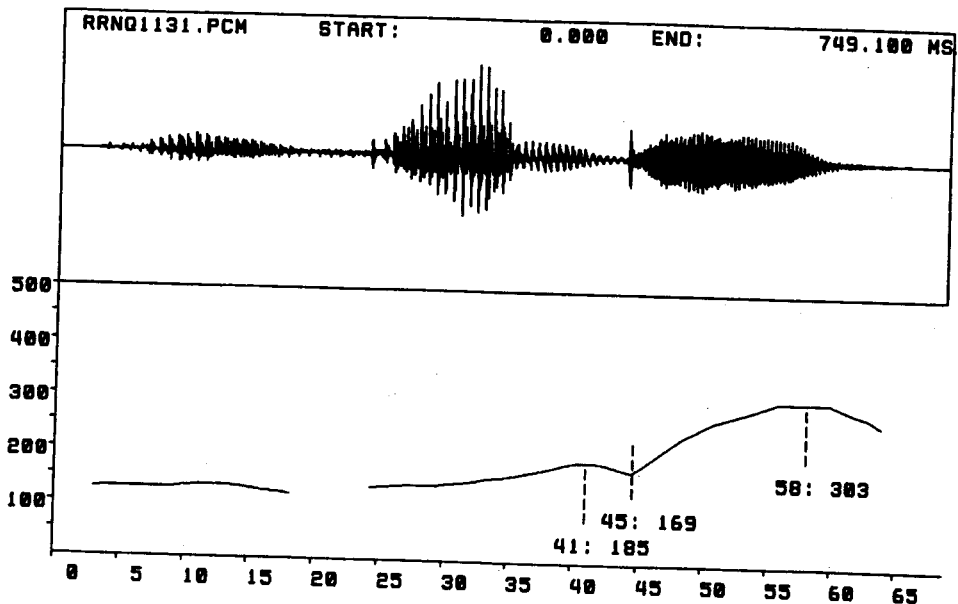


Figure 1. Fundamental-frequency contour for *November* indicating values at second-syllable peak  $F_0$  (frame 41: 185 Hz), third-syllable onset (frame 45: 169 Hz), and final-syllable peak (frame 58: 303 Hz).

at higher frequencies. The voicing threshold setting increases the likelihood that periodicity will be found throughout the token.)

The result of the acoustic analysis for each token was displayed in the form of an  $F_0$  contour, where each value in the contour represented an analysis frame corresponding to a 10-msec sampling window (Figure 1). The contours were inspected to ensure that the analysis provided plausible  $F_0$  values. When suspect values were encountered, the token was reanalyzed using  $F_0$  values obtained by counting pitch pulses in the waveform display as a guide to re-setting the ILS analysis parameters.

The following measurements were taken from the contour of each token: (1) second-syllable peak  $F_0$ , (2)  $F_0$  at third-syllable onset, and (3) third-syllable peak  $F_0$ . The amount of final rise was defined as the difference between  $F_0$  at the third-syllable onset and the third-syllable peak  $F_0$ . Syllable onset was determined by inspecting the  $F_0$  contour for a down-turn, which resulted from closure (for /b/ or /d/) at the boundary between the second and third syllables (see Figure 1). The frame containing the lowest point in the down-turn was taken as the onset of the third syllable. On the few occasions in which the  $F_0$  down-turn did not occur, the waveform was inspected and the average of the pitch periods over the first 10 msec of the third syllable, beginning with the first pitch period following the release of the /b/ or /d/, was used as the value of  $F_0$  at onset.

## Results

**Median peak  $F_0$  and final rise.** The distributions of peak  $F_0$  and final-rise values for each speaker in the sentence  $\times$  stress conditions were examined and typically found to be skewed or polymodal. Therefore, median peak  $F_0$  and final-rise values were determined for each of the 4 speakers in each condition. These values are presented in Table 1. Mean median peak  $F_0$  was higher for contrastive stress than for neutral stress in both utterances, but there seems to be no systematic effect of stress on the final rise.

Separate two-way (sentence  $\times$  stress) analyses of variance (ANOVAs) were performed on these median peak and final-rise values. As expected, median peak  $F_0$  was found to be significantly higher for questions with contrastive stress than for questions with neutral stress [210.7 Hz vs. 140.5 Hz;  $F(1,3) = 102.79, p < .001$ ]. This was consistent for all 4 speakers on both utterances. Median peak  $F_0$  did not differ between the utterances [172.4 vs. 186.2;  $F(1,3) = 0.09$ ], and the stress  $\times$  utterance interaction was not significant [ $F(1,3) = 0.79$ ].

Table 1  
Median Peak  $F_0$  and Final Rise

Speaker	<i>November</i>				<i>I did it</i>			
	Neutral		Contrastive		Neutral		Contrastive	
	Peak	Rise	Peak	Rise	Peak	Rise	Peak	Rise
L.R.	126.5	108.5	172.0	98.0	148.0	90.5	204.0	135.0
R.M.	159.5	78.5	278.5	73.5	129.5	72.0	159.0	188.0
J.S.	159.0	74.0	211.0	67.5	162.5	116.5	256.0	81.0
G.B.	123.0	93.5	204.2	111.0	116.0	81.0	202.0	143.5
Mean median	142.0	88.6	216.4	87.5	139.0	90.0	205.2	136.9

Note—Values are expressed in hertz.

**Table 2**  
Correlations Between Peak F0 and Final Rise for Two Utterances  
Produced with Neutral and Contrastive Stress

Speaker	November		I did it		Mean	
	Neutral	Contrastive	Neutral	Contrastive	Neutral	Contrastive
L.R.	.53	-.54	-.86	-.65	-.34	-.60
R.M.	.88	-.28	.69	-.93	.80	-.75
J.S.	-.36	-.88	.70	-.44	.39	-.73
G.B.	-.53	-.84	-.52	-.22	-.52	-.62
<i>M</i>	.41	-.70*	-.04	-.65*	.19	-.68†
<i>SE</i>	.40	.26	.54	.31	.40	.07

\* $p < .10$ , one-tailed. † $p < .001$ , one-tailed.

The analysis of final rise indicated no significant effects and no significant interactions.

**Peak F0-final rise relation.** Correlation coefficients (Pearson  $r$ ) between peak F0 and final rise were computed for the 4-5 tokens produced by each speaker in each condition and are shown in Table 2. In the neutral stress condition, mean correlations ranged from  $r = -.52$  to  $r = .80$ ; in the contrastive stress condition, all correlations were negative, ranging from  $r = -.60$  to  $r = -.75$ . Mean correlation coefficients across utterances for each speaker, and across speakers, were obtained by converting each speaker's correlation coefficient in each condition to  $z$ , using Fisher's transform

(Ferguson, 1981). The  $z$ , values were then averaged and reconverted to  $r$ s. Since each speaker produced only 4-5 tokens per sentence  $\times$  stress condition, individual speaker's mean correlations were not tested for significance. The significance of the deviation of the group mean  $r$ s from zero was tested for each utterance and for the average of the two utterances by  $t$  tests for samples of  $n = 4$ ,  $df = 3$ . Tests of significance in the contrastive stress condition were one-tailed, since the hypothesis was that these correlations would be negative. In the neutral stress condition, tests of significance were two-tailed.

