

Effects of Pitch and Speech Rate on Personal Attributions

William Apple
Columbia University

Lynn A. Streeter
Bell Laboratories
Murray Hill, New Jersey

Robert M. Krauss
Columbia University

In three experiments, subjects listened to recordings of male speakers answering two interview questions and rated the speakers on a variety of scales. The recordings had been altered so that the pitch of the speakers' voices was raised or lowered by 20% or left at its normal level, and speech rate was expanded or compressed by 30% or left at its normal rate. The results provided clear evidence that listeners use these acoustic properties in making personal attributions to speakers. Speakers with high-pitched voices were judged less truthful, less emphatic, less "potent" (smaller, thinner, faster), and more nervous. Slow-talking speakers were judged less truthful, less fluent, and less persuasive and were seen as more "passive" (slower, colder, passive, weaker) but more "potent." However, the effects of the acoustic manipulations on personal attributions also depended on the particular question that elicited the response.

Human speech provides a listener with at least two sources of information: a *verbal* channel, encoding the message's linguistic content, and a *vocal* channel, conveying paralinguistic information by variations in pitch, speech rate, loudness, and the like.

One important type of information communicated via the vocal channel concerns a speaker's affective state. The vocal characteristics associated with the expression of emotion are beginning to be understood (see, for example, Fairbanks, 1940; Hecker, Stevens, von Bismarck, & Williams, 1968; Williams & Stevens, 1972). It is now reasonably well established that stressful situations raise the voice's fundamental frequency (the number of glottal pulses per second) and that "active" emotions such as anger and fear tend to be reflected in increased mean pitch and

pitch variance, whereas "low energy" states such as sorrow and indifference are associated with a lower mean pitch and a slower speech rate.

Given that emotional states do differ reliably in their paralinguistic expression, to what extent do listeners use these vocal cues in judging the immediate affective state or more enduring personality traits of a speaker? An early investigation of the noncontent aspects of speech by Allport and Cantril (1934) demonstrated that listeners could judge, at better than chance levels, a speaker's age and at least some personality characteristics from voice alone: In four of six experiments, speakers' scores on a test of ascendance-submission (Allport's A-S reaction study) were judged with significant accuracy. However, reviewing much of the voice-attribution work, Kramer (1963) concluded that more common than accuracy in such judgment studies was the finding of "vocal stereotypes." That is, certain voices were reliably, though sometimes incorrectly, judged as belonging to certain personality types.

Unfortunately, as Kramer noted, the description of the vocal parameters that under-

We wish to thank John Lin and Nina Macdonald for their valuable assistance. We are also grateful to Myron Wish, Nancy Morency, and four anonymous reviewers for their helpful comments.

Requests for reprints should be sent to Robert M. Krauss, Department of Psychology, 400 Schermerhorn Hall, Columbia University, New York, New York 10027.

lie such stereotypes leaves much to be desired. While not all the earlier research on voice attributions can be faulted on these grounds, studies attempting to specify critical stimulus dimensions often have neglected or confounded important paralinguistic cues. For example, in reporting two recent field experiments, Miller, Maruyama, Beaver, and Valone (1976) concluded that the persuasiveness of a communication can be directly related to the rate at which it is delivered. More rapid speech was found to be more persuasive, presumably because a fast talker is viewed as more credible. Although Miller et al. have ruled out certain alternative explanations (such as the effect of having limited opportunity to counterargue against a rapid presentation), the stimulus materials used in their experiments warrant closer examination. In both experiments the persuasive messages were recorded by the same speaker at either a slow or a fast speech rate. This was accomplished "by simply instructing the speaker to practice delivering the same speech as rapidly and slowly as possible while controlling his level of enthusiasm and involvement" (Miller et al, 1976, p. 618). The stimulus recordings in the two conditions do, of course, differ in speech rate, but it is quite likely that they differ in other respects as well. In natural speech, such vocal parameters as amplitude, pitch, and rate tend to covary. For example, rapid speech is likely to be louder and higher pitched than normal speech (Black, 1961). Consequently, it is quite possible that subjects in the Miller et al. study were responding to pitch and/or loudness cues as well as to rate.

Because it is so difficult to assess, in a controlled way, the contribution of various vocal parameters to the attribution process using natural speech, a number of workers have attempted to deal with the problem by using nonspeech stimuli. For example, Scherer (1974) presented listeners with simple tone sequences generated on a Moog synthesizer, in which a minimal set of acoustic cues (pitch level and variation, amplitude level and variation, and tempo) were varied factorially; listeners rated the emotional quality of these tone sequences on a set of semantic

differential scales. Scherer found that judgments were most influenced by tempo and pitch variations: Fast tempo led to attributions of highly active and potent emotions (e.g., interest, anger, and happiness) and slow tempo to attributions of sadness, disgust, and boredom; extreme pitch variation with rising contours produced ratings of highly pleasant, active, and potent emotions (happiness, surprise, interest). Other parameters influenced the perception of specific emotional qualities. While Scherer's method effectively deals with the confounding of vocal qualities in natural speech and suggests the effect that independently varied acoustic cues can exert, in view of the artificial nature of the stimuli used, it is not clear how directly the results can be generalized to the processing of speech.

Other investigators, using natural speech, have achieved a degree of control over stimulus materials by means of editing techniques. For example, investigators have added or removed such speech disfluencies as filled pauses and repetitions from stimulus audiotapes and asked listeners to judge speakers' credibility and other personal attributes (Lay & Burron, 1968; Miller & Hewgill, 1964). Typically, highly hesitant or disfluent speakers are assigned relatively undesirable personality traits, and their communications are judged to be low in credibility.

