urnal of Personality and Social Psychology
977, Vol. 35, No. 5, 345-350

# Pitch Changes During Attempted Deception

Lynn A. Streeter
Bell Laboratories
Murray Hill, New Jersey

Robert M. Krauss, Valerie Geller, Christopher Olson, and William Apple
Columbia University

Two studies on speech samples from 32 male college students are reported. In the first, it was shown that the average voice fundamental frequency of the subjects was higher when lying than when telling the truth. In the second, judges rated the truthfulness of 64 true and false utterances either from an audiotape that had been electronically filtered to render the semantic content unintelligible or from an unfiltered tape. The truthfulness ratings of the judges who heard the content-filtered tape were negatively correlated with fundamental frequency, whereas for the unfiltered condition, truthfulness ratings were uncorrelated with pitch. Although ratings made under the two conditions did not differ in overall accuracy, accuracy differences were found that depended on how an utterance had been elicited originally.

Although the primary function of spoken language is to convey semantic information, speech also contains information about a speaker's affective state. In English, such paralinguistic information can be transmitted by variations in pitch, amplitude, and articulation rate. In an early investigation of the relation between pitch and affective state, Fairbanks (1940) examined pitch variations in male actors attempting to convey each of five emotions while reading a standard paragraph of irrelevant content. Considerable variation in pitch patterns was found. Attempts to convey anger were characterized by rapid and broad pitch excursions, whereas an attitude of indifference was marked by relatively little pitch variation and a low fundamental-frequency value (the number of glottal pulses per second). Median fundamental frequency was about 100 Hz higher for fear and anger than it was for contempt, grief, and indifference. Interestingly enough, studies of the dimensional structure of the emotional space (both for facial expressions of emotion and for emotion words) have found that the former pair of emotions tends to be high on arousal or activity, whereas the latter group tends to be low on this dimension (Abelson & Sermat, 1962; Osgood, 1966).

More recently, Stevens and Williams and their associates have demonstrated a general relation between the amount of stress that a speaker is experiencing and the fundamental frequency of his voice. In one experiment (Hecker, Stevens, von Bismarck, & Williams, 1968), subjects were required to read six meters, sum their values, announce the sum, and then read a brief phrase. Stress was induced by progressively decreasing the illumination period of the meters. For those subjects who did not reduce their vocal amplitude in the stressed state, fundamental frequency rose under stress. Rather more dramatic evidence for the relation of stress to fundamental frequency has come from an analysis of pilots' radio transmissions during serious flight difficulties compared with transmissions made before such difficulties were encountered (Williams & Stevens, 1969). In all cases, median fundamental frequency was higher in the emotionally stressful situation.

In view of these findings, one might expect an elevation in fundamental frequency to be associated with attempts at deception. The classical psychophysiological approach to the

Requests for reprints should be sent to Robert M. Krauss, Department of Psychology, Columbia University, New York, New York 10027.

detection of deception—or "lie detection," as it is more popularly known—attempts to discriminate between false and truthful statements by measuring changes in autonomic activity, on the assumption that lying is to some degree arousing (Barland & Raskin, 1975; Bersh, 1969; Lykken, 1974).

To the extent this assumption is true, one would expect to find higher fundamental frequencies for deceptive utterances than for truthful ones, and indeed such a difference has been reported by Ekman, Friesen and Scherer (1976). These investigators analyzed the speech of subjects in a previous experiment (Ekman & Friesen, 1974) during and after viewing a disturbing and rather repulsive surgical film. The subjects' task was to convince an interviewer that the film was a pleasant one. Fundamental pitch was higher for the deceptive responses than it was when subjects described a pleasant film they actually had seen. Although the findings of Ekman et al. are consistent with the line of reasoning outlined above, they do not provide an adequate test of the hypothesis, because the pitch elevation that these investigators found cannot unequivocally be attributed to the deceptive act per se. Note that in their procedure the act of deception was confounded intentionally with exposure to the stressful (unpleasant) film. Because the Ekman and Friesen (1974) experiment did not include a condition in which the subjects who had seen the unpleasant film truthfully described the film they had seen, it is not possible to determine whether the increase in fundamental frequency resulted from the act of deceiving or simply from having just undergone an unpleasant, stressful experience.