Group mean correlation coefficients (again determined through Fisher's transformation) and the standard

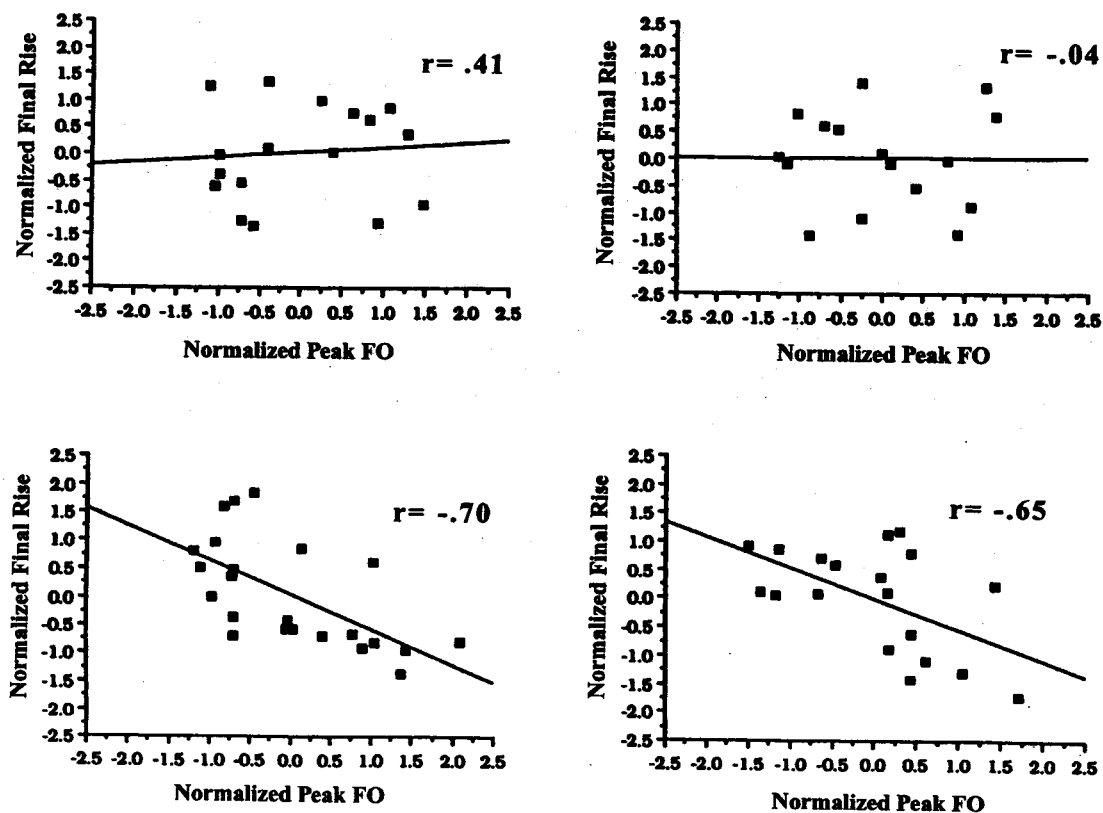


Figure 2. Relation between peak F0 and amount of final rise for *November* and *I did it* questions produced with neutral (top) and contrastive stress (bottom).

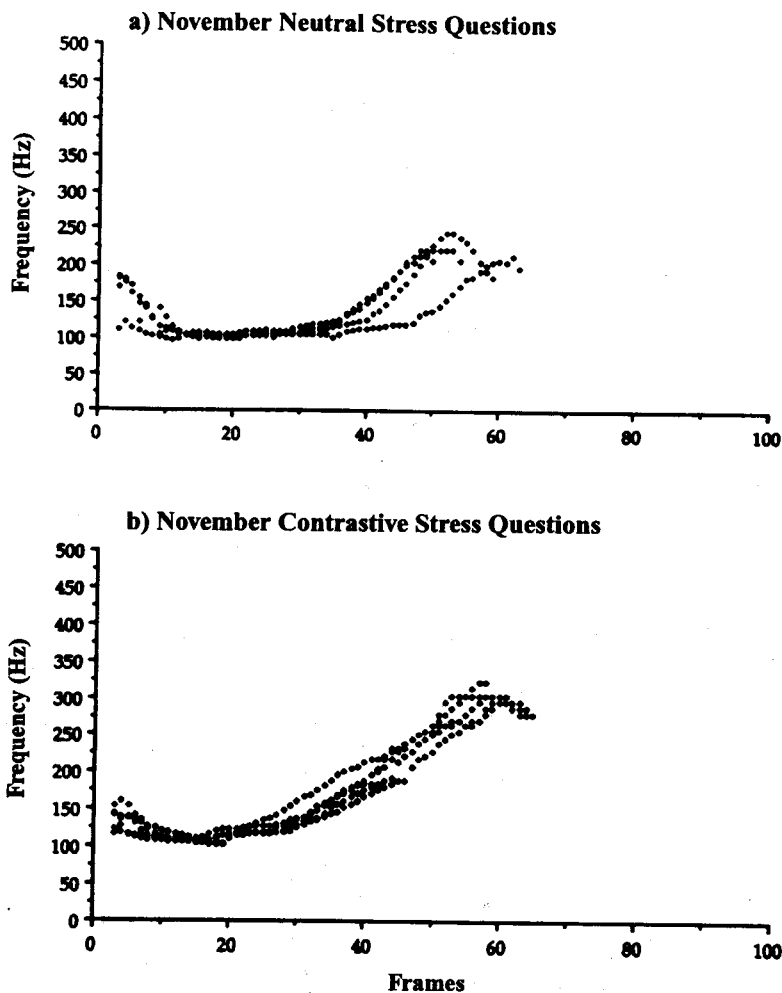


Figure 3. Continuously rising intonation contours produced by Speaker L.R. for the utterance *November*.

errors of the means are shown in Table 2. In the neutral stress condition, the mean correlation was near zero for *I did it* and positive for *November*. However, in the contrastive stress condition, significant negative correlations occurred for both utterances. The overall mean correlation (across utterances and speakers) was  $r = -.68$  ( $p < .001$ ) in the contrastive stress condition and  $r = .19$  (n.s.) in the neutral stress condition. The peak and final-rise values for each speaker's tokens in each sentence  $\times$  stress condition were converted to z-scores and are plotted in Figure 2 for *November* and *I did it* questions, respectively.

**Intonation contour shape.** The shape of the intonation contours used by the speakers was generally the continuously rising form typical of American English yes-no questions (e.g., Pike, 1945). All tokens of each speaker's neutral stress questions were of this form, as were most of their tokens of contrastive stress questions (see Figure 3). However, 2 speakers (J.S. and R.M.) used rise-fall-rise contours for some of their contrastive stress

questions. Speaker J.S. used this pattern for all of his contrastive stress tokens for both the *November* and the *I did it* utterance (see Figure 4); Speaker R.M. used this pattern for two contrastive stress tokens of *November*.

#### Discussion

The results of this experiment support the hypothesis that the trading relation observed by Studdert-Kennedy and Hadding (1973; Hadding-Koch & Studdert-Kennedy, 1965a, 1965b) in the perception of question intonation also occurs in production. A significant negative correlation was found between stressed-syllable peak  $F_0$  and the amount of final rise for questions produced with contrastive stress. The overall group mean correlation for questions with contrastive stress was  $r = -.68$ . The group mean correlations were of similar magnitude for each of the two utterances *November* and *I did it* ( $r = -.70$  and  $r = -.65$ , respectively), suggesting that the segmental makeup of the utterance is probably not a major determining factor. The occurrence of the trading

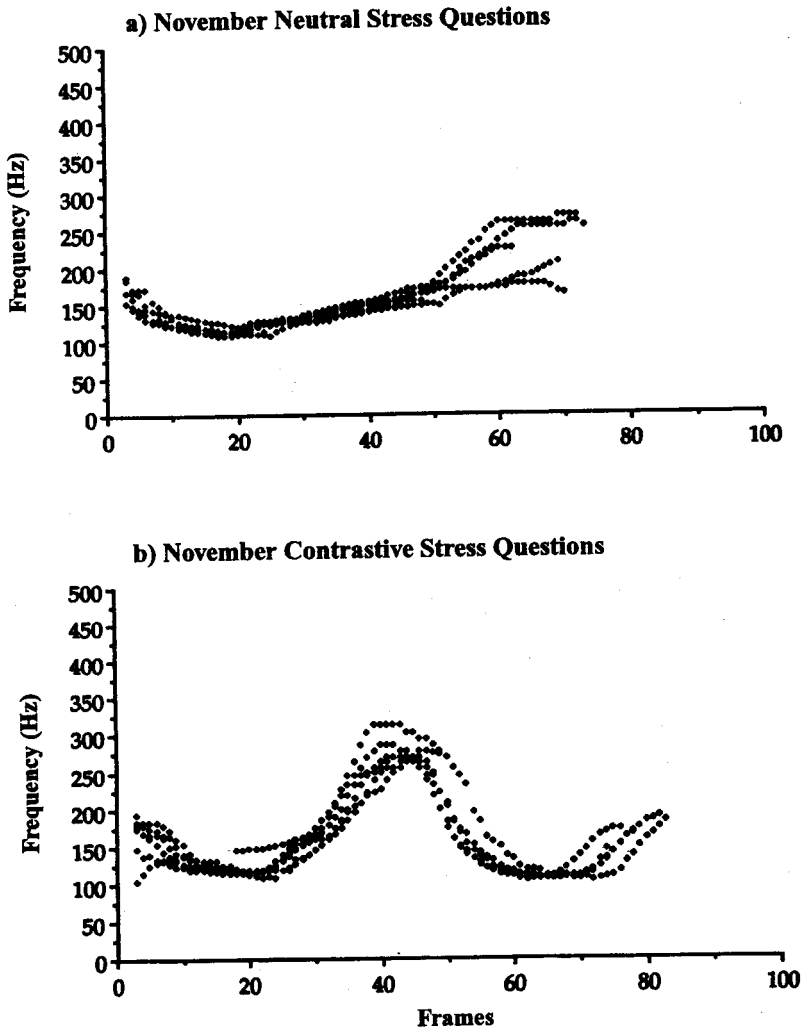


Figure 4. Continuously rising and rise-fall-rise intonation contours produced by Speaker J.S. for the utterance *November*.

relation in the contrastive stress condition was consistent across speakers: negative correlations were found for all 4 speakers.

