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*

Bruce L. Brown, William J. Strong, and Alvin C. Rencher

Language Research Center, Brigham Young University, Provo, Utah 84601
(Received 4 June 1973; revised 10 September 1973)

Utterances of two adults males were analyzed and synthesized by a fast Fourier Transforms method. Each of the two voices was synthesized in each of the twenty-seven combinations of three levels each of rate, mean *FØ*, and variance of *FØ* (a total of fifty-four "voices" generated from two). The effects of the rate, mean *FØ*, and variance of *FØ* manipulations, the interactive effects of rate and variance of *FØ*, and the effects due to speaker were all statistically significant predictors of personality ratings given the voices. They accounted for 86%, 4%, 3%, 2%, and 1% of the variance, respectively, in competence ratings and 48%, 1%, 6%, 1%, and 8% of the variance, respectively, in benevolence ratings. Increased speaking rate was found to decrease the benevolence ratings, and decreased rate was found to decrease competence ratings. Decreased variance of *FØ* was found to decrease the ratings on both competence and benevolence. Increased mean *FØ* in these male voices was also found to decrease competence and benevolence ratings.

Subject Classification: 70.30.

INTRODUCTION

There have been many studies of perception of persons from the nonverbal properties of their speech (for a review see Kramer, 1963[2]). Recently, attempts have been made to study this problem experimentally by utilizing techniques of speech synthesis by computer.[3–5] In these studies it was found that slowing the voices caused them to be rated less "competent" and speeding the voices caused them to be rated less "benevolent." These rate manipulations had the same effect upon every voice to which they were applied. There was also a significant effect for increased variance of fundamental frequency (*FØ*) to cause voices to be rated more benevolent and decreased variance of *FØ* to cause them to be rated less benevolent. However, this trend was not consistent over speakers. In these earlier studies only one acoustic dimension at a time was manipulated. The purpose of the study reported in this paper is to examine the interactive effects of rate, mean *FØ*, and variance of *FØ* when all three are manipulated in varying combinations for a given speaker. Since in the earlier studies the effects of variance-of-*FØ* manipulations were not consistent over various speakers (even though the overall effects were statistically significant), it would be expected that an interaction exists between variance of *FØ* and some other acoustic characteristics. One of the questions of interest in this study is the extent and nature of interactive effects between variance of *FØ* and the other two variables of this study, rate and mean *FØ* level, in the ways they alter the personality ratings given to various voices.

I. METHOD

The voices of two adult male college teachers speaking the sentence "We were away a year ago," were analyzed and synthesized by an automatic speech analysis–synthesis scheme. Each voice was synthesized in 27 forms (all possible combinations of three values of each of mean *FØ* level, rate, and variance of *FØ*). The three values of mean *FØ* level were: normal, decreased to 0.7 times normal (in hertz) and increased to 1.8 times normal. For rate they were: normal, decreased to 0.5 times normal (in msec duration) and increased to 1.5 times normal. For variance of *FØ* they were: normal, decreased to 0.2 times normal and increased to 1.8 times normal. These 54 synthetic voices, each repeating the sentence three times, were recorded on a testing tape in random order, with the first two voices in the sequence repeated later in the tape (in order to control for and evaluate practice effects). The tape was played to a group of 37 male and female judges who rated the voices on the following 15 adjectives with their paired opposites: intelligent, ambitious, polite, active, confident, happy, just, likeable, kind, sincere, dependable, religious, good-looking, sociable, and strong.

II. SPEECH ANALYSIS–SYNTHESIS PROCEDURES

The voices were manipulated in rate, mean *FØ* level, and variance of *FØ* by means of an automatic computer-based speech analysis–synthesis scheme.[6–7] In the analysis, new parameters were calculated each 10 msec. A spectrum was computed by means of a 512-point FFT operating on a speech segment windowed by a Hamming window of 40-msec duration. Fundamental frequency was measured with the cepstrum method. A smoothed spectrum was computed from the 26 low-order cepstral coefficients and peak-picked to determine five formant frequencies and amplitudes. Some smoothing was done on the fundamental frequency and formant frequency contours to eliminate gross discontinuities. All of the analysis and smoothing operations were completely automatic.