While the methods thus far described have all afforded some degree of stimulus control, recent developments in speech synthesis technology permit investigators to vary independently one or more parameters of natural speech. This approach has been exploited by Brown and his colleagues (Brown, Strong, & Rencher, 1973, 1974; Smith, Brown, Strong, & Rencher, 1975), who have manipulated speech rate, mean fundamental frequency, and fundamental frequency variance, using a computer-based analysis-synthesis system. Two major personality dimensions (termed "competence" and "benevolence") have emerged from factor analyses of judgments of such manipulated speech. Generally speaking, higher pitch seems to result in a speaker being judged less competent and less benevolent (Brown et al., 1974), whereas faster speech rates produce judgments of higher

competence but yielded an inverted-U relationship on the benevolence dimension (Smith et al., 1975).

The studies of Brown and his associates make an important contribution to our understanding of the significance that listeners ascribe to vocal qualities. Nevertheless, they are not free of methodological problems. They typically have relied on a small number of stimulus voices, all of whose acoustic permutations are presented to the same raters. For example, the Brown et al. (1974) study used only two adult male speakers uttering the same sentence ("We were away a year ago"), which was then manipulated to fill the 27 cells of a $3 \times 3 \times 3$ factorial design. It is not clear that judgments based on hearing 54 repetitions of the same content bear much similarity to the kinds of everyday personality ascriptions listeners make under more natural conditions. Additionally, the particular parametric values used in these studies are problematic. Brown et al. (1974) reported that in the high-pitch condition, the original fundamental frequency was multiplied by a factor of 1.8. Although they did not report average fundamental frequency values for their two speakers, assuming that they fell into the normal range for males, Brown et al.'s high-pitch manipulation would have raised a male voice into the female range. Similar criticisms apply to their choice of speech-rate scale factors.

The present article reports an exploratory series of experiments designed to demonstrate the effects of two acoustic parameters—average fundamental frequency and speech rate—on judgments of several state and trait variables. It was hoped that by using a large number of speakers and naturally produced utterances, together with more conservative values for acoustic alteration, and by independently manipulating the two parameters factorially, we could overcome the methodological problems noted above.

Experiment 1

Experiment 1 derives from a finding reported by Streeter, Krauss, Geller, Olson, and Apple (1977), who demonstrated that the

mean fundamental frequency of speakers' voices increased when they lied, relative to a truth-telling baseline, and that this effect was stronger when the speaker had been motivated to lie effectively. In addition, Streeter et al. found that although judges' ratings of truthfulness were essentially uncorrelated with pitch level, there *was* a significant negative correlation between judged truthfulness and average pitch for listeners who heard the speech after it had been passed through a content filter—a device that destroys intelligibility without affecting vocal features of the utterances. Apparently, the natural pitch increments during lying (on the average about 3 Hz) were too small to affect the judgments of listeners who had content available but were taken into account by listeners who could not understand the responses' verbal content. However, since pitch increments and speech rate decrements were correlated in their data, Streeter et al. could not rule out the possibility that listeners were attending to rate differences rather than pitch differences between true and false utterances. From the finding of Miller et al. (1976), one would expect slower speech to result in lower perceived speaker credibility. In Experiment 1 we assessed the effects of these two variables on truthfulness judgments.

Method

Stimulus materials. Forty male Columbia College undergraduates, all native speakers of English, answered questions for use as stimulus materials. They were individually recorded in a sound-isolated booth on high-quality audio equipment, and they received course credit for their participation. Each speaker answered six standard questions dealing with his opinion on a range of topics. The questions were administered in a fixed order, and speakers were instructed to answer all questions honestly and frankly. They were told to give brief answers, but to give more than a yes-or-no response. All speakers were debriefed concerning how their interviews might be used and signed informed-consent releases permitting the later use of the recordings.

Two of the six interview questions were selected for use in the experiment. One asked the subject's opinion of college admissions quotas designed to favor minority groups (Question 3); the other asked the subject what he would do if he suddenly won or inherited a large sum of money (Question 6). These two questions were chosen because both elicited a diversity of responses and because the two differed

on a dimension of personal salience for the respondents. (The question of quotas is a matter of genuine concern and frequent discussion among Columbia undergraduates.) Twenty-seven speakers, who gave responses that were not excessively long and represented the spectrum of opinions on the two questions, were selected.

Each of the 27 speakers was then randomly assigned to one of nine cells of a 3 (rate: slow, unmanipulated, fast) \times 3 (pitch: low, unmanipulated, high) completely crossed factorial design, with three speakers in each cell. All speech material was digitized and analyzed by the linear predictive coding (LPC) method of Atal and Hanauer (1971). (The LPC analysis calculates 14 parameters every 10 msec. Twelve of these parameters represent pseudoarea functions of the vocal tract; the other two are values for amplitude and fundamental frequency. The advantage of LPC analysis is that such variables as rate, pitch, and amplitude can be manipulated without changing other voice parameters. In addition, since the parameters are derived from the original speech, the quality of the resulting synthesis is high.) The two responses of each speaker were manipulated on a DDP-224 computer using an interactive program (Nakatani, Note 1) that displays the parameters as functions of time and allows the user to manipulate any or all of the 14 parameters as well as to linearly expand or compress the time base. The altered utterances can then be synthesized and recorded on audiotape.