For questions with neutral stress, the mean peak  $F_0$ -final rise correlation did not differ significantly from zero ( $r = .19$ ) overall or for either utterance individually. For these questions, individual speakers varied from a strong negative to a strong positive peak-final rise relation. This variability may have been due to the limited range of peak  $F_0$  values for questions without contrastive stress, or it may reflect the fact that the trading relation is related specifically to contrastive stress.

A potential concern with the results in the contrastive stress conditions is that, although most of the tokens in this condition had continuously rising contours, 2 speakers produced some tokens with rise-fall-rise contours. If the peak  $F_0$ -final rise relation differs for different contour shapes, combining data across these patterns could lead to invalid conclusions. Thus, for example, if the

amount of final rise for rise-fall-rise contours was generally less than that for continuously rising contours, then the finding of a negative peak  $F_0$ -final rise correlation could be spurious. This seems unlikely for two reasons. First, note that all the correlations in the contrastive stress condition for every subject were negative (see Table 2). Thus, it is not the case that the subjects who produced variant contour shapes were spuriously causing the negative correlation in the contrastive stress condition. Second, the correlations for the subjects who used the rise-fall-rise pattern were not consistently in a direction to influence the overall pattern of results. For example, Speaker J.S., who used the rise-fall-rise contour for all of his contrastive stress tokens, had peak-final rise correlations of  $r = -.88$  and  $r = -.44$ . Speaker R.M., who produced two tokens with rise-fall-rise contours for the *November* utterance, had a correlation of  $r = -.28$  for that utterance, but had a correlation of  $r = -.94$  in the *I did it* condition that included only

continuously rising contours. Thus, our finding of a negative correlation between peak  $F_0$  and final rise for yes-no questions with contrastive stress does not appear to be influenced by the presence of multiple contour shapes in the data.

The detailed pattern of results in Experiment 1 corresponds closely to that obtained by Studdert-Kennedy and Hadding (1973) for the perception of question intonation. For example, changes in peak  $F_0$  between 130 and 160 Hz had no consistent effect in perception, but changes from 160 to 200 Hz reliably reduced the amount of final rise that the listeners required to judge an utterance to be a question. In the present investigation, the mean median peak  $F_0$  in the neutral stress condition was approximately 140 Hz for each of the two utterances (*November* and *I did it*), and no trading relation was found. On the other hand, in the contrastive stress condition, the mean median peak  $F_0$  was greater than 200 Hz, and a trading relation was found. The close parallel between the present results in production and those of Studdert-Kennedy and Hadding in perception strongly suggests that the perceptual trading relation reflects listeners' attunement to patterns of speech production. This is consistent with the interpretation by Studdert-Kennedy and Hadding (1973) that their listeners perceived increases in peak  $F_0$  above 160 Hz as due to contrastive or emphatic stress, but it does not rule out the possibility that the trading relation occurs whenever peak  $F_0$  increases to a high value. Experiment 2 addressed this possibility by investigating whether the trading relation occurs when peak  $F_0$  increases to convey a difference in affect.

## EXPERIMENT 2

The goal of Experiment 2 was to replicate the trading relation found in Experiment 1 and explore its limiting conditions. Recall that in their perceptual studies, Studdert-Kennedy and Hadding found that the influence of peak  $F_0$  on the amount of final rise was limited to linguistic (question-statement) judgments. It was not found for judgments of the direction of pitch change (terminal rise or fall). Studdert-Kennedy and Hadding interpreted the dissociation between question-statement and rise-fall judgments with regard to the influence of peak  $F_0$  as evidence that the trading relation was a linguistic phenomenon, not a general psychoacoustic effect. Accordingly, in the present study, it was hypothesized that in production a trading relation would occur when peak  $F_0$  varied to convey a linguistic distinction, but not when it varied to convey an affective state. Trained actors were engaged to produce questions (and statements) with and without contrastive stress, as in Experiment 1. In addition, the same actors produced utterances without contrastive stress, but conveying positive and negative affects. It was predicted that the trading relation between peak  $F_0$  and amount of final rise in question intonation

would occur in the contrastive stress condition, but not in the affective conditions.

## Method

**Subjects.** Trained speakers were used as subjects. The speakers were 4 male actors, aged 23–28 years, recruited from the Yale Drama School. All were native speakers of American English from the northeastern and mid-Atlantic region. Each speaker was paid for his participation in a single 1½–2-h session.

**Test utterances.** The test utterances were those used in Experiment 1: *November* and *I did it*.

**Procedure.** A general description of the experiment was given to each speaker, followed by a 3–5-min calibration period (Cosmides, 1983), during which each speaker read a dramatic script from a science fiction novel (LeGuinn, 1968). The speakers were provided with written descriptions of the scenarios and given 10–15 min to study them and to prepare their productions. They were instructed to make four or five repetitions as nearly identical as possible for each scenario and to use facial gestures appropriate for each affect condition, since speaking with and without smiling has been shown to result in acoustic differences that listeners readily perceive (Tarter, 1980). Each speaker was seated in a sound booth approximately 24 in. from a microphone and a video camera. In an adjacent sound booth, the audio channel was recorded on a tape recorder, and the video channel was recorded on a VHS video recorder.

Each speaker produced the two sentences in a set of eight stress conditions and a set of eight affect conditions. The stress conditions resulted from crossing two degrees of stress (neutral and contrastive) with two sentence types (statement and question) and two listener-speaker distances (near and far). In the last condition, the speakers were instructed to speak first at an ordinary conversational level and then to raise the output level, as if to speak up over a short distance (e.g., across a table) or over moderate background noise; this condition was included to ensure that the range of peak  $F_0$  in the stress conditions was not unnaturally constrained. The affect conditions' result from crossing two affects (positive and negative) with two sentence types (question and statement) and two degrees of intensity (mild and moderate). The intensity conditions were again intended to ensure a full range of peak  $F_0$  values. As in Experiment 1, brief descriptive scenarios provided contextual support for each condition. Each speaker produced 4–5 repetitions in each condition for a total of approximately 128 tokens per subject and 512 total tokens, of which half were statements and half were questions. In the following analyses, only the 256 questions are considered and the data are collapsed across the near-far and mild-moderate dimensions of the design providing 8–10 trials for each subject in each condition.

The subjects completed all stress conditions before going on to the affect conditions. Within the stress conditions, all the *November* tokens were spoken first (neutral-contrastive statements, neutral-contrastive questions), followed by the *I did it* tokens in the same order of conditions. Within the affect conditions, positive affect tokens of both test sentences were spoken first in the order statements-questions, then negative affect tokens of both test sentences in the same order. After the stress conditions, and after the affect conditions, the speakers reviewed the tapes of their productions. If a speaker expressed dissatisfaction with any of his productions, an opportunity was provided to produce additional tokens.

**Acoustic analysis.** The data were treated as in Experiment 1. Thus, the following measurements were made on each token: (1) second-syllable peak  $F_0$ , (2)  $F_0$  at third-syllable onset, and (3) third-syllable peak  $F_0$ . The amount of final rise was defined as the difference between  $F_0$  at third-syllable onset and the third-syllable peak  $F_0$ . Third-syllable onset was determined as in Experiment 1.



