Pitch Changes During Attempted Deception

Lynn A. Streeter
Bell Laboratories
Murray Hill, New Jersey
Robert M. Kraus, Valerie Getler, Christopher Olson, and
William Apple
Columbia University

Two studies in speech samples from 50 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 context-distilled 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 partial pulses per second). Median fundamental frequency was about 100 Hz higher for joy 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, 1961; Osgood, 1960).

More recently, Stevens and Williams and their associates have demonstrated a general relation between the amount of stress that a speaker perceives and the fundamental frequency of his voice. In one experiment (Hecker, Stevens, von Bismarch, & Williams, 1961), subjects were required to read six metters, 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 increased 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. Kraus, 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 & Roskin, 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 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 truthfully designated the interrogator 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: political, religious, personal future, and values. All questions were designed either the interviewee's personal beliefs about his place in the future. The interviewees were instructed to ask the questions by using the wording and the order given in the interview schedule; however, they were not to improve additional questions. Two of the four topics were designated as interview deception topics, and on these the interviewee was instructed to falsify his responses—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 interviewee was told that the interview would be 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 videotaped and were so informed.

Crosscutting this in the original experiment was a second manipulated variable that was called around, but which could probably be described "bogus" well by a number of other terms. The intent was to engage the subjects' motivation to deceive well by making the deception and consequences simpler than simple compliance with the experimenters' request. Therefore, the interviews in the present study (but not those in the nonmanipulated condition) were told that previous research had shown the ability to deceive to be correlated with a particular cognitive process. They were told after this deception was revealed in a step-by-step postexperimental debriefing. It is conceivable that 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 interviewees 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 interviewers judgments in the original experiment. In the present experiments we also found effects attributable to the arousal ma-
PITCH CHANGES DURING ATTEMPTED DECEPTION

Table 1

<table>
<thead>
<tr>
<th>Interview</th>
<th>Face-to-face</th>
<th>Intercom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aroused</td>
<td>-4.8</td>
<td>-7.3</td>
</tr>
<tr>
<td>Non-aroused</td>
<td>-1.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>M</td>
<td>-2.5</td>
<td>-3.3</td>
</tr>
</tbody>
</table>

Face-to-face: 7
Intercom: 7

Note: n = 3 per cell. Values in parentheses indicate number of interviewers in that cell who had higher pitch values while lying than while truth-telling.

Once in pitch between lying and truth-telling was greater for the subjects who underwent an unusual 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 50 Hz hum also removed legitimate data points, because the average fundamental frequency of male voices ranges between 100 and 200 Hz (Bromhahan & Malmberg, 1973). 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 shading), 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 unmonitored.

Experiment 2

Method

Procedure. The 64-segment audiotape from Experiment 1 was presented to groups of subjects who
Table 2
Mean-Recallability Index by Experimental Condition

<table>
<thead>
<tr>
<th>Audiotape</th>
<th>Interview</th>
<th>Filtered</th>
<th>Unfiltered</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face-to-Face</td>
<td>.18</td>
<td>.18</td>
<td>.09</td>
<td>.19</td>
</tr>
<tr>
<td>Intercom</td>
<td>.14</td>
<td>.19</td>
<td>.20</td>
<td>.19</td>
</tr>
<tr>
<td>M</td>
<td>.15</td>
<td>.16</td>
<td>.14</td>
<td></td>
</tr>
</tbody>
</table>

Note. *n* = 15 rater per cell. Values in cells represent mean of summed truthfulest ratings given to true statements minus summed ratings given to false statements.

correlated with the interviewer's average fundamental frequency on the segment being rated (r = -.29, 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, which each repeated 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 (Face-to-Face/Intercom) × 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 vocal features 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 × Arousal/Nonarousal interaction was also found, F(1,28) = 4.46, p < .05. Interviewers in the intercom condition were more detectable when they had been aroused than when they had not been around; 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 truth-telling 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 rats do not ordinarily use pitch variations as a cue for deception detection, they will do so to some extent when semantic content is unattainable.

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 biased 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 (by, 13%) as the contribution of paralinguistic features to the ratings of the unfiltered tape. However, such an interpretation assumes that the two sources of information—paralinguistic and semantic content—are orthogonal, and such an assumption is probably unreasonable. Certainly, one would expect this relationship to stress, which are expressed paralinguistically by variations in pitch for instance, would also be reflected in semantic content by a greater word selection, the presence of nonstandard 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 contributions to our unfiltered ratings constitute no more than 13% 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 unacceptable did appear to have done so. Again, however, we cannot judge 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 pitch variation rate (number of syllables per second) between true and false utterances were moderately correlated with pitch differences for the same segments (r = .38, P < .01, one-tailed). This does seem reasonable to conclude that pitch variations were other features that convey with it connotative tone to the judgments of the subjects hearing content-filtered speech than to those of the subjects hearing unfiltered speech.

It is interesting that the rats who heard the content-filtered speech were as more accurate in detecting deception than were those who heard the unfiltered speech. The condition do differ, however, depending upon whether the deceptive statement was made in a face-to-face interview or in an interview situation. The rats who heard the unfiltered speech were better able to detect deception on the part of the interrogator interviewers than deception by those in the face-to-face condition, a result that initially replicates a previous finding (Kraus et al., Note 1). For responses elicited in the face-to-face condition, the rats who heard content-filtered speech were somewhat more accurate than the rats who had access to context. 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 only rely heavily on content (Kraus, Note 3), and it is probably reasonable to assume 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 ascertaining 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 trended 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 we see in counter with the more conventional, psychophysiological-based methods (Lykken, 1974).

Reference Notes

1. Kraus, R. N., Gallier, E., & Olof, C. T. Mediality and cue in the detection of deception. Paper...


References


Received November 29, 1976

---

The examine with the pschologic terrestial rides has zation o desire o far (Lerner, 1967: L per cent and the over on star, 1974 Accar have a they des series o his cole

---

This is National second a to Phils, comment the authors other pe Chicago Rogers Norman University