In the present article we report data on pitch changes taken from an experimental situation in which the act of deception and exogenous stress were varied independently. The utterances of 32 pairs of male college undergraduates in a previously reported study by Krauss, Geller, and Olson (Note 1) were used.

## Experiment 1

### *Method*

The experiment that provided the speech samples used was structured as an interview situation. One member of each randomly assembled dyad was arbitrarily designated the interviewer and the other the interviewee. Care was taken to insure that dyad members were previously unacquainted. The interviewer was given an interview schedule containing five questions in each of four topic areas: politics, religion, personal future, and values. All questions concerned either the interviewee's personal beliefs or his plans for the future. The interviewers were instructed to ask the questions by using the wording and the order given in the interview schedule and not to improvise additional questions. Two of the four topics were designated for the interviewee as deception topics, and on these the interviewee was instructed to falsify his response—that is, to give an answer that did not correspond to his true belief or intention. The topics on which interviewees lied and those on which they told the truth were counterbalanced across dyads. The interviewer was told that the interviewee would be lying on two of the four topics but not, of course, on which ones. After each response, the interviewer rated its truthfulness on a 7-point scale.

In about half of the dyads, the interviews took place across a table; in the remainder, the subjects were located in separate rooms and communicated over a high-quality audio hookup. The two conditions will be referred to as face-to-face and intercom, respectively. In both conditions, both members of the dyad were videotape recorded and were so informed.

Crosscutting this in the original experiment was a second manipulated variable that was called *arousal*, but which could probably be described equally well by a number of other terms. The intention was to engage the subjects' motivation to deceive well by making the deception more consequential than simple compliance with the experimenter's request. Therefore, the interviewees in the arousal (but not those in the nonarousal) condition were told that previous research had shown the ability to deceive to be a general social skill, highly correlated with intelligence. They were also informed that the videotapes of their performance would be "evaluated" by a team of psychiatrists at Columbia's Psychiatric Institute. The nature of this deception was revealed in a step-by-step postexperimental debriefing. It is unclear how well this manipulation served its intended purpose. The aroused interviewees did not rate themselves as more nervous than the nonaroused interviewees on a postexperimental questionnaire, nor were they judged to be more nervous by their interviewers. Similarly, an independent sample of judges viewing the videotapes did not rate the aroused interviewers as more nervous than the nonaroused. Despite this, judges viewing the videotapes were significantly better at discriminating between the true and false statements of the aroused interviewees compared to those of the nonaroused interviewees, although this was not the case for interviewer judgments in the original experiment. In the present experiments we also found effects attributable to the arousal ma-

nipulation. Hence, although it was not clear that the manipulation had the effect that it was intended to have, it was also clear that it had some effect, and since we have been unable to think of a term that is more descriptive, we have held to the previous usage, which at least has the virtue of consistency.

The audiotape used in the present studies represents a balanced (across treatments) sample of 32 male dyads from the original study. There were two utterances from each interviewee, once telling the truth and once lying. (For further procedural details, see Krauss et al., Note 1.)

The 64 response segments were digitized and edited to remove silent portions by means of a waveform editing program (Nakatani, Note 2). The mean edited response length was 12.86 sec with a standard deviation of 10.34 sec. All segments were analyzed for fundamental frequency every 40 msec by means of a cepstrum method (Noll, 1964, 1967), a highly reliable pitch detection algorithm. Spectral analysis of the audiotape revealed the presence of an intermittent 60 Hz hum with a rich harmonic structure, which had been caused by fluorescent fixtures and air-conditioning equipment in an adjacent room. To deal with this problem, all potentially specious pitch values of 60, 120, and 180 Hz plus or minus 2 Hz were excluded from the analysis.