The synthesis was accomplished with a five-pole parallel synthesizer which was simulated on a DEC PDP-

15 computer. Eleven parameters ($F\emptyset$, five formant frequencies, and five formant amplitudes) were input to the synthesizer at 10-msec intervals and the input parameters were linearly interpolated as needed in the synthesizer. The speech was output by means of a 12-bit digital-to analog converter and recorded on audio tape for testing.

III. RESULTS

The averages of judges' ratings of the 54 voices were factor analyzed by the principle axes method with a varimax rotation. The resulting factor pattern (Fig. 1) was almost identical with the ones from the earlier studies, and again Factors I and II were labeled "competence" and "benevolence," with pseudo-eigenvalues (on rotated factors) of 7.086 and 6.702, respectively.

Figures 2–4 are plottings of the factor scores of ratings given the 54 altered forms of the two synthesized voices, grouped in such a way as to show the effects of increasing rate (I) and decreasing rate (D) on the factor score positions of voices at each level of variance of $F\emptyset$. Each of the three figures contains this display for a different one of the levels of mean $F\emptyset$. Figures 5–7 give the same plottings, but grouped in such a way as to show the effects of increasing (I) and decreasing (D) variance of $F\emptyset$ at each level of rate, with each of the three figures displaying this for one of the levels of mean $F\emptyset$. The normal voice (no manipulations) for each of the two speakers are shown in Figs. 3 and 6 (which both show effects of manipulations at normal mean $F\emptyset$ level). The unmanipulated voice of Speaker 1 is rated higher than the unmanipulated voice of Speaker 2 on both competence and benevolence.

The most obvious things to notice from a comparison of Figs. 2–4 with Figs. 5–7 are that rate manipulations have much greater effects on the factor scores of voice ratings and much more consistent ones than variance of $F\emptyset$ manipulation. In every case decreasing the rate decreased the competence rating and increasing rate decreased the benevolence rating. In most cases decreased rate also decreased the benevolence rating, but in two of the 18 slowed voices it increased benevolence ratings while decreasing competence (Fig. 4). These were both Speaker 2 under condition of increased mean $F\emptyset$, in which cases the corresponding normal rate forms of these voices had very low benevolence ratings to begin with, leaving little room for a decrease in benevolence. As a general rule, it appears that slowing the voice decreases competence markedly and benevolence slightly, unless the voice is already low in benevolence rating, in which case competence still decreases markedly, but benevolence may increase. A similar secondary effect may be noticed with respect to increasing the rate of the voice. Although the primary effect is always for increased rate to decrease benevolence rat-

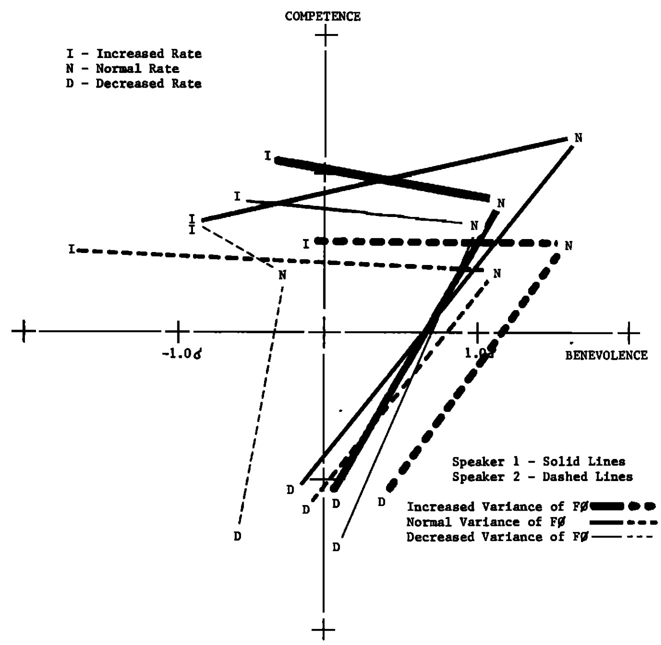

FIG. 2. The effects of rate manipulation on decreased mean $F\emptyset$ voices for each of three levels of variance of $F\emptyset$.