## Results

**Median peak F0 and final rise.** As in Experiment 1, median peak F0 and final-rise values for each speaker's questions were determined and are shown for the stress conditions in Table 3. Mean median peak F0 was greater for tokens with contrastive stress than for those with neutral stress. This was the case for each speaker and each sentence, with one exception (Speaker P.N.'s median peak F0 is about equal for *November* tokens with neutral and contrastive stress). Mean median final rise was greater for tokens with contrastive stress than for those with neutral stress for the sentence *November*, but the reverse was true for the sentence *I did it*. Two-way ANOVAs—utterance (*November* vs. *I did it*)  $\times$  condition (neutral stress vs. contrastive stress or positive vs. negative)—were performed separately on the median peak F0 and final-rise data for the stress conditions and the affect conditions. For peak F0 in the stress conditions, the main effect of stress was significant [ $F(1,3) = 14.92, p < .05$ ], reflecting the fact that median peak F0 was higher in the contrastive stress condition than in the neutral stress condition. The main effect of utterance was not significant [ $F(1,3) = 4.98, p > .10$ ], nor was the condition  $\times$  utterance interaction [ $F(1,3) = 0.04$ ]. For final rise, a two-way ANOVA—utterance (*November* vs. *I did it*)  $\times$  condition (neutral stress vs. contrastive stress)—indicated no significant main effects and no significant interactions.

Median peak F0 and final-rise values for each speaker's questions in the affect conditions are shown in Table 4. Mean median peak F0 was greater for positive affect than for negative affect for the sentence *I did it*; however, the mean median peak F0 was slightly greater for negative affect tokens than for positive affect tokens for the sentence *November*. Mean median final rise was greater

for positive affect tokens than for negative affect tokens for both utterances, a result that is seen for all but 1 speaker (Speaker R.R. produced slightly greater final rise for negative affect tokens than for positive affect tokens of *I did it*). A two-way ANOVA (utterance  $\times$  affect) performed on median peak F0 indicated no significant effects and no significant interactions. For final rise, the ANOVA indicated a main effect for utterance [ $F(1,3) = 13.29, p < .05$ ], reflecting more final rise for *I did it* questions than for *November* questions. No other main effects or interactions were significant.

**Peak F0-final rise relation.** Correlations between peak F0 and final rise for the 8–10 tokens produced by each speaker in each of the four stress conditions and each of the four affect conditions are shown in Tables 5 and 6, respectively. Mean correlation coefficients were computed across conditions within speaker and across speakers using Fisher's transform. The significance of the deviation of the group mean  $r$  from zero was tested on samples of  $n = 4, df = 3$ . As in Experiment 1, one-tailed significance tests were performed in the contrastive stress condition, and two-tailed tests were performed in all other conditions.

For individual speakers, mean correlations across the test sentences in the contrastive stress condition ranged from  $r = -.42$  to  $r = -.78$ , and the correlations in the neutral stress condition ranged from  $r = -.11$  to  $r = .35$  (see Table 5). In the affect conditions, mean correlations for individual speakers averaged across test sentences ranged from  $r = -.36$  to  $r = .83$  for the two affects separately (see Table 6). Overall mean correlations for individual speakers averaged across the test sentences and affects ranged from  $r = -.27$  to  $r = .55$ . In summary, every speaker demonstrated negative mean correlations in the contrastive stress conditions, whereas

Table 3  
Median Peak F0 and Final Rise for Stress Conditions

Speaker	November				I did it			
	Neutral		Contrastive		Neutral		Contrastive	
	Peak	Rise	Peak	Rise	Peak	Rise	Peak	Rise
E.O.	104.5	60.0	153.0	40.5	131.0	48.0	146.5	45.0
P.N.	124.5	62.5	123.5	69.0	118.5	112.5	136.0	48.5
R.R.	133.5	75.7	164.0	149.9	162.0	112.0	173.5	101.5
S.B.	110.0	95.0	135.5	52.0	101.0	48.0	149.0	96.0
Mean median	118.1	73.3	144.0	77.8	128.1	80.1	152.2	72.7

Note—Values are expressed in hertz.

Table 4  
Median Peak F0 and Final Rise for Affect Conditions

Speaker	November				I did it			
	Positive		Negative		Positive		Negative	
	Peak	Rise	Peak	Rise	Peak	Rise	Peak	Rise
E.O.	129.5	48.5	137.5	12.5	138.5	45.5	134.0	26.0
P.N.	260.5	72.0	256.5	65.0	219.5	140.0	133.5	23.5
R.R.	200.0	77.0	192.0	106.0	263.0	94.0	186.0	97.0
S.B.	139.0	53.5	171.5	48.0	137.0	98.0	169.5	57.5
Mean median	182.2	62.7	189.4	57.9	189.5	94.4	155.7	51.0

Note—Values are expressed in hertz.

Table 5  
Correlations Between Peak *F0* and Final Rise for Each Speaker in Each Stress Condition

Speaker	November		<i>I did it</i>		Mean	
	Neutral	Contrastive	Neutral	Contrastive	Neutral	Contrastive
E.O.	.18	-.65	.05	-.78	.11	-.72
P.N.	-.20	-.51	-.03	-.33	-.11	-.42
R.R.	.25	-.84	.27	-.47	.26	-.70
S.B.	.69	-.94	-.12	-.33	.35	-.78
<i>M</i>	.26	-.79*	.04	-.51*	.15	-.67†
<i>SE</i>	.22	.26	.08	.17	.10	.13

\* $p < .05$ , one-tailed. † $p < .01$ , one-tailed.

correlations in the neutral stress and affect conditions ranged over positive and negative values.

The group mean correlations between peak *F0* and final rise in the stress conditions and their standard errors appear in Table 5. In the neutral stress condition, the overall mean correlation was  $r = .15$ , which did not differ significantly from zero. For the two utterances separately, both correlations were positive, but neither was significant. However, in the contrastive stress condition, the overall mean correlation was  $r = -.67$  ( $p < .001$ ), and significant negative correlations were found for each test sentence.

Table 6 presents the corresponding correlational data for the affect conditions. The mean peak *F0*-final rise correlation across speakers, test sentences, and affects was  $r = .34$ , which is not significantly different from zero. Mean correlations across the two affect conditions for each of the test sentences were  $r = .44$  (n.s.) and  $r = .05$  (n.s.) for *November* and *I did it*, respectively. Mean correlations across test sentences for the positive and negative affect conditions were  $r = -.11$  (n.s.) and  $r = .52$  (n.s.), respectively. For the two test sentences individually, mean correlations were  $r = .33$  and  $r = .55$  for *November* and  $r = -.57$  and  $r = .50$  (all n.s.) for *I did it* in positive and negative affect conditions, respectively. Thus, across subjects in the affect conditions, there were no correlations that differed significantly from zero.

The peak *F0* and final-rise values for each speaker in each sentence  $\times$  stress and sentence  $\times$  affect condition were converted to *z* scores and are plotted in Figures 5 and 6 for the stress and affect conditions, respectively.

**Intonation contour shapes.** All tokens included from all 4 speakers in both the neutral stress and the contrastive stress conditions had the continuously rising pattern typical of American English yes-no questions.

In the affect conditions, all of the positive affect contours and most of the negative affect contours were also of the continuously rising form. In the negative affect condition, some tokens produced by Speakers E.O. and S.B. had contours that might best be described as rise-flat, or rise-fall-flat, whereas some tokens produced by Speaker P.N. had the rise-fall-rise contour for the utterance *November*.

### Discussion

The purpose of Experiment 2 was to confirm the findings of Experiment 1 and to investigate whether the trading relation would also occur when peak *F0* varied to convey affect. To that end, the experiment included the same test sentences and the same neutral and contrastive stress conditions as those in Experiment 1. In addition, conditions were added in which speakers conveyed positive and negative affect.

The results of Experiment 2 replicate the findings of Experiment 1. The trading relation between peak *F0* and final rise occurred among questions with contrastive stress, where significant negative mean correlations were found between peak *F0* and final rise overall ( $r = -.67$ ,  $p < .001$ ) and for each test sentence ( $r = -.79$ ,  $p < .05$ ;  $r = -.51$ ,  $p < .05$ ). However, among questions without contrastive stress, the overall correlation was nonsignificant ( $r = .15$ ).

