PERCEPTION OF SPEECH RATE AS A FUNCTION OF VOCAL INTENSITY AND FREQUENCY*

STANLEY FELDSTEIN AND RONALD N. BOND
University of Maryland, Baltimore County

The study was designed to examine the possibility that vocal frequency and vocal intensity influence the perception of speech rate. One 30-second segment of spontaneous speech was used to produce nine stimulus segments that factorially varied three levels of vocal frequency and three levels of vocal intensity but were identical in speech rate. The segments were recorded backwards in pairs such that the first member of each pair was the original segment and the second was the altered segment. Eighty-eight judges were then asked to compare the speech rate of the second member of each pair with that of the first in terms of a seven-point scale that varied from "much slower" to "much faster." The results indicate that vocal frequency and intensity both separately and jointly influenced the perception of speech rate.

A good many studies (see Grosjean and Lane, 1974, for a brief review) have cumulatively demonstrated that the perception of speech rate is a function of three components of the speech stream, i.e., articulation rate and the number and duration of pauses. The purpose of the study reported here was to examine the possibility that the perception of speech rate is also influenced by the vocal intensity and frequency of speech sounds. There is evidence (e.g., Black, 1961) that in natural speech, rate, pitch, and loudness tend to be directly related to each other. Thus, slow speech is likely to be uttered more softly and at a lower pitch than rapid speech. Does this covariation in the production of speech occur in the perception of speech? No previously reported study appears to have explored this possibility although many studies have involved, in one way or another, the perception of speech rate as an important variable.

For example, Scherwitz, Berton, and Leventhal (1977) tried to determine which cues helped interviewers to classify the susceptibility of interviewees to coronary heart disease. The study had judges rate the interviewees' responses in terms of a number of variables, two of which were "speed of speaking" and "speed of answering." Another variable rated by the judges was "voice emphasis," to some extent a measure of loudness. A similar study was conducted by Schucker and Jacobs (1977), who measured a group of lexical and stylistic variables and had judges estimate speed of speech and voice volume. By comparison with the other measures, voice emphasis and speech rate in the former

* *The report is based, in part, upon a paper presented at the annual meeting of the Eastern Psychological Association, New York, April, 1981. The authors are grateful to Dr. Marilyn Wang for her notion of "presentation" order, and to Drs. Jonathan Finkelstein and Thomas Blass and an anonymous reviewer for their concern with explanations. They are also indebted to the Statistics Center of the University of Maryland Baltimore County and the Computer Science Center of the University of Maryland College Park for their computer support.*

study and voice volume and speech rate in the latter study accounted for most of the variance of the susceptibility ratings.

Other studies have investigated the relation of speech rate to various types of interpersonal attributions. These studies have measured and manipulated speech rate and most have, at the same time, independently manipulated pitch or fundamental frequency. Very few, however, have also controlled for loudness or vocal intensity. Scherer (1974), for instance, obtained ratings of simple, synthesized tone sequences in which tempo, pitch, and intensity were independently varied and found that they were associated with different attributions of emotional qualities. It is not clear, however, that the results are generalizable to spontaneous speech.

Two experiments by Miller, Maruyama, Beaber, and Valone (1976) investigated the role of speech rate in judgments of persuasiveness. The authors manipulated rate of speech and found that the more rapid rates were perceived as more persuasive. However, neither vocal intensity nor frequency was taken into account and it is possible, as Apple, Streeter, and Krauss (1979) have pointed out, that the listeners responded to intensity and frequency as well as to speech rate. Brown, Strong, and Rencher (1972, 1973, 1974) conducted three experiments to explore the effects of speech rate, fundamental frequency, and the variance of fundamental frequency upon personality attributions. In each of the studies, they found rate to be the best predictor of the personality ratings. In order to extend the generality of this finding, they and a colleague (Smith, Brown, Strong, and Rencher, 1975) conducted still another experiment in which they used more voices than had been used previously and nine rates of speech. Analyses of the results yielded a curvilinear relation between speech rate and a factor they called "benevolence" and a strong linear relation between speech rate and another factor called "competence." In other words, their listeners perceived both faster and slower rates as indicative of less benevolence than a "normal" rate. But they perceived the faster rates as indicative of greater competence than the slower rates. Although Brown and his associates used synthesized speech stimuli to produce variations in speech rate, fundamental frequency, and the variance of fundamental frequency (intonation), they controlled for vocal intensity and all other vocal characteristics of the voices used in the experiments. Apple, Streeter, and Krauss (1979), who investigated the influence of rate and pitch on interpersonal attributions, also controlled for vocal intensity by means of a random assignment of their stimulus speakers.