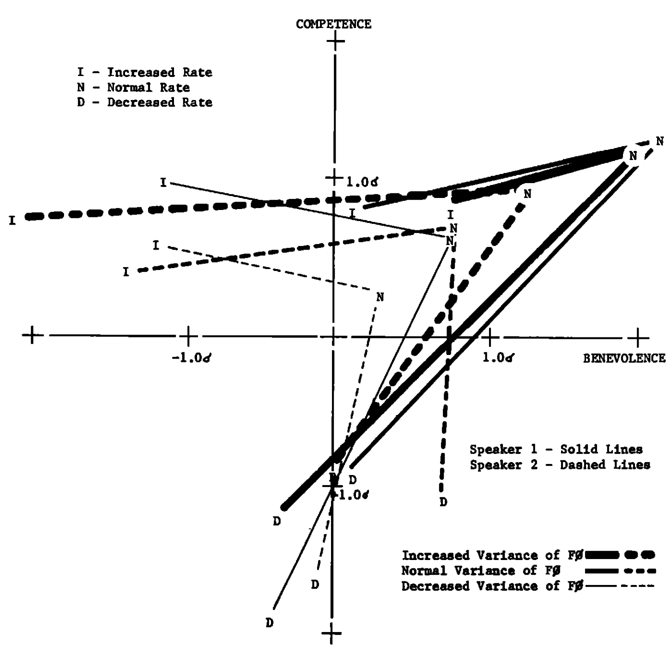

FIG. 3. The effects of rate manipulation on normal mean $F\emptyset$ for each of three levels of variance of $F\emptyset$.

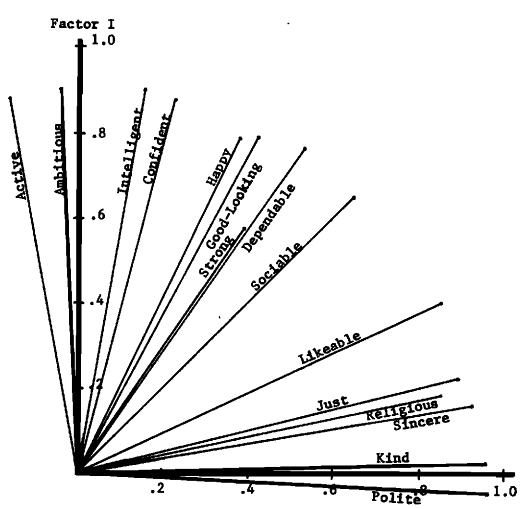

FIG. 1. Factor pattern for the adjective ratings of 54 voices.

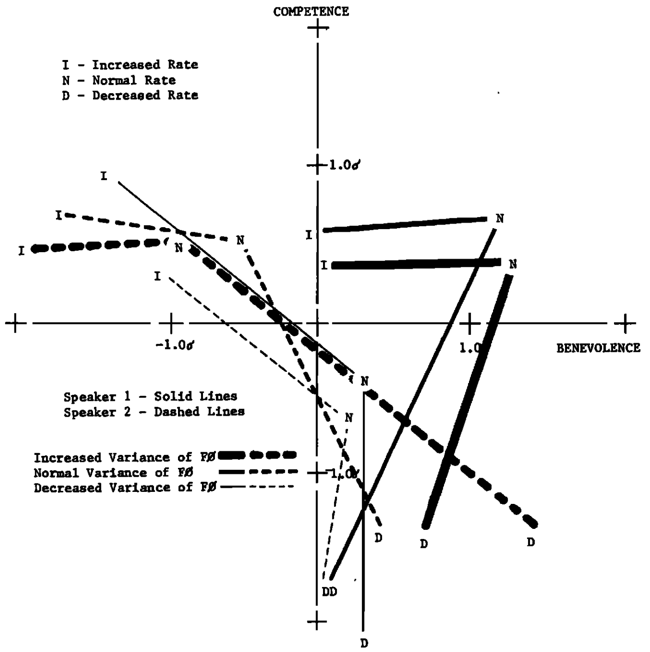

FIG. 4. The effects of rate manipulation on increased mean $F\emptyset$ voices for each of three levels of variance of $F\emptyset$.