The scale factors chosen for the low- and high-pitch manipulations were 80% and 120%, respectively, of the speaker's unmanipulated fundamental frequency. The values chosen for the slow and fast speech rate manipulations were 70% compression and 130% expansion, respectively, of the utterance's time base, resulting in speech rates that were 77% and 143% of the unmanipulated rates.¹ These scale values were chosen because the resulting speech still sounded natural and the acoustic properties remained more or less within the normal range of values. (See Hanley, 1951; Mysak, 1959; Peterson & Barney, 1952; and Terango, 1966, for data on pitch and Goldman Eisler, 1968, for normative speech rate data.) The mean premanipulation fundamental frequency and speech rate for our speakers were 109.79 Hz ($SD = 14.82$) and 3.27 syllables/sec ($SD = .66$). Following manipulation, all utterances were resynthesized to produce a stimulus audiotape of the answers to the quota question (in one random order) and a tape of the answers to the money question (in a different order). The tapes were formatted to allow 15 sec of silence between each response.

Procedure. Twenty undergraduates, 14 males and 6 females, were paid for their participation.² Raters were run in groups of three to seven. They were told that the purpose of the study was to determine how well people can tell, from the sound of the speaker's voice, whether someone is lying or telling the truth. They were informed that approximately half the

speakers had been instructed to tell the truth, whereas the remaining half had been told to lie—that is, to give answers that did not correspond to their actual beliefs or feelings.

The two stimulus tapes were played over high-quality audio equipment with the order of tape presentation counterbalanced. A 500-Hz warning tone preceded and followed each answer by 1 sec. Raters were given 15 sec in which to rate the truthfulness of the answer on a 7-point scale ranging from "not at all truthful" (1) to "entirely truthful" (7).

Because of variations in the quality of the recordings, raters were also told that recordings had been made using a variety of equipment in several different environments.³ The experimental session lasted about an hour, and there was a 5-min. rest period between tapes. Participants in the three rating studies were sent a report describing the purpose and detailing some of the findings of the study several weeks after its conclusion.

Results

Some of the premanipulation characteristics of the stimulus materials are summarized in Table 1. Note that the more involving topic, college admissions quotas, resulted in significantly longer, slower, and higher pitched responses. Since the quota topic always preceded the money topic in the original order, we cannot rule out the possibility that these differences are due to serial position effects.

¹ The rate manipulation algorithm linearly expanded or compressed consonant and vowel information alike. In natural speech, rate variation tends to be reflected more in changes of vowel duration than in consonant duration. However, the rate manipulation used resulted in reasonably natural-sounding speech.

It may not be immediately apparent why 70% compression of an utterance's time base results in a speech rate that is 143% of the original. Consider a response consisting of n syllables spoken over s seconds of time; that utterance's rate would be n/s syllables/sec. Seventy-percent compression changes the rate to $n/(.7s)$, or $1.43 n/s$ —a faster speech rate that is 143% of the original. Likewise, a 130% expansion of an utterance's time base results in a slower speech rate that is 77% of the original rate.

² One additional rater was run, but his data were discarded after he expressed some suspicion regarding possible splicing of the tapes. No other raters voiced any suspicion that the stimuli had been altered.

³ There is some quality variation across speakers in the LPC synthesis; in particular, speakers who tend to mumble and have a high degree of nasalization seem to suffer the greatest quality degradation.

Table 1
Premanipulation Characteristics of Stimulus Tapes

| | Quota tape | | Money tape | | <i>t</i> |
|-----------------------------|------------|-----------|------------|-----------|----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | |
| Response length (syllables) | 111.1 | 40.7 | 84.6 | 42.7 | 3.89** |
| Response time (sec) | 36.8 | 13.8 | 25.6 | 13.2 | 4.68** |
| Speech rate (syllables/sec) | 3.0 | .50 | 3.4 | .74 | 2.12* |
| Fundamental frequency (Hz) | 112.2 | 14.9 | 107.4 | 14.6 | 5.60*** |

Note. *N* = 27 segments per tape.

* $p < .05$, two-tailed. ** $p < .01$, two-tailed. *** $p < .001$, two-tailed.

However, it seems more plausible to regard these differences as reflections of differences in our subjects' personal involvement with the two questions and/or the cognitive complexity of the answers they called for (see Goldman Eisler, 1968, and Williams & Stevens, 1969).

To check the adequacy of the random assignment, premanipulation speech rate (syllables/sec) and average pitch were subjected to analyses of variance (3 rate levels \times 3 pitch levels \times 2 questions) with speakers nested within the rate and pitch factors. In addition to the between-questions differences already noted, the analysis of the premanipulation pitch failed to reveal other significant effects. However, there was a marginally significant difference in premanipulation speech rate across the three assigned rate conditions, $F(2, 18) = 3.00$, $p < .08$, primarily because of some slower speakers' random assignment to the slow rate condition. Therefore, prior to all further analyses, we adjusted the raw data for covariation on speakers' premanipulation speech rate and fundamental frequency: The appropriate beta weights were derived from linear regression of the data collapsed across raters. This adjustment has the virtue of controlling for spurious rate and question effects in the analyses of variance.⁴ Having thus transformed each dependent variable, we computed min-*F* ratios (and their approximate degrees of freedom) for the analyses of variance. The *F'* statistic we report (Winer, 1971, pp. 375-378) treats both speakers and raters as random effects, permitting simultaneous generalization over both groups.⁵

A 3 (pitch levels) \times 3 (rate levels) \times 2

(questions) analysis of variance was performed on the mean truthfulness ratings with repeated measures on the last factor. Premanipulation speech rate and premanipulation fundamental frequency were used as covariates, and all means reported below have been adjusted for these covariates. A significant main effect was found for the pitch manipulation, $F''(3, 34) = 3.37$, $p < .05$, and a marginally significant effect for rate, $F''(3, 51) = 2.26$, $p < .10$. The mean truthfulness ratings for the three pitch conditions (going from low to high pitch) were 4.62, 4.40, and 4.09, respectively, indicating that lower pitch enhanced credibility. The corresponding means for the rate manipulation were, going from slow to fast rate, 4.10, 4.63, and 4.37; the unmanipulated rate was judged most credible and the slow rate least credible.