These studies that called for the attribution of interpersonal characteristics and emotions did not require their listeners to estimate rate of speech explicitly. It was, however, the listeners' perception of speech rate that presumably accounted at least partly for their attributions. If the results of such studies can be generalized beyond the laboratory, and if interpersonal perception is thought to play a significant role in the process of daily living (Feldstein, in press), then the question of whether the perception of the rate of a speech sample is influenced by its vocal intensity and frequency is clearly important.

METHOD

The general procedure used in the experiment was to have individuals judge the speech rate of speech stimuli that were, in fact, identical in rate but systematically different in terms of vocal frequency and intensity.

Instruments

Vocal intensity changes. Alterations in vocal intensity were effected by electronically sending a speech signal from the playback mechanism of one audio-cassette recorder (a Sony TC-520CS in this experiment) to the recording unit of another cassette deck (a Marantz 5240) of which the recording level was varied systematically. Because of the variability in the intensity level of a speech signal, a 400-Hz tone was used to calibrate the recording level of the Marantz deck such that the speech stimuli used in the experiment were recorded at three approximately equally-spaced levels of intensity.

Vocal frequency changes. The device used to alter vocal frequency does so by changing both fundamental and higher formant frequencies, and permits shifts in either direction that are no greater than 50% of the original frequency. The device consists of two variable delays. Each delay stores a sample of speech (approximately $^1/_{10}$ of a second in duration) at one rate and outputs of it at a different rate. The two delays alternate in order to provide continuous output. That is, as one delay is in the process of storing its sample, the other is in the process of outputting its sample. The critical feature of the device is its ability to vary frequency without altering either intensity or speech rate. The device accepts a speech signal from one recorder and sends the transformed signal to another recorder.

It should be noted that the way in which the device alters (fundamental) frequency also changes the power spectrum (formant frequencies) such that increasing the frequency yields speech, or a vocal signal, that sounds as if it were produced by a shortened vocal tract, whereas decreasing the frequency has the effect of a lengthened vocal tract.

Rate perception scale. The only dependent variable in the study was perceived rate of speech, as measured by a seven-point, bipolar, "rate perception scale." The scale ranges from −3 ("much slower") to +3 ("much faster"), with −2 and +2 being "moderately slower" and "moderately faster," respectively, −1 and +1 being "slightly slower" and "slightly faster," respectively, and 0 indicating "no difference." It was chosen because: (a) a seven-point scale seemed to offer an optimal number of choices; (b) the use of positive and negative numbers seemed to clarify the ratings of faster and slower, respectively; and (c) "0" seemed to most obviously represent "no difference."

Stimulus material

One 30-second segment of speech excised from a woman's description of a TAT card (Murray, 1938) was used to produce the speech stimuli. The segment contained 72 words, a rate that falls comfortably within a "normal," or moderate, range (Allen, Anderson, and Hough, 1968). In order to control for the effects of content, the segment was rerecorded backwards. It was then transformed, by means of two audio-cassette

recorders and the device for altering frequency, to produce a set of nine stimuli that factorially varied three levels of frequency and three levels of intensity. That is, in addition to the intensity and frequency levels of the original, unmanipulated segment, a lower and higher level of each variable was produced. The original frequency (F) of the segment was decreased by about 13% for the lower level (F−), and increased by about 14% for the higher level (F+). The three levels of intensity, as heard by the judges, were approximately 65 dB (I−), 70 dB (I), and 75 dB (I+) as measured on an A-weighted filter.[1] The changes in each variable, and the increases and decreases sounded acoustically "natural."