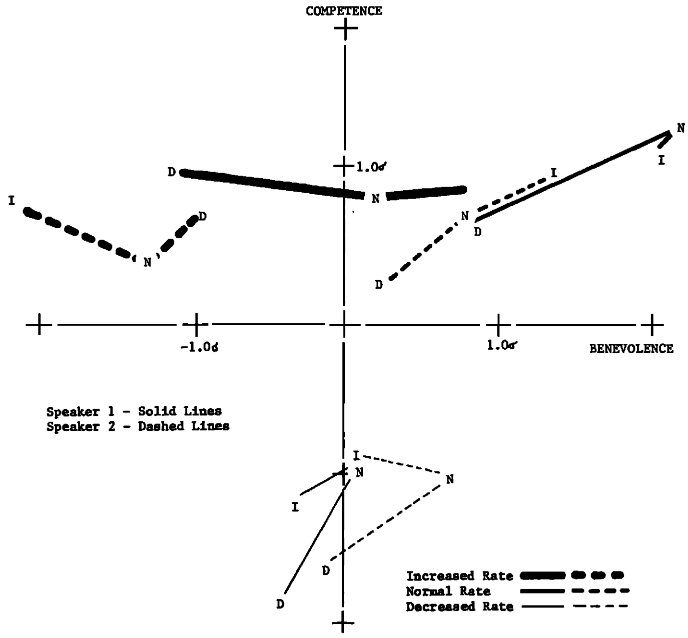

FIG. 6. The effects of variance of $F\emptyset$ manipulation on normal mean $F\emptyset$ voices for each of three levels of rate.

ings, it appears to sometimes increase competence and sometimes decrease it. In general, it increases competence when the competence rating of the normal rate voice is rather low and decreases competence when the competence rating of the normal rate voice is high. It looks like a kind of "regression toward the mean" phenomenon.[8]

One generality may also be applicable to the variance of $F\emptyset$ manipulation. In 16 out of the 18 decreased variance of $F\emptyset$ voices, either competence rating or benevolence rating decreased. However, when the effect of decrease in variance of $F\emptyset$ is looked at for competence and benevolence separately, the results are not very general. A decrease in variance of $F\emptyset$ decreased competence 12 out of 18 times (with the other being either an increase in competence or no change), and decreased benevolence 11 out of 18 times. Such qualitative observations as these do not take into account the magnitude of the changes, and perhaps a better way to generalize is with the variance comparisons of a multivariate analysis of variance.

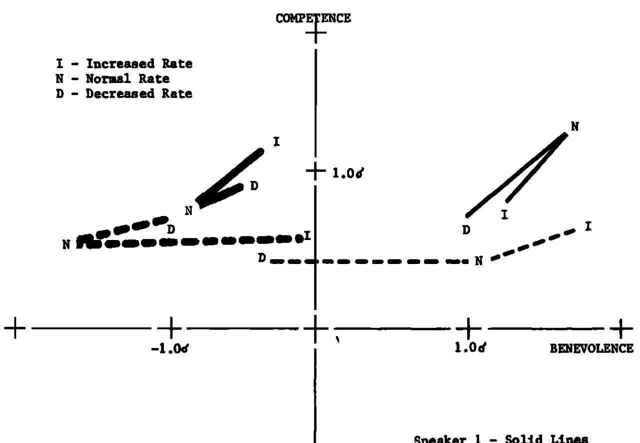

FIG. 5. The effects of variance of $F\emptyset$ manipulation on decreased mean $F\emptyset$ voices for each of three levels of rate.