In the affect conditions, no evidence of a trading relation was found. Rather, the peak *F0*-final rise relation among questions conveying affect was characterized by an overall mean correlation that did not differ significantly from zero ( $r = .34$ ). Since the trading relation did not occur in the affect conditions despite a higher mean peak *F0* among these affectively toned questions than among questions with contrastive stress, it may be seen that the effect is not a result of raising peak *F0* generally.

Table 6  
Correlations Between Peak *F0* and Final Rise for Each Speaker in the Affect Conditions

Speaker	November		<i>I did it</i>		Mean		Overall Mean
	Positive	Negative	Positive	Negative	Positive	Negative	
E.O.	-.23	-.34	-.41	-.10	-.32	-.23	-.27
P.N.	.37	.83	-.25	.83	.06	.83	.55
R.R.	.72	.88	-.93	.66	-.36	.79	.34
S.B.	.31	.25	.07	.30	.19	.28	.24
<i>M</i>	.33	.55	-.57	.50	-.11	.52	.34
<i>SE</i>	.23	.41	.37	.28	.14	.34	.09

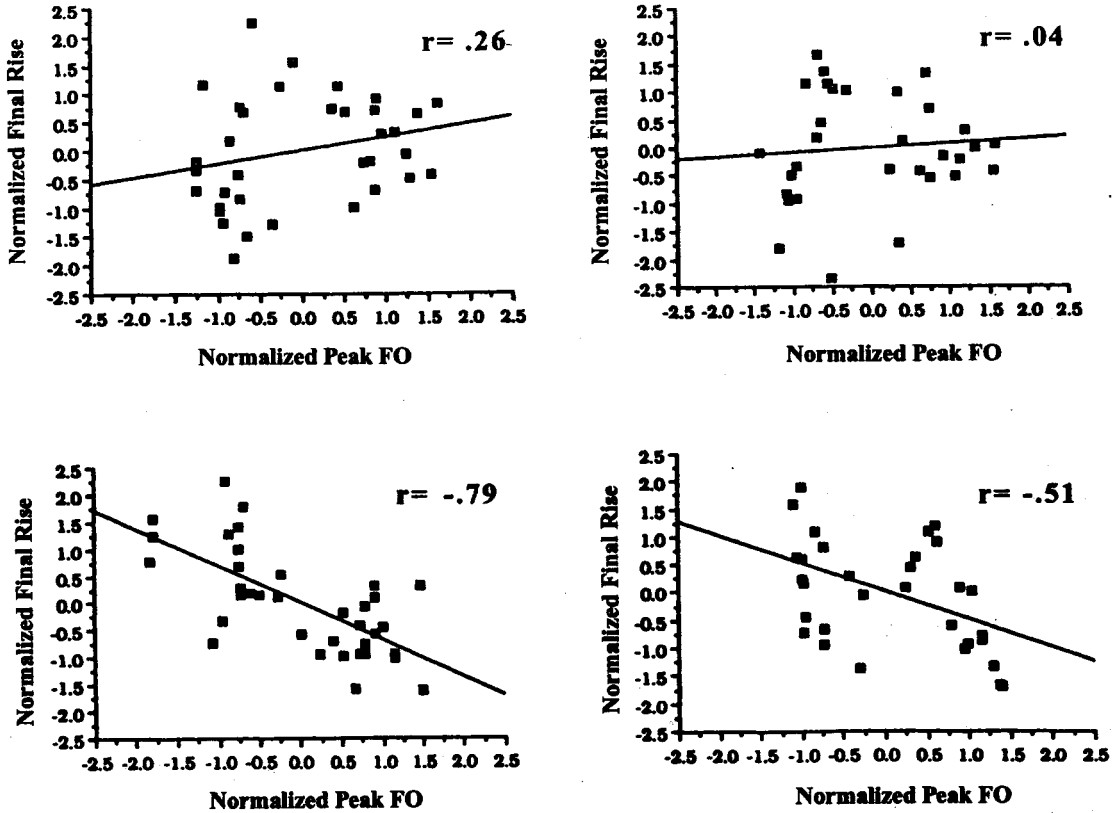


Figure 5. Relation between peak  $F_0$  and amount of final rise for *November* and *I did it* questions produced with neutral (top) and contrastive stress (bottom).

Instead, the trading relation appears to be specific to the use of  $F_0$  to convey contrastive stress.

As in Experiment 1, these results are clear in their indications. Each speaker in the contrastive stress condition demonstrated negative correlations for both utterances ( $r = -.33$  to  $r = -.94$ ), whereas speakers in the neutral stress condition demonstrated correlations that varied from small negative to moderate positive ( $r = -.20$  to  $r = .69$ ). Similarly, in the affect condition, speakers' overall correlations ranged from small negative to moderate positive ( $r = -.27$  to  $r = .55$ ). Considering the two emotions separately for each speaker, five out of eight mean correlations were positive.

These results demonstrate a dissociation in the trading relation between linguistically driven and affectively driven variations of  $F_0$ . That is, when  $F_0$  is increased to convey contrastive stress, the amount of final rise used to signal a question decreases. However, when  $F_0$  increases to convey affect, the amount of final rise tends to increase. This dissociation supports the hypothesis that the trading relation is specifically a linguistic phenomenon. This specificity raises the possibility that the articulatory actions used to raise  $F_0$  for contrastive stress may differ from the actions that raise  $F_0$  to convey positive and negative affects. In short, elevation of  $F_0$  for contrastive stress appears to interfere with the

production of the final-syllable rise in  $F_0$ , whereas elevation of  $F_0$  to convey affect does not impede production of final rise.

### EXPERIMENT 3

The goal of Experiment 3 was to validate the results of Experiment 2 by ascertaining that naive listeners could apprehend the speakers' intent. Therefore, in Experiment 3, listeners were asked to rate each token from Experiment 2 as either a question or a statement. The affective polarity and degree of affect of each token were also rated. Of special interest was whether the differential occurrence of the trading relation would be maintained in those specific tokens for which listeners correctly apprehended the speakers' intentions. The subset of tokens for which the majority of listeners apprehended the speakers' intent with regard to question-statement and affect was examined for the relation between peak  $F_0$  and final rise.

#### Method

**Subjects.** Eighteen adult native English-speaking subjects participated in Experiment 3. The subjects were graduate and undergraduate students from the University of Connecticut. All subjects indicated they had normal hearing.

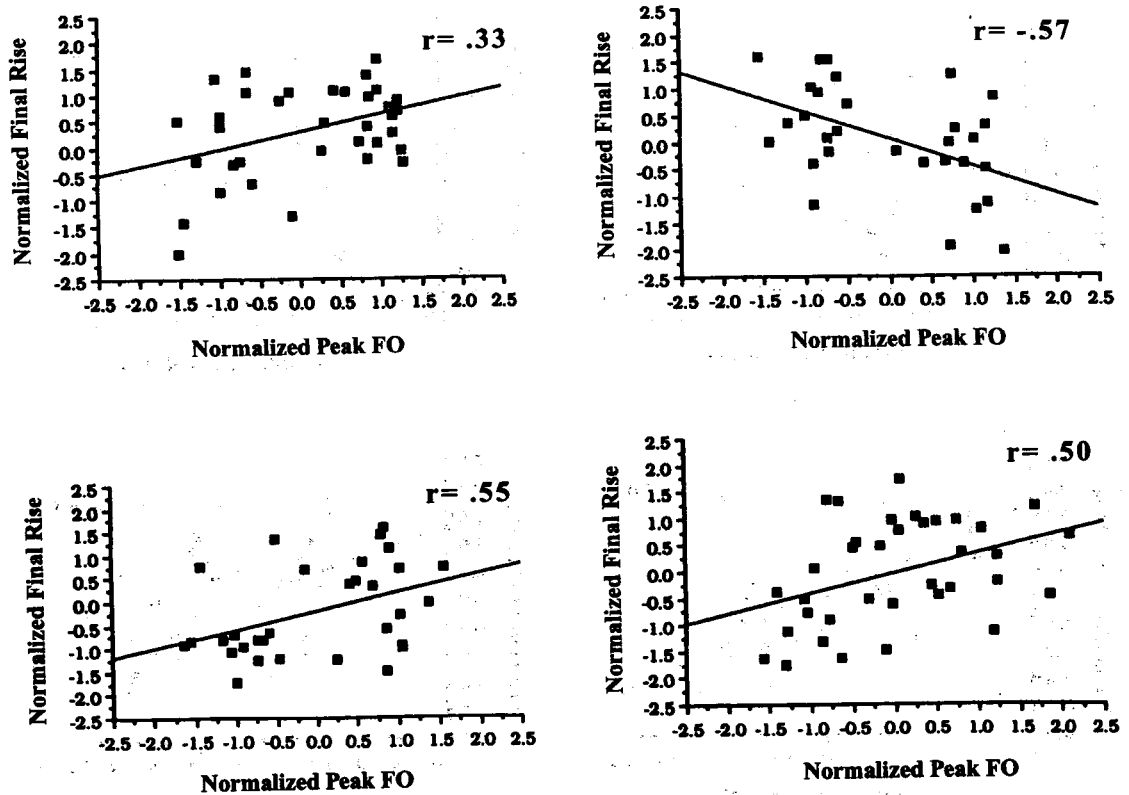


Figure 6. Relation between peak  $F_0$  and amount of final rise for *November* and *I did it* questions produced with positive (top) and negative affect (bottom).