No significant effect was found for the

⁴ To ensure that our subjects' ratings were not affected by variations in acoustic quality across our nine experimental conditions, we had six undergraduates rate the 54 recorded segments for intelligibility and correlated these ratings with our subjects' ratings of truthfulness. The two sets of ratings were not significantly correlated ($r = .15$). Similarly, to ensure that response content was well distributed across conditions, we had 12 undergraduates rate from a transcript how pro- or antiquota (for the quota question) or generous or selfish (for the money question) each response was. Their ratings were then subjected to a 3 (pitch levels) \times 3 (rate levels) analysis of variance. For neither question were significant main effects or interactions found.

⁵ In cases where the min-*F* value (*F'*) is marginal, we will also report the conventional *F* ratios for raters (considering speakers as a fixed effect) and speakers (considering raters as a fixed effect). These values represent less conservative tests.

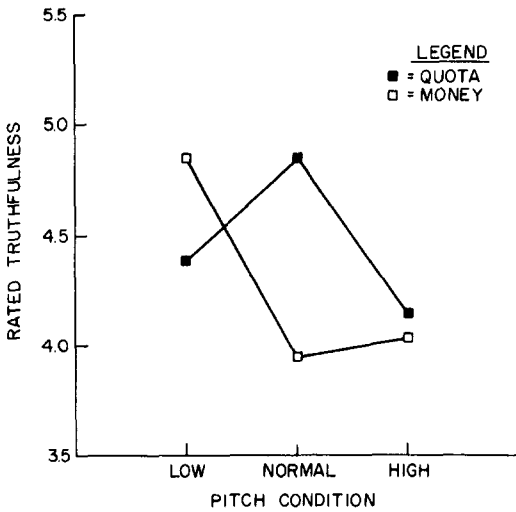


Figure 1. Average rated truthfulness plotted as a function of pitch condition (low, normal, high) for each of the two question topics.

Pitch \times Rate interaction. However, there was a nearly significant Pitch \times Question effect, $F''(2, 32) = 3.02, p < .10$; raters' $F(2, 38) = 10.25, p < .001$; speakers' $F(2, 18) = 3.84, p < .05$. For the quota question, truthfulness ratings and pitch were curvilinearly related; for the money question, low-pitched speakers were judged as most truthful. The two sets of means are shown in Figure 1.⁶

Discussion

The results demonstrate that the acoustic manipulations performed on the speech stimuli affected judgments of truthfulness. Consistent with the findings of Streeter et al. (1977), judges rated high-pitched voices as less truthful than lower pitched voices. Perhaps listeners perceived high pitch to be an indication of stress and attributed such stress to attempted deception. The pitch manipulations used, although not extreme enough to place voices outside the normal male pitch range, were evidently large enough to produce attributions of lying from naive listeners. It will be recalled that the smaller, naturally occurring pitch increments accompanying deception in the Streeter et al. experiment did not evoke such attributions, except in the filtered listening condition, in which verbal content was unintelligible.

While the rate effect was only marginally significant, it appears that the effect of rate on truthfulness judgments is not linear; rather, the pattern is an inverted-U function, with slower speech perceived as least credible. These results are consistent with the findings of Miller et al. (1976), who demonstrated that a faster speaker was perceived as more intelligent, knowledgeable, and objective than a slower speaker. Since Miller et al. used only two levels of speech rate, it is, of course, not possible to establish the effect of intermediate rates on credibility judgments in their study. It is relevant, however, that Brown and co-workers have reported a similar inverted-U relationship between manipulated speech rate and their "benevolence" dimension, on which the adjective pair sincere-insincere loads significantly (see Smith, Brown, Strong & Rencher, 1975). To the extent that ratings of sincerity correspond to this study's truthfulness measure, the two results are consistent.

The most plausible explanation of the Question \times Pitch interaction (Figure 1) is that listeners took question content into consideration in making their truthfulness judgments: When listening to a potentially "loaded" topic (quota question), raters were willing to call both low- and normal-pitched voices more truthful than high-pitched voices. For the less involving question (money), raters were willing to call only low voices more truthful. This interaction argues for a pitch threshold, above which deception is signaled to raters. Such a threshold would interact with response content, so that for an emotionally involving topic, it would be set higher than for a less involving topic. With an emotionally involving topic, some of the vocally reflected stress can be attributed to the topic; given a non-involving topic, the high-pitch responses are likely to be attributed to attempted deception. However, since only two questions were used, further research is needed to test this content-attribution hypothesis.

The truthfulness ratings reflect judgment

⁶ The corresponding analysis performed on the intelligibility ratings revealed a significant effect only for the rate manipulation, with fastest speech suffering greatest loss in intelligibility.

processes that give rise to attributions of a speaker's transient state. If listeners had not been told that certain speakers were lying, differences in the acoustically manipulated variables might have been seen as enduring vocal properties reflecting stable personal predispositions. Such properties have been referred to by the linguist Trager (1958) as the "voice set" and involve "the physiological and physical peculiarities resulting in the patterned identification of individuals as . . . persons of a certain sex, age, state of health, body build, rhythm state" (p. 4). The voice set, therefore, acts as a relatively permanent background against which transient vocal changes are superimposed. In the absence of situational factors (e.g., the possibility that the speaker was lying) that could explain the voice qualities produced by our acoustic manipulations, listeners would be likely to ascribe such qualities to the voice set. How such variables affect person perception and contribute to vocal stereotypes was explored in Experiment 2.

Experiment 2

Method

Stimulus materials. The quota and money tapes from Experiment 1 were used.