The factor scores of voices were analyzed with a four-way fixed-effects multivariate analysis of variance with repeated measures on three of the factors (treatments) and the fourth (speakers) independent. Figure 8 is a plotting of the centroids of the factor scores for the simple effects of rate, mean $F\emptyset$ and variance of $F\emptyset$. Rate manipulations are seen to have by far the greatest effect upon ratings of voices, and again the trend was for decreased rate to decrease competence ratings and increased rate to decrease benevolence ratings. The Wilks' lambda value for this treatment is 0.0081 ($p < 0.001$ with $df = 2$ for treatments and $df = 26$ for error). The variance-of-$F\emptyset$ treatment (lambda $= 0.3964$) and the mean $F\emptyset$ treatment (lambda $= 0.3595$) are both significant beyond the 0.01 level (treatment $df = 2$ and error $df = 26$). Consistent with earlier findings, decreasing the variance of $F\emptyset$ causes a voice to be rated less competent and less benevolent, and increasing it causes it to be rated slightly more benevolent. Raising the mean $F\emptyset$ causes the voice to be rated less competent and slightly less benevolent.

The only interaction that is statistically significant is the rate vs variance-of-$F\emptyset$ interaction. Figure 9 shows the form of that interaction averaged over the three mean $F\emptyset$ levels. The most obvious characteristic of this interaction is the tendency for rate increase (and to a lesser extent rate decrease) to obscure any differences in received ratings among the three variance-of-$F\emptyset$ levels. (Notice how much more closely the three variance of $F\emptyset$ levels are grouped for increased rate than for normal rate.) Manipulation of variance of $F\emptyset$ appears to have its greatest effect at the middle ranges of rate. Rate, on the other hand, has great effect at any level of variance of $F\emptyset$.

This interaction can be viewed another way. From Fig. 8 it appears that increased speaking rate causes the voice to be rated lower on benevolence but generally the same on competence. A close examination of Figure 9 shows that when the normal rate voice is already rated high in competence, that is the case; but when the normal rate voice is lower in competence rating (as is the average of the decreased-variance-of-$F\emptyset$ voices in Fig. 9), the effect of increased rate is to improve competence rating. Since the two voices used in this study to generate the 54 were both "high competence" and "high benevolence" voices, the form of the effects of rate shown in Fig. 8 may not be representative of male speakers. Indeed, it was found in the study by Smith *et al.* (see Footnotes 5 and 8) that the average effect on a more representative sample of voices (covering a wider range on competence) is for competence ratings to increase as the rate of speaking is increased. It may be then, that the effects of rate and variance of $F\emptyset$ are not interactive with one another, but additive, and that the statistically significant interaction is due to a ceiling effect in the extremity of competence ratings.

In order to evaluate the extent to which the five sources of variance found to be significant (rate, variance of $F\emptyset$, mean $F\emptyset$, interaction of rate and variance-of-$F\emptyset$, and speaker) account for ratings of the synthesized voices, a model was constructed by transforming each centroid to a deviation from the grand centroid. In order to get a predicted competence–benevolence score for each rate, mean $F\emptyset$, and variance-of-$F\emptyset$