### Results

A 2 (Truth/Lie) × 2 (Face-to-Face/Intercom) × 2 (Arousal/Nonarousal) mixed analysis of variance with repeated measures on the first factor was computed on the mean fundamental frequency of the 64 segments. There was a significant main effect for the Truth/Lie factor, $F(1, 28) = 4.47$, $p < .05$. On the average, fundamental frequency was 3.3 Hz higher when the subjects lied than when they told the truth. The analysis also revealed a marginally significant interaction between the Truth/Lie variable and Arousal, $F(1, 28) = 3.05$, $p < .10$. The average difference in fundamental frequency between truthtelling and lying was greater for the aroused subjects than for those who were not aroused by an average of 5.4 Hz. Mean truth-minus-lie fundamental frequency differences for the two between-subjects conditions are shown in Table 1.

Table 1
*Mean Truth-Minus-Lie Pitch Differences (in Hz) by Experimental Condition*

| Interview | Interviewee: Aroused | Interviewee: Nonaroused | *M* |
|---|---|---|---|
| Face-to-face | −4.8 (5) | −.1 (4) | −2.5 |
| Intercom | −7.1 (7) | −1.0 (6) | −4.1 |
| *M* | −6.0 | −.6 | −3.3 |

*Note.* n = 8 per cell. Values in parentheses indicate number of interviewees in that cell who had higher pitch values while lying than while truthtelling.

### Discussion

These results provide clear support for the hypothesis that higher pitch is associated with deceptive utterances. Moreover, this difference in pitch between lying and truthtelling was greater for the subjects who underwent an arousal manipulation than for those who did not. The significance levels reported here are probably conservative estimates of the reliability of pitch differences in the experimental situation. The necessity of removing all values that could have been produced by a 60 Hz hum also removed legitimate data points, because the average fundamental frequency of male voices ranges between 100 and 200 Hz (Brosnahan & Malmberg, 1975). Other things being equal, recordings made in an acoustically more satisfactory environment should yield a larger number of data points and, hence, more reliable differences.

Given that increases in fundamental frequency are reliably associated with deceptive responses, it seems worthwhile to ask whether listeners use this information to detect deceptive responses. Interestingly, although listeners readily list a number of behaviors as cues to deception (e.g., nervousness, gaze aversion, facial shielding), increased pitch is not among them. Since semantic information and paralinguistic cues are both carried in the vocal channel, we decided to run a condition in which subjects heard speech that had been rendered unintelligible, in addition to speech that was unmutilated.

## Experiment 2

### Method

*Procedure.* The 64-segment audiotape from Experiment 1 was presented to groups of subjects who

Table 2
*Mean Detectability Index by Experimental Condition*

| Interview | Audiotape: Filtered | Audiotape: Unfiltered | *M* |
|---|---|---|---|
| Face-to-face | .18 | −.01 | .09 |
| Intercom | .08 | .33 | .20 |
| *M* | .13 | .16 | .14 |

*Note.* n = 15 raters per cell. Values in cells represent means of summed truthfulness ratings given to true statements minus summed ratings given to false statments.

rated the truthfulness of each response on a 7-point bipolar scale, ranging from (1) "not at all truthful" to (7) "completely truthful." Two versions of the tape were presented. One was an undoctored copy of the original; the other had been rerecorded through a circuit described by Rogers, Scherer, and Rosenthal (1971). The procedure, termed *content filtering*, effectively renders speech unintelligible while preserving such vocal features as loudness, intonation, tempo, length, and, of course, pitch. The Rogers et al. circuit permits bandpass filtering as well as the introduction of distortion where desirable. In the present study, all frequencies above 2000 Hz were removed and all frequencies below 650 Hz were distorted by peak clipping. The result sounded similar to speech heard through a wall—identifiable as speech, but unintelligible.

*Subjects.* A total of 30 undergraduate volunteers (24 males and 6 females) served as subjects. They were paid $3 for participating. Half of them heard the content-filtered version of the 64-segment tape; the remainder heard the intact version. The subjects were run in groups of three to seven.

The 64 segments were arranged in random order and the order was the same for all presentations. For the subjects who heard the content-filtered version, the interviewer's question (which preceded the interviewee's response) was not distorted. In both conditions, the subjects listened to the audio tapes over high-quality speakers.