**Stimuli.** The stimuli were all of the *November* questions and statements produced by each of the 4 speakers in Experiment 2. Each speaker's productions were randomized separately and arranged in blocks of 10 tokens with 7-sec interstimulus intervals and 15-sec interblock intervals.

**Procedure.** The subjects were tested individually in a sound-attenuated room. Each subject heard the productions of 2 speakers in separate tests with a 5-min break between tests. Before each test, the subjects were instructed that they would hear a single speaker saying the word *November* as either a question or a statement and in various tones of voice. Each token was rated as a question or a statement and then rated for the polarity and degree of affect on an 11-point scale ( $-5$  to  $+5$ ), where  $-5$  indicated *very negative* and  $+5$  indicated *very positive*. Each token was presented once without feedback. Stimuli were presented over headphones at a comfortable volume level. Each session lasted approximately 40 min.

## Results

Those tokens correctly labeled by the majority of listeners as questions and having an average rating consistent with the speaker's intended affect were selected for further consideration. The criterion for inclusion of tokens rated on the affective scale was an average rating of  $+1.5$  to  $+5$  for positive affect,  $-1.5$  to  $-5$  for negative affect, and  $+1.0$  to  $-1.0$  for tokens from the nonaffective stress conditions. In general, the listeners' ratings were in good agreement with the speakers' intent. Question-statement judgments were in very high agreement with

the speakers' intent, with 98% of the questions and 100% of the statements rated appropriately by the majority of listeners. In addition, 50% of the questions from the neutral and contrastive stress conditions, 62% of the positive affect questions, and 75% of the negative affect questions met the affective rating criteria for inclusion in the following analysis.

The mean correlations between peak  $F_0$  and final rise for the questions that met the inclusion criteria are shown in Table 7, along with the correlations for the total number of tokens from Experiment 2. Only in the contrastive stress condition was the correlation negative ( $r = -.59$ ); the correlation in both the positive and the negative affective conditions was positive ( $r = .26$  and  $r = .43$ , respectively). Thus, the differential relation between peak  $F_0$  and final rise for linguistic versus affective distinctions was maintained for this selected subset relative to the values obtained in Experiment 2, where no selection criteria were imposed.

## Discussion

The results of Experiment 3 show that the listeners were generally able to apprehend simultaneously the affective and linguistic intentions of speakers, despite the fact that the utterances were very brief (consisting of only three syllables). With respect to the trading relation between peak  $F_0$  and final rise in questions, the listener

judgments in this experiment provide further evidence supporting Studdert-Kennedy and Hadding's hypothesis that the trading relation is specifically linguistic in nature. The dissociation of the trading relation between questions conveying linguistic and affective contrasts found in Experiment 2 is confirmed when only those tokens are considered for which listeners apprehended the speakers' intent.

### GENERAL DISCUSSION

The two major hypotheses of this study were supported by the results. The prediction of a trading relation between peak  $F_0$  and the amount of final rise in questions conveying contrastive stress was confirmed by the results of both Experiments 1 and 2. In addition, the differential occurrence of the trading relation demonstrated in Experiment 2 and upheld by listeners in Experiment 3 suggests that increases in  $F_0$  associated with contrastive stress are produced differently from the increases used to convey affect.

Studies of trading relations in speech have hitherto been solely concerned with perception (e.g., Bailey, Summerfield, & Dorman, 1977; Best, Morriongiello, & Robson, 1981; Fitch, Halwes, Erickson, & Liberman, 1979; Summerfield, 1975). Although many investigators have tended to regard trading relations as a reflection of perceptual attunement to the dynamics of articulation (e.g., Repp, 1982), there has been little attempt to test this notion or develop its implications, so that the relation between perception and production has remained a matter of speculation.

Investigation of the perception-production link is important because perceptual trading relations are particular to synthetically produced or altered speech, in which acoustic variables are manipulated without regard to the constraints of natural speech production. The perceptual studies by Studdert-Kennedy and Hadding (1973; Hadding-Koch & Studdert-Kennedy, 1964) used hybrid stimuli formed by imposing synthetic intonation contours on a single natural utterance. Although the range of  $F_0$  values used to construct the contours was based on spectrographic data from natural productions (Hadding-Koch, 1961), the combination of  $F_0$  values in any particular stimulus was arbitrary, and all the stimuli had the same duration, to which a stylized contour was fitted. As a result, much of the variability in frequency spectrum, amplitude envelope, duration, and  $F_0$  of naturally produced speech was missing. By contrast, the present study analyzed utterances constrained only by the scenarios

devised to elicit questions with appropriate patterns of stress and affective inflections. Given the difference in experimental conditions, the closeness with which the results of the present production study match those of the perceptual studies attests to the functional reality of the trading relation between peak  $F_0$  and final rise. In demonstrating the trading relation in production, these results give evidence that perceptual trading relations can indeed reflect listeners' sensitivity to patterns of speech production. Moreover, this study provides specific information about the organization of  $F_0$  in speech by establishing the set of circumstances under which production patterns result in a trading relation.

Two types of explanation for these findings are possible. On the one hand, both the trading relation and its limitation to a context of contrastive stress may be linguistic conventions. That is, they may be arbitrary associations, established by members of a linguistic community and learned as rules by speakers as they acquire the language (in this case, English and, presumably, Swedish). This explanation is less than satisfying, however, because it fails to address why or how the trading relation originated. Its presence in both languages means either that it evolved in each language separately or that English and Swedish share the trading relation as a result of their distant common ancestry in Germanic. Furthermore, this explanation requires that speakers acquire a rule for decreasing the final rise as peak  $F_0$  increases, only when  $F_0$  peaks increase to convey contrastive stress, not when they are used to convey affect. While this is not impossible, it is difficult to motivate.

On the other hand, the trading relation may reflect the articulatory dynamics of raising  $F_0$ . Two articulatory hypotheses will be considered. Both assume that the production of the final rise is accomplished by the action of laryngeal muscles, primarily cricothyroid (CT), which control the tension of the vocal folds and thus  $F_0$  (Atkinson, 1973; Lieberman, Sawashima, Harris, & Gay, 1970). They differ on the mechanism that underlies the production of the  $F_0$  rise associated with stress. Lieberman (1967) proposed that the production of questions and statements in English involves the distinction between marked and unmarked breath groups. In the unmarked breath group used for statements,  $F_0$  is strictly a function of subglottal pressure ( $P_s$ ).  $P_s$  and  $F_0$  are typically highest at the beginning of an utterance; as  $P_s$  is used up over the course of the utterance to maintain voicing,  $P_s$  and  $F_0$  decrease over the duration of the breath group. In the marked breath group used for yes-no questions, a rise in  $F_0$  at the end of the breath

Table 7  
Mean Correlations Between Peak  $F_0$  and Final Rise  
for November Utterances Judged as Questions

Experiment	Condition		Affect		<i>M</i>
	Neutral Stress	Contrastive Stress	Positive	Negative	
2	.15	-.67	-.11	.52	.34
3	.25	-.59	.26	.43	.35

group is effected despite falling  $P_s$  by an increase in laryngeal tension. Lieberman further proposed that a speaker effects any  $F_0$  rise in the middle of an utterance by increasing the flow of air through the glottis, thus using up some of the finite subglottal-pressure reserve. In the case of a yes-no question, this loss of  $P_s$  results in a final rise that is reduced relative to what it would have been without the (previous) stress.