Procedure. Eleven college students, nine males and two females, were paid for their participation as raters. The procedure was essentially the same as that used in Experiment 1, with the following differences: Raters were instructed that the study's purpose was to investigate how listeners form impressions of speakers from the things they say as well as from the way they say them. Accordingly, raters were told to focus both on content and delivery when making their judgments.

The speaker of each recorded segment was rated on nine bipolar adjective pairs taken from the semantic differential (Osgood, Suci, & Tannenbaum, 1957); scales were chosen that had high loadings on one of Osgood et al.'s semantic space factors and relatively low loadings on the other two. The scales for the evaluation factor were sour-sweet, awful-nice, and bad-good. Scales for the potency factor were thin-thick, small-large, and weak-strong. Those for the activity factor were slow-fast, cold-hot, and passive-active.

A warning tone followed each recorded segment, signaling judges to begin making the nine ratings on 7-point scales. (The second adjective of each pair was scored as 7.) Instructions stressed that ratings should reflect the listener's impression of the speaker

and not agreement or disagreement with the content of the answer. Subjects were told that all speakers had answered the questions truthfully.

Results

Means of the nine scales were computed for each recorded segment, and the intercorrelation matrix was factor analyzed using principal factoring followed by a varimax rotation. A three-factor solution accounted for 84.5% of the total variance and roughly corresponded to the three dimensions of Osgood et al.'s (1957) semantic space. We chose scales loading greater than .60 (in absolute value) as representative of the respective factors. Using that cutoff, Factor 1 consisted of all three activity scales (slow-fast, cold-hot, and passive-active) as well as the strong-weak scale; it accounted for 54.4% of the variance. Factor 2 was a pure evaluation dimension (with only the three evaluative scales loading appreciably: sour-sweet, awful-nice, bad-good); it accounted for 20.2% of the variance. The third factor consisted of two potency scales (thin-thick, small-large) and an activity scale (slow-fast); it accounted for 9.9% of the variance.

Each rater's data were reduced to three factor scores weighting the original scales by the factor loadings; the factor scores were adjusted for covariates (as in Experiment 1) and entered into univariate analyses of variance of the same design as that used in Experiment 1.

For Factor 1, a significant rate effect, $F''(2, 32) = 10.49$, $p < .001$, and a nearly significant Rate \times Question interaction, $F''(2, 31) = 3.02$, $p < .10$, were obtained; raters' $F(2, 20) = 9.72$, $p < .01$; speakers' $F(2, 18) = 3.87$, $p < .05$. The configuration of means is shown in Figure 2. In both cases, slow speakers were perceived as less active. No significant main effects of interactions were found for Factor 2. For Factor 3, significant main effects were found for rate, $F''(2, 36) = 12.87$, $p < .001$, and pitch, $F''(2, 33) = 7.94$, $p < .01$. In both cases the relationship was monotonic, with increasing pitch and rate resulting in judgments of decreasing potency. No other significant effects were found.

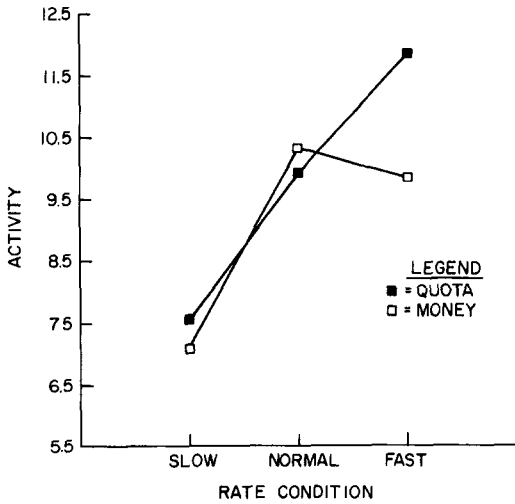


Figure 2. Averages for the activity factor plotted as a function of rate condition (slow, normal, fast) and question topic.

Discussion

These results extend the findings of Experiment 1 to judgments of more stable speaker dispositions. Men speaking in higher pitched voices were perceived as less potent (smaller, thinner, slower) and slow-speaking men were perceived as more passive (slower, colder, more passive, weaker) and more potent.

These findings are to some extent consistent with correlational evidence provided by Scherer, Koivumaki, and Rosenthal (1972). In their experiment, listeners rated taped segments, taken from a recorded play, on semantic differential scales similar to the ones used here, as well as on scales reflecting the segments' acoustic properties (e.g., bass-treble, soft-loud). Unlike the present study, raters judged the emotion portrayed and not their impression of the speaker, and a variety of listening conditions were used to degrade semantic content and prosodic features. Nevertheless, Scherer et al. found a marginally significant relationship between pitch rating (bass-treble) and potency ratings (strong-weak) paralleling the main effect reported above: Lower pitched speech was placed toward the "stronger" pole in the potency dimension of the emotional-meaning space. In contrast to our results, they also found a marginal correlation between articulation rate

(slow-fast) and potency ratings: Segments with slower speech were rated as "weaker." Not surprisingly, Scherer et al. also found that segments with faster speech were heavily loaded on the activity dimension.

Our findings are likewise consistent with the results of Brown et al. (1974). Although the methodological differences previously noted preclude direct comparison, Brown found that high fundamental frequency decreased competence ratings—a scale probably related to our potency dimension.

The Rate \times Question interaction (Figure 2) again suggests the influence that content exerts on raters' judgments: When speakers talked about quotas, ratings on the passive-active dimension were linear with manipulated rate; for the money question, they were not.