### *Results*

Mean truthfulness ratings were computed for each of the 64 segments, and the mean ratings given to the filtered and unfiltered versions were correlated. The resulting correlation of .39 ($p < .01$) indicated that there was better than chance agreement between the two groups.

Overall, the truthfulness ratings of the content-filtered responses were significantly correlated with the interviewee's average fundamental frequency on the segment being rated ($r = -.26$, $p < .05$); the higher the average pitch, the less truthful it was judged. For raters who heard the unfiltered tape, the corresponding correlation was essentially zero ($r = -.01$).

It will be recalled that the stimulus tapes contained the responses of 32 respondents who each appeared twice, once telling the truth and once lying. A detectability index was computed by subtracting the truthfulness ratings given to false responses from the ratings given to true responses. A 2 (Face-to-Face/Intercom) × 2 (Arousal/Nonarousal) × 2 (Filtered/Unfiltered) mixed analysis of variance was computed on the detectability index. No main effect for rating condition was found; overall, raters were no better at discriminating true from false responses on the basis of content-free speech than they were on the basis of the unfiltered recordings. However, a significant Filtered/Unfiltered × Face-to-Face/Intercom interaction, $F(1, 28) = 4.20$, $p < .05$, indicated that the relative accuracy of the two rating conditions depended to some extent upon whether the original interview was conducted face-to-face or over an intercom. As is shown in Table 2, for intercom interviews, the detectability index was higher (i.e., true responses were better distinguished from false responses) for judgments of unfiltered speech than for judgments of content-filtered speech. For interviews conducted face-to-face, the reverse was true. A significant Face-to-Face/Intercom × Aroused/Nonaroused interaction was also found, $F(1, 28) = 4.46$, $p < .05$. Interviewees in the intercom condition were more detectable when they had been aroused than when they had not been aroused; the opposite was true for interviewees in the face-to-face condition.

## General Discussion

The results of our first experiment indicate that a speaker's pitch tends to be higher during lying than during truthtelling and that the magnitude of this difference is marginally greater when the act of deceiving is arousing or stressful. Our second experiment indicates

that although raters do not ordinarily use pitch variations as a cue for deception detection, they will do so to some extent when semantic content is unintelligible.

The significant positive correlation between the truthfulness ratings made in the filtered and unfiltered conditions might be interpreted to indicate that judgments in the unfiltered condition are not based exclusively on the semantic information contained in the utterance. Because the unfiltered tape transmitted all of the paralinguistic information contained in the content filtered tape as well as the additional semantic information, it is tempting to interpret the common variance (i.e., 15%) as the contribution of paralinguistic features to the ratings of the unfiltered tape. However, such an interpretation assumes that the two sources of information —paralanguage and semantic content—are orthogonal, and such an assumption is probably unreasonable. Certainly, one would expect that reactions to stress, which are expressed paralinguistically by variations in pitch for instance, would also be reflected in semantic content by uncertainty in word selection, the presence of nonjuncture pauses, and the like. We cannot tell from our data whether the subjects who listened to the unfiltered tapes used cues of this sort rather than paralinguistic cues for their judgments. What we can say is that the paralinguistic contribution to our unfiltered ratings constitutes no more than 15% of the total variance.

Although the subjects who heard the unfiltered tape do not appear to have made use of variations in pitch to detect deception, the subjects for whom the content was inaccessible apparently did. Again, however, we cannot state with certainty that the content-filtered subjects were responding to pitch cues per se because within-speaker changes in voice frequency are likely to be correlated with other acoustic cues as well. For example, differences in articulation rate (number of syllables per second) between true and false utterances were moderately correlated with pitch differences for the same segments ($r = -.24$, $p < .10$, one-tailed). What does seem reasonable to conclude is that pitch variation or other features that covary with it contributed more to the judgments of the subjects hearing content-filtered speech than to those of the subjects hearing unfiltered speech.