This hypothesis has some advantages over the "convention" explanation. For example, since it proposes an explanation based on an articulatory constraint, it does not require speakers to acquire a complex rule for when and how much to reduce the final rise. However, this proposal also has several problems. First, stress-related  $F_0$  rises are not produced solely by using  $P_s$ , as Lieberman originally proposed. Rather, some combination of  $P_s$  and CT activity are involved (Atkinson, 1973; Gelfer, 1987).<sup>2</sup> Second, even if it was the case that stress production involves primarily  $P_s$ , it is unclear that the effect Lieberman suggests would occur on utterances as short as three syllables (such as those used in this study). If it did, it would suggest a gross lack of planning in speech production: how would speakers ever get to the end of longer utterances with multiple stresses? Finally, it is not clear how Lieberman's hypothesis would account for the dissociation of the trading relation seen in Experiment 2.

An alternative articulatory account of the trading relation is based on EMG studies of speech production that indicate that, in addition to  $P_s$ , CT activity is involved in the  $F_0$  rise associated with stress (Atkinson, 1973; Gelfer, 1987), as well as in the final  $F_0$  rise signaling questions (Atkinson, 1973; Hirano, Ohala, & Vennard, 1969; Lieberman et al., 1970). Perhaps the need for CT activity on two adjacent syllables underlies the reduction in final rise following a syllable with contrastive stress. For example, on the basis of studies of single-unit motor activity in humans, it is known that the peak effect of CT contraction on  $F_0$  has a latency of 70–80 msec (Baer, 1978, 1981). Furthermore, the relaxation time of CT is longer than its contraction time (Löfqvist, Baer, McGarr, & Story, 1989). Thus, CT contractions on adjacent syllables required to produce the continuously rising question contours seen in Experiments 1 and 2 would very likely overlap in time, with the initiation of the second contraction occurring before the muscle had completely relaxed from the first contraction. This would reduce the force of the second contraction, resulting in a reduction in the amount of  $F_0$  rise. The lack of a trading relation for the affectively inflected questions suggests that the rise in  $F_0$  on the second syllable of these utterances may be produced by a mechanism different from that used for affectively neutral contrastive stress. Perhaps the most likely mechanism (especially in the case of positive affects, such as happiness) is an overall increase in  $P_s$  across the utterance, sufficient to override the effect of overlapping CT contractions.

The mechanism proposed above to account for the trading relation in intonation suggests that the produc-

tion of intonational units may share certain features with the production of segmental units of speech. Because speech production is dynamic, upper vocal tract articulatory gestures for phonetic segments often overlap temporally with the gestures for preceding or following segments. One result of this coarticulation is context-dependent variation of articulatory trajectories and, ultimately, of the acoustic output. Thus, for example, the direction of second and third formant transitions for stop consonants may rise or fall depending on the relevant formant frequencies of the preceding or following vowel (Ohman, 1965). This is due to a physical (physiological) constraint on the trajectory of the tongue in moving between the positions required to produce the vowels and consonants. Ohman concluded that the neural commands for consonants and vowels in VCV sequences must be active simultaneously. The suggestion we mean to make is that an analogous phenomenon may occur in the case of intonation. That is, in the context of a preceding  $F_0$  prominence associated with stress, the production of final rise is altered because of a physical (physiological) constraint on the ability of the CT muscle to produce temporally overlapping gestures that raise  $F_0$ . The result is a reduction in the magnitude of the second gesture and, therefore, a reduction in the final rise.

This proposal has several advantages over both linguistic convention and Lieberman's  $P_s$  hypothesis as an explanation for the data presented in this paper. As with Lieberman's hypothesis, this proposal reduces the acquisition problem to one of discovering articulatory dynamics, for which infants seem to be quite well suited. It also eliminates the need to posit a scenario regarding how the trading relation and its specific use arose in languages as diverse as English and Swedish; presumably, these languages use the same mechanisms to raise  $F_0$  for stress and question final rises. This hypothesis has advantages over Lieberman's  $P_s$  hypothesis, not the least of which is that it is consistent with current data on physiological mechanisms underlying the production of stress. In addition, it offers a plausible and parsimonious explanation for both the trading relation and its specific occurrence. It may also be related to other phenomena, such as the tendency to avoid equally stressed adjacent syllables<sup>3</sup> and the use of rule-based substitutions in tone languages.

Although the present experiments were not specifically designed to test the question, the results are consistent with other data suggesting that speech  $F_0$  may be controlled differently for linguistic and affective purposes. For example, a variety of evidence suggests that the two hemispheres of the brain have somewhat different roles in the control of affective behavior both in humans and in other mammalian species (Denenberg, 1981). In humans, dichotic studies with normal subjects have shown a left-ear (right-hemisphere) advantage for affective prosody (Ley & Bryden, 1982; Shipley-Brown et al., 1988; Zurif, 1974), whereas a right-ear (left-hemisphere) advantage is usually observed for segmental judgments. Perceptual

disorders involving emotional prosody following right-hemisphere brain injury have also been demonstrated, but deficits of linguistic prosody apparently occur following injury to either hemisphere (e.g., Blumstein & Cooper, 1972; Heilman et al., 1984; Tucker et al., 1977; Weintraub et al., 1981). Some studies of the production of linguistic intonation have found that patients with right-hemisphere damage produce normal linguistic stress and sentence intonation (Behrens, 1988; Emmorey, 1987), but others have reported deficits after right-hemisphere damage (Shapiro & Danly, 1985; Weintraub et al., 1981). Although far from unanimous in their indications, these studies point to a different degree of involvement of the left and right hemispheres in linguistic and affective prosody, including intonation. The results of the present study are consistent with the hypothesis that the control of linguistic intonation is functionally separable from the affective use of F0.