To increase reliability and to control for within-judge order effects, the set of nine stimuli (F−I−, FI−, F+I−, F−I, FI, F+I, F−I+, FI+, F+I+) was recorded twice on an audiocassette such that the order of the second set (Stimulus Set 2) was the reverse of that of the first set (Stimulus Set 1). In addition, each of the 18 stimuli was preceded by a recording of the original segment, which served as the standard with which the transformed segment was to be compared. To counterbalance the order of Sets 1 and 2, the two sets were rerecorded in reverse order (i.e., Set 2 followed by Set 1) on another cassette.

Judges and procedure

Thirty-one male and 57 female undergraduate university students volunteered to serve as judges in the study as a way of fulfilling a course requirement. The judges were assembled into groups of from 3 to 10. The seating arrangement consisted of two rows of five chairs each, with approximately two feet (0.61 meters) between any two chairs within each row. Each row had its own Sony speaker positioned approximately 6 feet (1.83 meters) from the center chair of the row. The two rows were on opposite sides of the room arranged such that they faced each other. It was intended, and presumed, that the judges in a given row attended primarily to the speaker in front of their row. The judges were told that they were to "listen to 18 pairs of speech samples and to rate the second sample of each pair as it compares with the first in terms of speech rate" and that they would have 15 seconds to rate each pair. Half of the judges listened to one of the audiocassettes and half listened to the other.

RESULTS

The 1584 ratings were initially subjected to a univariate, split plot ANOVA with five independent variables. Inasmuch as the analysis yielded no significant main effect of

[1] *The "A" of the Af scale on the sound-level meter filters indicates that an A weighting network was used to process the incoming speech signal. The weighting functions for the A, B, and C scales available on such filters have been established by the American Standards Institute, and the function for the A scale increasingly attenuates a signal from 1000 Hz to 20 Hz in order to provide an approximation of human auditory perception. The "f" of the Af scales indicates the use of a fast meter response, which has to do with the value of the filter time constant.*

gender, or meaningful interactions involving it, and since the numbers of males and females were quite different, the ANOVA was recomputed using only the remaining four factors: counterbalanced order, stimulus sets, vocal intensity, and vocal frequency.

Stimulus sets, which provided the control for within-judges order effects, yielded a significant main effect and a two-way and four-way interaction effect with intensity alone, and with intensity, frequency, and counterbalanced order. Further analyses indicated that these effects were primarily a function of the order in which the sets were presented to the judges. For some reason or reasons, which may have included individual differences and reactions to backward speech, the judges who listened to the first stimulus set first and those who listened to the second stimulus set first rated the two sets quite differently: the average rating of the first set was 0.45, whereas that of the second set was 0.12. However, the two sets were rated similarly when they formed the *last* nine pairs in the sequence of stimuli presented to each group of judges: the first set yielded an average rating of 0.21 and the second set an average rating of 0.25. Thus, separate ANOVAs of the two "presentation" orders were computed.

As expected, the analysis of only those ratings provided by the judges in their first encounter with the two stimulus sets (the first "presentation" order) yielded a significant main effect of counterbalanced order. However, both presentation orders yielded, as did the full ANOVA, significant main effects of intensity and frequency and a significant interaction of intensity and frequency. Thus, the findings suggest that vocal frequency and vocal intensity separately and jointly influence the perception of speech rate. The results of the full ANOVA (see Table 1 for means and standard deviations) indicate that both intensity ($F_{(2,172)} = 23.773$, $p < 0.001$) and frequency ($F_{(2,172)} = 102.274$, $p < 0.001$) were positively related to perceived speech rate, but that the impact of frequency upon the perception of speech rate was greatest at the lower level of intensity and the impact of intensity, although less, was greatest at the lower level of frequency ($F_{(4,344)} = 7.439$, $p < 0.001$). Omega squares were computed to assess the strength of these relationships and indicate that vocal frequency accounted for 14.9% of the variance associated with perceived speech rate, while vocal intensity accounted for 2.4%, and the joint influence of the two factors accounted for only 0.7%.