Experiment 3

Experiment 3 returned this work to the area of judgments of the speaker's affective state. Impressions of nervousness, emphaticness, seriousness, fluency, and persuasiveness illustrate how these acoustic variables serve to convey a speaker's self-presentation under conditions in which raters believe that answers are being given honestly. These state ratings (with the exception of persuasiveness) were chosen because Krauss, Geller, and Olson (Note 2) found significant correlations between them and truthfulness ratings in a previous study of deception interactions.

Method

Stimulus materials. The quota and money tapes from Experiment 1 were used.

Procedure. Ten college-student subjects, two males and eight females, were paid to rate all segments on five state variables: fluency, emphaticness, persuasiveness, nervousness, and seriousness. The rating scales spanned 7 points, with 1 indicating that the smallest amount of a variable was judged and 7 indicating that the largest amount was judged. For example, anchors of the nervousness scale were "Not at all nervous" and "Very nervous." Subjects were encouraged to adopt their own criteria for all ratings; no external standards were given. Instructions were virtually identical to those used in Experiment 2. Again, listeners were asked to take into consideration both the content of an answer and the manner in which it was delivered.

Table 2
Min-F Values for the Five State Variables

| State variable | Rate (R) effect | Pitch (P) effect | Question (Q) × R | Q × P |
|----------------|--------------------|---------------------|---------------------|--------|
| Persuasiveness | 3.93** | 2.94* | 3.54** | |
| Fluency | 6.03*** | | 4.07** | 2.73* |
| Emphaticness | 6.66*** | | 3.12** | 3.33** |
| Nervousness | 6.16*** | 5.90*** | | |
| Seriousness | | | | 4.60** |

* $p < .10$. ** $p < .05$. *** $p < .01$.

All other procedural details were identical to those of Experiment 2, except that this study was run at Bell Laboratories, using students from a number of colleges who were home on summer vacation.

Results

Segment means were computed for all state measures and analyzed using covariance analyses as described above. Table 2 summarizes the findings. Note that all variables with the exception of seriousness show main effects for the rate manipulation. These effects are all of the inverted-U type with the normal (unmanipulated) speakers judged most fluent, persuasive, and so forth, and slow speakers judged lowest on these scales. (Ratings of nervousness go in the direction opposite to the other three scales.)

Only nervousness yielded a significant main effect for pitch; the pitch manipulation was marginally significant for persuasiveness. Rated nervousness increased with higher pitch, whereas rated persuasiveness decreased.

In addition, ratings of emphaticness and seriousness showed significant Question × Pitch interactions. The shape of these interactions, shown in Figure 3, is quite similar to the corresponding effect on judged truthfulness (Figure 1). For the quota question, only the highest pitched group was "underrated" on emphaticness and seriousness, whereas for the money question, both the normal- and high-pitched groups were "underrated." The fluency measure showed a comparable interaction.

There were Question × Rate interactions for fluency, persuasiveness, and emphaticness. All interactions had a similar shape; for the

quota question only the slowest group suffered low ratings, while for the money question both the slow and fast groups received low ratings. The effect for seriousness judgments was marginally significant, but of the same shape.

Discussion

These results again demonstrate the effect acoustic variables have on person perception processes. Decreasing speech rate has a particularly deleterious effect on a speaker's perceived persuasiveness, fluency, and emphaticness. Similarly, increased pitch lowers ratings of persuasiveness and increases greatly the impression of nervousness.

Our findings also suggest that context plays a role in the attribution process, as evidenced by the question interactions. When a speaker

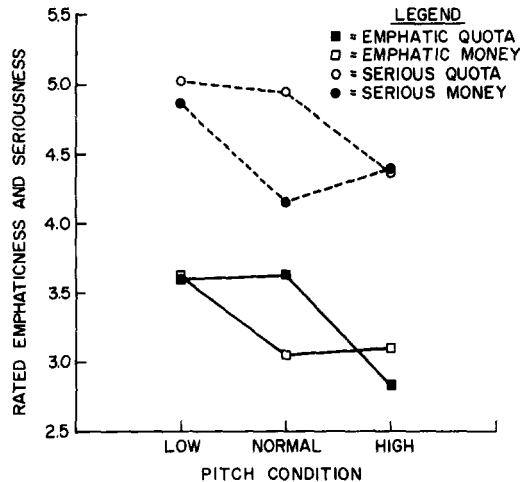


Figure 3. Average emphaticness and seriousness ratings plotted as a function of pitch condition (low, normal, high) and question topic.

answered the quota question, higher pitch levels could be discounted by attributing them to the stressfulness of the topic. Similarly, it may have been inappropriate on the money topic to talk too slowly, (perhaps because of the topic's relative simplicity) or too rapidly (perhaps because rapid speech is perceived as an attempt to be inappropriately serious or persuasive). On the other hand, rapid speech on the quota question might have been attributed to the speaker's conviction regarding his argument.

General Discussion

The three experiments taken as a whole provide clear evidence that acoustic properties of a message have considerable impact on judgments of a variety of state and trait variables. The impressions of high-pitched or slow-talking speakers seem particularly negative. For example, men with high-pitched voices are judged less truthful, less persuasive, weaker, and more nervous. Similarly, slow-talking men are judged to be less truthful, fluent, emphatic, serious, and persuasive, and more passive, although they are also seen as more potent.

By using a large number of speakers and factorially varying speech rate and pitch, we have a reasonable measure of confidence in the validity of our findings. However, since no female voices were used, we cannot generalize these results to women; it is conceivable that these same acoustic variables would produce different effects on perceptions of women speakers.