It is interesting that the raters who heard the content-filtered speech were no more accurate in detecting deception than were those who heard the unfiltered speech. The conditions do differ, however, depending upon whether the deceptive statement was made in a face-to-face interview or in an intercom situation. The raters who heard the unfiltered speech were better able to detect deception on the part of the intercom interviewees than deception by those in the face-to-face condition, a result that incidentally replicates a previous finding (Krauss et al., Note 1). For responses elicited in the face-to-face condition, the raters who heard content-filtered speech were somewhat more accurate than the raters who had access to content. However, there is nothing in our results to suggest that it is advisable to ignore the content of an utterance when trying to decide whether it is true or false. Subjects typically rely heavily on content (Kraut, Note 3), and it is probably reasonable that they should. Nevertheless, the present data suggest that speech does contain a measure of useful information apart from its semantic content that is not ordinarily utilized when assessing the truthfulness of an utterance.

Although all of our subjects were males, we see no reason in principle that similar results should not be obtained with females. Care should be taken, however, not to overinterpret these results in terms of their practical implications. Although deceptive responses tended to be accompanied by an elevation in pitch, the extent of this rise in pitch varied as a function of experimental condition and, even under the best conditions, did not hold for all subjects (see Table 1). Therefore it seems to us that the use of pitch change in practical lie-detection schemes will be subject to the same problems that are encountered in the more conventional, psychophysiologically based methods (Lykken, 1974).

## Reference Notes

1. Krauss, R. M., Geller, V., & Olson, C. T. *Modalities and cues in the detection of deception.* Paper

presented at the meeting of the American Psychological Association, Washington, D.C., September 1976.

2. Nakatani, L. H. *A guide to editor: An interactive program for editing digital tapes.* Unpublished manuscript, 1974.
3. Kraut, R. E. *Verbal and nonverbal cues in the perception of lying.* Paper presented at the meeting of the American Psychological Association, Washington, D.C., September 1976.

## References


Abelson, R. P., & Sermat, V. Multidimensional scaling of facial expression. *Journal of Experimental Psychology,* 1962, *63,* 546–554.

Barland, G. H., & Raskin, D. C. An evaluation of field techniques in detection of deception. *Psychophysiology,* 1975, *12,* 321–330.

Bersh, P. J. A validation study of polygraph examiner judgments. *Journal of Applied Psychology,* 1969, *53,* 399–403.

Brosnahan, L. F., & Malmberg, B. *Introduction to phonetics.* Cambridge, England: Cambridge University Press, 1975.

Ekman, P., & Friesen, W. V. Detecting deception from body or face. *Journal of Personality and Social Psychology,* 1974, *29,* 288–298.

Ekman, P., Friesen, W. V., & Scherer, K. Body movements and voice pitch in deceptive interaction. *Semiotica,* 1976, *16,* 23–27.

Fairbanks, G. Recent experimental investigations of vocal pitch in speech. *Journal of the Acoustical Society of America,* 1940, *11,* 457–466.

Hecker, M. H. L., Stevens, K. N., von Bismarck, G., & Williams, C. E. Manifestations of task-induced stress in the acoustic speech signal. *Journal of the Acoustical Society of America,* 1968, *44,* 993–1001.

Lykken, D. T. Psychology and the lie detector industry. *American Psychologist,* 1974, *29,* 725–739.

Noll, A. M. Short-time spectrum and cepstrum techniques for vocal pitch detection. *Journal of the Acoustical Society of America,* 1964, *36,* 296–302.

Noll, A. M. Cepstrum pitch determination. *Journal of the Acoustical Society of America,* 1967, *41,* 293–304.

Osgood, C. E. Dimensionality of the semantic space or communication via facial expression. *Scandanavian Journal of Psychology,* 1966, *7,* 1–30.

Rogers, P. L., Scherer, K. R., & Rosenthal, R. Content filtering human speech: A simple electronic system. *Behavior Research Methods and Instrumentation,* 1971, *3,* 16–18.

Williams, C. E., & Stevens, K. N. On determining the emotional state of pilots during flight. An exploratory study. *Aerospace Medicine,* 1969, *40,* 1369–1372.


Received November 29, 1976 ■