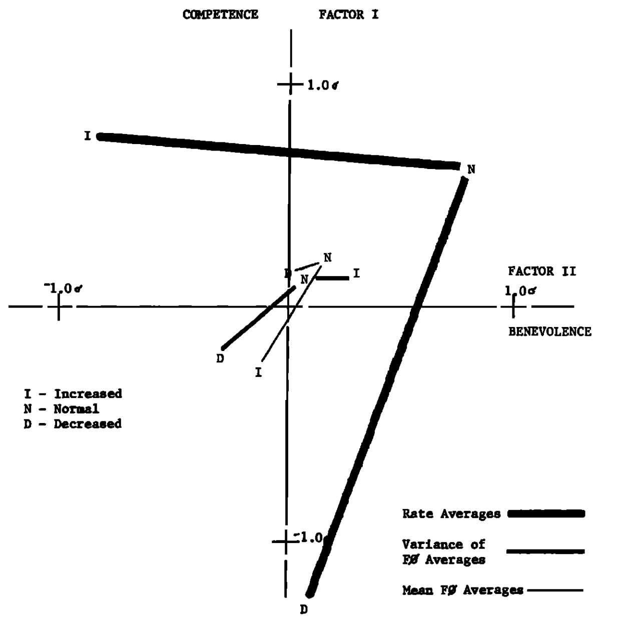

FIG. 8. Grand centroids of factor scores for the separate effects of rate, mean $F\emptyset$, and variance of $F\emptyset$ manipulations.

combination, the separate effects of each manipulation made on a given voice plus the interaction effect were added to the grand centroid. The predicted and actual factor scores for each of the 54 voices (27 manipulation combinations for each speaker) were then compared. A display of the disparities between the 54 predicted factor score plottings and their corresponding 54 actual or observed factor score plottings would be very dense and complex. Figure 10 displays a small subset of these: the predicted and corresponding observed factor scores for the nine voices which are the nine combinations of rate and variance of $F\emptyset$ generated from Speaker 1 at the decreased level of mean $F\emptyset$.

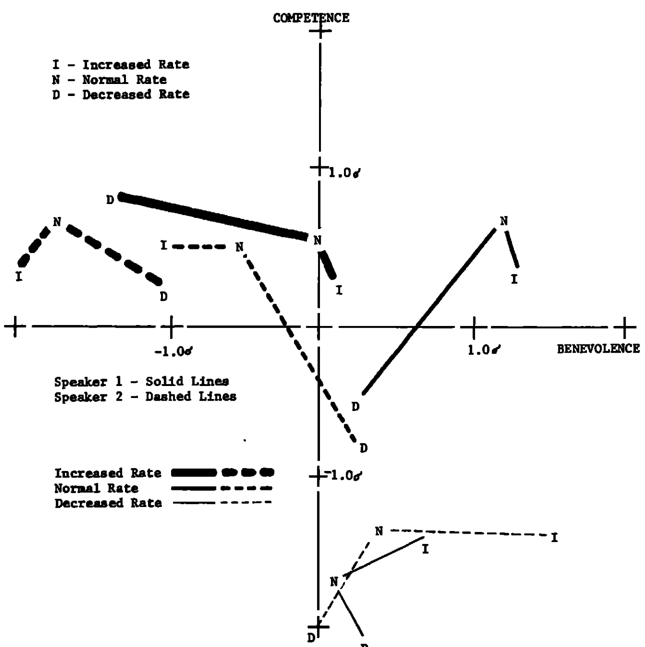

FIG. 7. The effects of variance of $F\emptyset$ manipulation on increased mean $F\emptyset$ voices for each of three levels of rate.

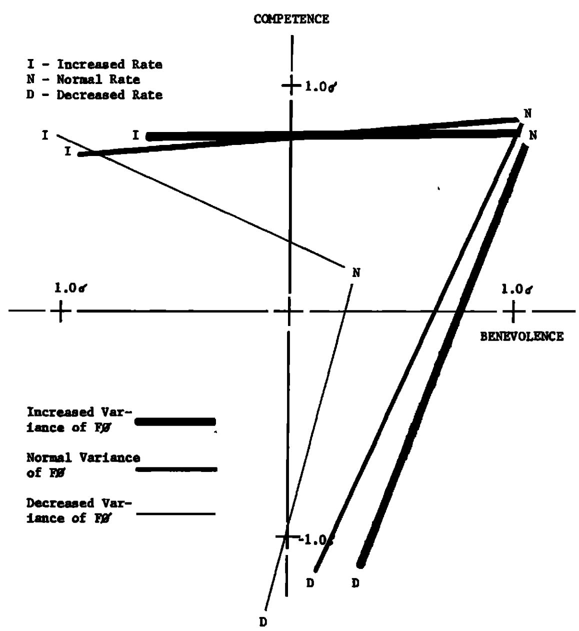

FIG. 9. Centroids of factor scores for all nine rate—variance of $F\emptyset$ combinations averaged over the three mean $F\emptyset$ levels—showing the interaction of rate and variance of $F\emptyset$.