Message context, presumably mediated by the two question topics, was also demonstrated to influence attribution processes. However, since there was only one question of each type, what follows must remain somewhat speculative. The question interactions suggest an interpretation along the lines of Kelley's (1971) discounting principle. Kelley suggests that the more factors a situation contains—any one of which might plausibly have resulted in an observed outcome—the less likely is any one factor to be perceived as the cause of that outcome. With fewer possible causes present, the cause-to-effect attribution is more compelling. We hoped that

our acoustic manipulations would be potent enough to affect speaker state or trait attributions. However, it is evident that listeners may have, at least partially, attributed our acoustic alterations to the question topic. The quota question called for an answer that was both complex and emotionally involving for our college-student speakers and raters. As Table 1 shows, before any manipulations were done, the quota answers were longer and slower (suggesting greater cognitive complexity; see Goldman Eisler, 1968) and higher pitched (suggesting greater stressfulness; see Hecker et al., 1968). Because of this, slower and higher pitched answers might have been perceived as more appropriate for quota responses than for money responses.⁷ But even when premanipulation speech rate and pitch were covaried, the question interactions remained, supporting the discounting principle. For example, higher pitch levels seemed to be discounted on ratings of truthfulness, fluency, emphaticness, and seriousness for speakers answering the quota, but not the money, question.

It should be noted that manipulating fundamental frequency by multiplying by a scale factor as we did has the effect of multiplying the *variance* of the fundamental frequency by the square of the scale factor. Thus, high-pitched segments were both high-pitched and high pitch-variance segments, and vice versa. Thus, we cannot rule out the interpretation that the pitch effects observed could be pitch-variance effects. However, this interpretation seems unlikely considering the findings of Brown et al. (1974), who did the appropriate factorial experiment and found that increased mean fundamental frequency lowered judgments of speakers' competence and benevolence, while decreased variance also lowered these ratings. Thus, it seems that average pitch and pitch variance affect judgments on

⁷ Subsequently, in connection with another study, we had 12 undergraduates rate a long list of potential interview topics (including the two used in the present study) as to how stressful and how complex they would be for a typical Columbia undergraduate to discuss. Subjects judged the quota question to be significantly more stressful and more complex than the money question.

a wide range of scales in opposite ways. In the present study, in which both pitch and pitch variance were positively correlated, the pitch-correlated variance could only have attenuated the effects of average pitch.

Given that the experimental manipulations used here affected the perception of speakers' personality and state, it remains to be explained why these particular data patterns were observed. Miller et al. (1976), for example, conclude that the effect of speech rate on message persuasiveness is mediated by the effect that variable has on the perception of a speaker's credibility. This, they assert, is a "less rationalistic view" of attitude change than other interpretations (e.g., change mediated by comprehension effects or counterargument disruption—two hypotheses their experiments ruled out).

Certainly such a conclusion would be justified if it were the case that variations in voice quality bore no relation to the actual internal state or predisposition of the speaker. However, there is considerable evidence that stressful situations do produce discernible changes in voice quality (Fairbanks, 1940; Hecker et al., 1968; Williams & Stevens, 1969, 1972), and it seems reasonable to assume that listeners in the present study used such variations appropriately to infer a speaker's state from the quality of his speech. For example, Hecker et al. demonstrated that task-induced stress raised the fundamental frequency of those speakers who did not talk more softly under stress. In our experiment, attributions of increased nervousness to higher-pitched speakers are quite "rationalistic," given the similarity of our pitch manipulations to the effect of real-life stress. Similarly, Streeter et al. (1977) demonstrated that pitch increments accompany deceptive responses; listeners' truthfulness judgments in Experiment 1 appropriately reflect this relationship.

The effects of speech rate can be similarly interpreted in light of Goldman Eisler's (1968) finding that rate and the cognitive complexity of the topic were negatively related. Listeners may have assumed that lying increases speakers' cognitive load, resulting in slower rates. Unpublished speech rate data

from the Streeter et al. (1977) study lend plausibility to this argument. A marginally significant interaction ($p < .08$) indicated that subjects did, in fact, speak more slowly when lying than when telling the truth, provided they had been given instructions engaging their motivation to lie effectively. However, speakers not receiving such instructions spoke more rapidly when lying.

Judgments of what is or is not "rationalistic" are probably to a large extent matters of personal preference, but it does seem to us that a listener would be ill-advised to ignore reliable information concerning a speaker's internal state in evaluating the speaker's message, especially when the internal state seems incongruent with the situation or with the message's content.

Furthermore, the significant Question \times Manipulation interactions for ratings of state variables suggest some qualifications on the findings of Miller et al. (1976). Fast speakers are not always more persuasive; talking too quickly in response to the money question produced lower persuasiveness ratings than did responses at a normal rate. Apparently, listeners take more into account than meets the ear—at least, more than simply the acoustic data.

Evidence of veridicality for vocally based attributions of enduring personality traits is less firm. We have found no reliable data to indicate that fast talkers actually are more active people or that higher pitched men are weaker than their lower pitched counterparts. Apart from studies of psychiatric patients, much of the work in this area deals with the traits of introversion-extraversion and dominance. Mallory and Miller (1958) found small but significant negative correlations between *judged* pitch and rate and the dominance scale of the Bernreuter Personality Inventory. While these findings support our results on judged potency along the pitch dimension (Experiment 2), it is possible that Mallory and Miller's acoustic judgments are either inaccurate or subject to biasing effects from other sources, since raters judged speakers in a live situation. Furthermore, our results for judged potency on the speech-rate variable appear opposite to Mallory and

Miller's. A more recent study (Ramsay, 1966) found that the speech of subjects classified as extraverts on the Eysenck Personality Inventory had longer unbroken phonation times and shorter silences than those of introverts across a variety of speaking tasks; no data on speech rates were presented.