Discussion

The results seem to indicate that the perception of speech rate is influenced not only by articulation rate and the number and duration of pauses, but also by the pitch and loudness of the perceived speech. Although there are no immediately obvious reasons why pitch and loudness should covary with perceived speech rate, there are at least three tenable alternative explanations of the obtained results. The findings may be a function of experience with the outcome of speech production processes. That is, the perception of covariation among pitch, loudness, and speech rate may be learned from almost always hearing such covariation in naturally occurring (produced) speech. A second explanation is akin to the halo effect of expecting that someone who is adept at one thing is adept at most things. It may be, in other words, that when a change in one charac-

TABLE 1

Descriptive Statistics of Perceived Speech Rate at the Lower, Unmanipulated, and Higher Levels of Vocal Intensity and Vocal Frequency

Intensity		Frequency F–	Frequency F	Frequency F+	Over Frequency
I–	*M*	–0.915	0.199	0.813	0.032
	SD	1.127	1.051	1.125	
I	*M*	–0.631	0.222	0.881	0.385
	SD	1.190	1.036	1.132	
I+	*M*	0.051	0.682	1.011	0.864
	SD	1.243	1.108	1.173	
Over Intensity	*M*	–0.498	0.367	0.902	

Note. The number of observations for each table cell is 176. Judgments of speech rate were made in terms of a 7-point, bipolar scale.

teristic is perceived, it is expected – and thus perceived – that all other characteristics of the speech change in a similar direction.

A third methodological explanation is that the results are a function of having only asked the judges for their perception of speech rate. It might be argued that the judges attributed whatever changes they perceived to the only characteristic they were permitted to rate. The implication here is either that they could not specifically identify the nature of any particular change and therefore assumed it to be a change in the characteristic they were asked to rate, or that they decided to report any change they detected in terms of the only scale available to them even though they could identify which feature of the speech had changed. The argument does not seem to be a strong one, but the design of the study cannot eliminate it as a possibility.

A number of other aspects of the study deserve some elaboration. As was mentioned earlier, the device that was used to alter vocal frequency yields speech signals whose pitch differences sound as if they were produced by changing the length of the vocal tract. Thus, the findings of the study with respect to the influence of frequency on perceived speech cannot strictly be attributed to fundamental frequency alone.

It can be argued that the use of backward speech was unnecessary as a way of controlling for semantic content because all the speech stimuli were produced from the

TABLE 2

Product-Moment Correlation Coefficients Yielded by Comparisons of Stimulus Sets (SS) Across Two Counterbalanced Orders (CO)

		CO 1	
		SS 1	SS 2
	SS 2	0.85	0.87
CO 2			
	SS 1	0.92	0.94

Note. The *df* for each of the coefficients is 7. In CO 2, SS 2 was presented to the judges first.

same 30-second speech segment. Thus, the content was the same throughout the study and allowed for the use of the traditional "content standard" control technique. It seems quite conceivable, however, that the content of the segment as well as the gender of the speaker could have interacted with the changes in frequency and intensity in their influence on the perception of speech rate. The design of the present study does not permit the detection of such an effect. The results, therefore, that might have been yielded by use of the content standard technique may not, in the present case, have been generalizable much beyond the content or type of content of the segment and the gender of its speaker. Playing English speech backwards effectively masks its content and, to some extent, speaker sex, and sounds very much like Scandinavian speech rather than like nonspeech. It may well be, therefore, that the generalizability of the present results is not especially more limited than it might have been with the use of standard content.