## REFERENCES

- ATKINSON, J. E. (1973). *Aspects of intonation in speech: Implications from an experimental study of fundamental frequency*. Unpublished doctoral dissertation, University of Connecticut.
- BAER, T. (1978). The effect of single-motor-unit firings on fundamental frequency. *Journal of the Acoustical Society of America*, **64**, S90(A).
- BAER, T. (1981). Investigation of the phonatory mechanism. In *ASHA Reports: Vol. 11. Proceedings of the Conference on the Assessment of Vocal Fold Pathology*. Hart, MO: Ludlow.
- BAILEY, P., SUMMERFIELD, Q., & DORMAN, M. (1977). *On the identification of sine-wave analogues of certain speech sounds* (Status Report No. SR 51/52, pp. 1-25). New Haven, CT: Haskins Laboratories.
- BEHRENS, S. (1988). The role of the right hemisphere in the production of linguistic stress. *Brain & Language*, **33**, 104-127.
- BEST, C., MORRONGIELLO, B., & ROBSON, R. (1981). The perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, **29**, 191-211.
- BLUMSTEIN, S., & COOPER, W. (1972). Hemispheric processing of intonation contours. *Cortex*, **10**, 146-158.
- BLUMSTEIN, S., & LIEBERMAN, P. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge: Cambridge University Press.
- BOLINGER, D. L. (1958). A theory of pitch accent in English. *Word*, **14**, 109-149.
- BOLINGER, D. L. (1978). Intonation across languages. In J. P. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language: Volume 2. Phonology* (pp. 471-522). Stanford, CA: Stanford University Press.
- COHEN, A. (1972). Some observations on the pitch of questions. In A. Valdman (Ed.), *Papers in memory of Pierre Delattre* (pp. 97-103). London: Mouton.
- COHEN, A., COLLIER, R., & T'HART, J. (1982). Declination: Construct or intrinsic feature of speech pitch? *Phonetica*, **39**, 254-273.
- COOPER, W., & SORENSON, J. (1981). *Fundamental frequency in sentence production*. New York: Springer-Verlag.
- COSMIDES, L. (1983). Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 864-881.
- CRYSTAL, D. (1969). *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.
- DAVITZ, J. R. (1964). A review of research concerned with facial and vocal expressions of emotion. In J. R. Davitz (Ed.), *The communication of emotional meaning* (pp. 13-31). New York: McGraw-Hill.
- DENENBERG, V. (1981). Hemispheric laterality in animals and the effects of early experience. *Behavioral & Brain Sciences*, **4**, 1-49.
- EMMOREY, K. (1987). The neurological substrates for prosodic aspects of speech. *Brain & Language*, **30**, 305-320.
- FAIRBANKS, G., & HOAGLIN, L. (1941). An experimental study of the duration characteristics of the voice during the expression of emotions. *Speech Monographs*, **8**, 85-90.
- FAIRBANKS, G., & PRONOVOST, W. (1939). An experimental study of the pitch characteristics of the voice during the expression of emotions. *Speech Monographs*, **6**, 87-104.
- FERGUSON, G. (1981). *Statistical analysis in psychology and education*. New York: McGraw-Hill.
- FITCH, H., HALWES, T., ERICKSON, D., & LIBERMAN, A. M. (1979). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics*, **27**, 343-350.
- FRIES, C. C. (1964). On the intonation of yes-no questions in English. In D. Abercrombie (Ed.), *In honour of Daniel Jones* (pp. 242-254). London: Longmans.
- FRY, D. B. (1955). Duration and intensity as acoustic correlates of linguistic stress. *Journal of the Acoustical Society of America*, **35**, 765-769.
- FRY, D. B. (1958). Experiments in the perception of stress. *Language & Speech*, **1**, 126-152.
- GELFER, C. (1987). *A simultaneous physiological and acoustic study of fundamental frequency*. Unpublished doctoral dissertation, City University of New York.
- HADDING-KOCH, K. (1961). *Acoustico-phonetic studies of intonation in southern Swedish*. Lund, Sweden: Gleerups.
- HADDING-KOCH, K., & STUDDERT-KENNEDY, M. (1964). An experimental study of some intonation contours. *Phonetica*, **11**, 175-185.
- HADDING-KOCH, K., & STUDDERT-KENNEDY, M. (1965a). Intonation contours evaluated by American and Swedish test subjects. In W. Bethge (Ed.), *Proceedings of the Vth International Congress of Phonetic Sciences* (pp. 326-331). Basel, Switzerland: S. Karger.
- HADDING-KOCH, K., & STUDDERT-KENNEDY, M. (1965b). A study of semantic and psychophysical test responses to controlled variations in fundamental frequency. *Studia Linguistica*, **17**, 65-76.
- HEILMAN, K., BOWERS, D., SPEEDIE, L., & COSLETT, H. (1984). Comprehension of affective and non-affective prosody. *Neurology*, **34**, 917-921.
- HIRANO, M., OHALA, J., & VENNARD, W. (1969). The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation. *Journal of Speech & Hearing Research*, **12**, 616-628.
- JASSEME, W. (1972). The question-phrase fall-rise in British English. In A. Valdman (Ed.), *Papers in memory of Pierre Delattre* (pp. 241-252). Paris: Mouton.
- KRAMER, E. (1964). Elimination of verbal cues in judgments of emotion from voice. *Journal of Abnormal & Social Psychology*, **68**, 390-396.
- LEE, W. R. (1956). English intonation: A new approach. *Lingua*, **4**, 361.
- LEE, W. R. (1960). *An English intonation reader*. London: Macmillan.
- LEE, W. R. (1980). A point about the rise-endings and fall-endings of yes-no questions. In L. R. Waugh & C. H. van Schooneveld (Eds.), *The melody of language intonation and prosody* (pp. 165-168). Baltimore: University Park Press.
- LEGUINN, U. (1968). *A wizard of earthsea*. Boston: Houghton Mifflin.
- LEY, R., & BRYDEN, M. (1982). A dissociation of right and left hemisphere effects for recognizing emotional tone and verbal content. *Brain & Cognition*, **1**, 3-9.
- LIEBERMAN, P. (1967). *Intonation, perception and language*. Cambridge, MA: MIT Press.
- LIEBERMAN, P., SAWASHIMA, M., HARRIS, K., & GAY, T. (1970). The articulatory implementation of the breath-group and prominence: Cricothyroid muscular activity in intonation. *Language*, **46**, 312-327.
- LÖFQVIST, A., BAER, T., MCGARR, N., & STORY, R. (1989). The cricothyroid muscle and voicing control. *Journal of the Acoustical Society of America*, **85**, 1314-1321.
- MCCCLUSKEY, K., ALBAS, D., NIEMI, R., CUEVAS, C., & FERRER, C. (1975). Cross-cultural difference in the perception of the emotional content of speech: A study of the development of sensitivity in Canadian and Mexican children. *Developmental Psychology*, **11**, 551-555.
- OHMAN, S. E. G. (1965). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, **39**, 151-168.

- PIERREHUMBERT, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, **66**, 363-369.
- PIKE, K. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.
- REPP, B. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, **92**, 81-110.
- SCHERER, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, **99**, 143-165.
- SCHERER, K. R., LADD, D. R., & SILVERMAN, K. E. (1984). Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America*, **76**, 1346-1356.
- SHAPIRO, B., & DANLY, M. (1985). The role of the right hemisphere in the control of speech prosody in propositional and affective contexts. *Brain & Language*, **25**, 19-36.
- SHIPLEY-BROWN, F., DINGWALL, W., BERLIN, C., YENI-KOMSHIAN, G., & GORDON-SALANT, S. (1988). Hemispheric processing of affective and linguistic intonation contours in normal subjects. *Brain & Language*, **33**, 16-26.
- SIGNAL TECHNOLOGY, INC. (1978). *Interactive laboratory system* [Computer program]. Goleta, CA: Author.
- STARKWEATHER, J. A. (1961). Vocal communication of personality and human feelings. *Journal of Communication*, **11**, 63-72.
- STUDDERT-KENNEDY, M., & HADDING, K. (1973). Auditory and linguistic processes in the perception of intonation contours. *Language & Speech*, **16**, 293-313.
- SUMMERFIELD, Q. (1975). *Information-processing analyses of perceptual adjustments to source and context variables in speech*. Unpublished doctoral dissertation, Queen's University of Belfast.
- TARTTER, V. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics*, **27**, 24-27.
- TUCKER, D., WATSON, R., & HEILMAN, K. (1977). Discrimination and evocation of affectively intoned speech in patients with right parietal disease. *Neurology*, **27**, 947-950.
- ULDALL, E. T. (1960). Attitudinal meanings conveyed by intonation contours. *Language & Speech*, **3**, 223-234.
- ULTAN, R. (1978). Some general characteristics of interrogative systems. In J. P. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language: Volume 2. Phonology* (pp. 41-63). Stanford, CA: Stanford University Press.
- WEINTRAUB, S., MESULAM, M., & KRAMER, L. (1981). Disturbances in prosody—A right hemisphere contribution to language. *Archives of Neurology*, **38**, 742-744.
- WILLIAMS, C. E., & STEVENS, K. N. (1969). On determining the emotional state of pilots during flight: An exploratory study. *Aerospace Medicine*, **40**, 1369-1372.
- WILLIAMS, C. E., & STEVENS, K. N. (1972). Emotions and speech: Some acoustic correlates. *Journal of the Acoustical Society of America*, **52**, 1238-1250.
- WILLIAMS, C. E., & STEVENS, K. N. (1981). Vocal correlates of emotional states. In J. K. Darby (Ed.), *Speech evaluation in psychiatry* (pp. 221-240). London: Grune & Stratton.
- ZURIF, E. (1974). Auditory lateralization: Prosodic and syntactic factors. *Brain & Language*, **1**, 91-404.

## NOTES

1. Indeed, it might be argued that any use of final rise to convey questions reflects the emotional state of the speaker, since in most languages syntactic devices are available that can be used to indicate questions without the use of final rise. For example, question words, tags, word order, and particles are among the structures used to express interrogation in many of the world's languages, including English (Bolinger, 1958; Ultan, 1978). Alternatively, it may be that the use of final rise was, at some stage of prehistory, an unequivocal indication of the affective state of the speaker, but has undergone historical change, such that it is now used as a formal device, an alternative to syntactic devices, such as wh-movement, without necessarily implicating the speaker's emotional state.

2. Lieberman seems to have accepted some role for CT in stress-related F0 rises in his book with Blumstein (Blumstein & Lieberman, 1988).

3. We do not mean to suggest that this is the only reason for avoiding equally stressed adjacent syllables. Metrical factors are also likely to be relevant.

(Manuscript received January 21, 1992;  
revision accepted for publication August 17, 1994.)