The acoustic stimulus, of course, contains more information than average pitch and rate. In addition, there is sequential information (provided by intonation contours and duration pattern), loudness, and variability of both pitch and loudness over time. There is also voice quality information (e.g. "breathy" or "raspy" voices) that may not be so readily specified in terms of physical parameters. All of these factors can be expected to enter into the person perception process via stereotypes with larger or smaller kernels of truth.

On none of the measures we examined was the Rate \times Pitch interaction statistically significant. The median F interaction was 1.12—close to its expected value under the null hypothesis. This absence of interaction argues for an additive model, in which pitch and rate exert independent effects on listeners' judgments. It remains to be seen whether support for such a model will continue as the role of additional vocal factors is explored.

Reference Notes

1. Nakatani, L. H. *SYNLOG: An interactive system for manipulating speech*. Unpublished manuscript, Bell Laboratories, Murray Hill, N.J., 1976.
2. 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.

References

- Allport, G. W., & Cantril, H. Judging personality from voice. *Journal of Social Psychology*, 1934, 5, 37-55.
- Atal, B. S., & Hanauer, S. L. Speech analysis and synthesis by linear prediction of the speech wave. *Journal of the Acoustical Society of America*, 1971, 50, 637-655.
- Black, J. W. Relationships among fundamental frequency, vocal sound pressure, and rate of speaking. *Language and Speech*, 1961, 4, 196-199.
- Brown, B. L., Strong, W. J., & Rencher, A. C. Perceptions of personality from speech: Effects of manipulations of acoustical parameters. *Journal of the Acoustical Society of America*, 1973, 54, 29-35.
- Brown, B. L., Strong, W. J., & Rencher, A. C. Fifty-four voices from two: The effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech. *Journal of the Acoustical Society of America*, 1974, 55, 313-318.
- Fairbanks, G. Recent experimental investigations of vocal pitch in speech. *Journal of the Acoustical Society of America*, 1940, 11, 457-466.
- Goldman Eisler, F. *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press, 1968.
- Hanley, T. D. An analysis of vocal frequency and duration characteristics of selected samples of speech from three American dialect regions. *Speech Monographs*, 1951, 18, 78-93.
- 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.
- Kelley, H. H. *Attribution in social interaction*. Morristown, N.J.: General Learning Press, 1971.
- Kramer, E. Judgment of personal characteristics and emotions from nonverbal properties of speech. *Psychological Bulletin*, 1963, 60, 408-420.
- Lay, C. H., & Burron, B. F. Perception of the personality of the hesitant speaker. *Perceptual and Motor Skills*, 1968, 26, 951-956.
- Mallory, E. B., & Miller, V. R. A possible basis for the association of voice characteristics and personality traits. *Speech Monographs*, 1958, 25, 255-260.
- Miller, G. R., & Hewgill, M. A. The effect of variations in nonfluency on audience ratings of source credibility. *Quarterly Journal of Speech*, 1964, 50, 36-44.
- Miller, N., Maruyama, G., Beaver, R. J., & Valone, K. Speed of speech and persuasion. *Journal of Personality and Social Psychology*, 1976, 34, 615-624.
- Mysak, E. Pitch and duration characteristics of older males. *Journal of Speech and Hearing Research*, 1959, 2, 46-54.
- Nakatani, L. H. Computer-aided signal handling for speech research. *Journal of the Acoustical Society of America*, 1977, 61, 1057-1062.
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.
- Peterson, G. E., & Barney, H. L. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 1952, 24, 175-184.
- Ramsay, R. W. Personality and speech. *Journal of Personality and Social Psychology*, 1966, 4, 116-118.
- Scherer, K. R. Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences. In S. Weitz (Ed.), *Nonverbal communication*. New York: Oxford University Press, 1974.

- Scherer, K. R., Koivumaki, J., & Rosenthal, R. Minimal cues in the vocal communication of affect: Judging emotions from content-masked speech. *Journal of Psycholinguistic Research*, 1972, 1, 269-285.
- Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. Effects of speech rate on personality perception. *Language and Speech*, 1975, 18, 145-152.
- Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., & Apple, W. Pitch changes during attempted deception. *Journal of Personality and Social Psychology*, 1977, 35, 345-350.
- Terango, L. Pitch and duration characteristics of the oral reading of males on a masculinity-femininity dimension. *Journal of Speech and Hearing Research*, 1966, 9, 590-595.
- Trager, G. L. Paralanguage: A first approximation. *Studies in Linguistics*, 1958, 13, 1-12.
- Williams, C. E., & Stevens, K. N. On determining the emotional state of pilots during flight. An exploratory study. *Aerospace Medicine*, 1969, 40, 1369-1372.
- Williams, C. E., & Stevens, K. N. Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America*, 1972, 52, 1238-1250.
- Winer, B. J., *Statistical principles in experimental design* (2nd ed.), New York: McGraw-Hill, 1971.

Received January 18, 1978 ■

Corrections to Koretzky, Kohn, and Jeger

In the article "Cross-Situational Consistency Among Problem Adolescents: An Application of the Two-Factor Model" by Martin B. Koretzky, Martin Kohn, and Abraham M. Jeger (*Journal of Personality and Social Psychology*, 1978, Vol. 36, No. 9, pp. 1054-1059), there are errors in two correlations reported in the first paragraph on page 1058. The Factor I and Factor II scores for the Koretzky (1976) study cited there should be reversed. The correct correlation for Factor I is .40, and for Factor II it is .60. Thus, the sentence should read, "Consistency correlations between classroom and residence settings were even stronger in this experiment for Factor II ($r = .60$) and were also respectable for Factor I ($r = .40$)."

Martin B. Koretzky's affiliation was erroneously given as the Veterans Administration Hospital, Bronx, New York. His affiliation at the time of the original research was the State University of New York at Stony Brook.