The notion of "presentation" order is worth further comment. Few studies involving judgments of speech stimuli (including those cited in this report) appear to have been concerned with the possible effects of the order in which the stimuli are presented to the judges. In the present study, presentation order had an effect upon the absolute values of the stimulus ratings from one Stimulus Set to another either within or across the two counterbalanced orders. Recall that each Stimulus Set consisted of nine manipulations and that the order of the manipulations (stimuli) in Set 2 was the reverse of that of Set 1. Nevertheless, a comparison of the two Sets in the first counterbalanced order (Set 1 followed by Set 2) yielded a product-moment r of 0.90 and in the second counterbalanced order (Set 2 followed by Set 1), an r of 0.87. Comparisons of the Sets across the two orders yielded the coefficients listed in Table 2. The magnitude of the

coefficients suggests that regardless of the sequence of the manipulations and of the order in which they are presented to the judges, their relative ratings with respect to each other tend to remain similar.

Finally, given that the present findings were obtained with the use of an actual speech rate that could be considered moderate, it seems fair to ask whether they are generalizable to faster and slower rates of speech. Do frequency and intensity interact with each other in their influence on perceived rate only at the actual rate examined or at other actual rates? It also seems possible that actual rate may interact with frequency and intensity in their influence on perceived rate.

Apart from these issues, and regardless of which explanation is used to account for the findings, it is clear that they seriously question the viability of using the perception of speech rate as a dependent or mediating variable in research without controlling both intensity and frequency, and they suggest that the results of those studies that have so used rate perception must be reevaluated.

REFERENCES

ALLEN, R.R., ANDERSON, S. and HOUGH, J. (1968). *Speech in American Society* (Columbus, Ohio).

APPLE, W., STREETER, L.A. and KRAUSS, R.M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, **37**, 715-727.

BLACK, J.W. (1961). Relationships among fundamental frequency, vocal sound pressure, and rate of speaking. *Language and Speech*, **4**, 196-199.

BROWN, B.L., STRONG, W.J. and RENCHER, A.C. (1972). Manipulations of vocal qualities by speech synthesis: A new way to study person perception. *Proceedings of the 80th Annual Convention of the American Psychological Association*, **7**, 197-198.

BROWN, B.L., STRONG, W. and RENCHER, A.C. (1973). Perceptions of personality from speech: Effects of manipulations of acoustical parameters. *Journal of the Acoustical Society of America*, **54**, 29-35.

BROWN, B.L., STRONG, W.J. and RENCHER, A.C. (1974). 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*, **55**, 313-318.

FELDSTEIN, S. (In press). Impression formation in dyads: The temporal dimension. In M. Davis (ed.), *Interaction Rhythms: Periodicity in Communicative Behavior* (New York).

GROSJEAN, F. and LANE, H. (1974). Effects of two temporal variables on the listener's perception of reading rate. *Journal of Experimental Psychology*, **102**, 893-896.

MILLER, N., MARUYAMA, G., BEABER, R.J. and VALONE, K. (1976). Speed of speech and persuasion. *Journal of Personality and Social Psychology*, **34**, 615-624.

MURRAY, H.A. (1938). *Explorations in Personality* (New York).

SCHERER, K.R. (1974). Acoustic concomitants of emotional dimensions: Judging affects from synthesized tone sequences. In S. Weitz (ed.), *Nonverbal Communication* (New York), pp. 105-111.

SCHERWITZ, L., BERTON, B.S. and LEVENTHAL, H. (1977). Type A assessment and interaction in the behavior pattern interview. *Psychosomatic Medicine*, **39**, 229-240.

SCHUCKER, B. and JACOBS, D.R. (1977). Assessment of behavioral risk for coronary disease by voice characteristics. *Psychosomatic Medicine*, **39**, 219-228.

SMITH, B.L., BROWN, B.L., STRONG, W.J. and RENCHER, A.C. (1975). Effects of speech rate on personality perception. *Language and Speech*, **18**, 145-152.