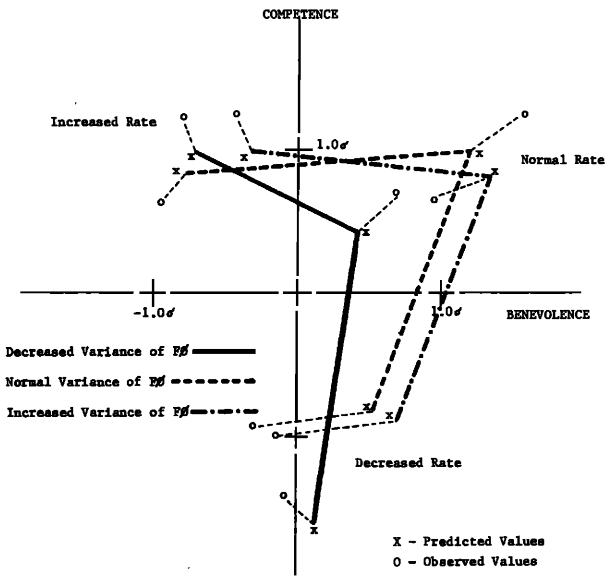

FIG. 10. Comparison of predicted and observed factor scores for the nine rate and variance of $F\emptyset$ combinations for Speaker 1 at decreased mean $F\emptyset$ level.

The estimation of proportion of variance in voise ratings accounted for[9] by each of the five significant variance sources was also computed. It is estimated that rate manipulations are responsible for 86% of the variance in competence ratings, mean $F\emptyset$ for 4%, variance of $F\emptyset$ for 3%, interaction of rate and variance-of-$F\emptyset$ for 2%, and individual speaker characteristics for 1%, leaving 4% to be accounted by for all other sources combined. Variance in benevolence ratings is accounted for 48% by rate, 1% by mean $F\emptyset$, 6% by variance of $F\emptyset$, 1% by the interaction of rate and variance-of-$F\emptyset$, 8% by speaker, and 36% by all other sources. These variance estimates have value primarily in evaluating the relative contributions of rate, mean $F\emptyset$ and variance of $F\emptyset$ to voice ratings. The amount of variance due to differences between the individual speakers that were analyzed and then synthesized (which would be mediated by acoustic characteristics other than the ones manipulated) is probably underestimated, since these two speakers are a very small sample and were not selected at random from a population of male speakers. (Both were adult male college professors.)

IV. DISCUSSION

The major findings of this and the earlier studies in this series are for the most part summarized in Figs. 8 and 9:

(1) Decreasing speaking rate causes a decrease in competence ratings (and, according to the Smith *et al.* study,[5] a decrease in benevolence ratings).

(2) Increasing speaking rate causes a decrease in benevolence ratings (and, for male voices on the whole, an increase in competence ratings[5,8]). With the six diverse voices and the nine levels of rate used in the Smith, *et al.* study,[5] competence was found to increase monotonically with measured rate, while benevolence had an inverted "*U*" relationship, the highest benevolence ratings being given to voices in the middle range of measured rate.

(3) Increased variance of $F\emptyset$ has a tendency to increase benevolence ratings and decreased variance of $F\emptyset$ causes a decrease in both competence and benevolence ratings.

(4) Increased mean $F\emptyset$ causes a decrease in both competence and benevolence ratings.

(5) The effects of rate manipulations are much more sizeable and consistent than the effects of mean $F\emptyset$ or variance of $F\emptyset$ manipulations.

(6) There is an interaction between the effects of rate manipulations and those of variance of $F\emptyset$, but this effect can be explained as a tendency for voices that are already extreme on competence or benevolence ratings to resist being moved further in that direction by manipulations that would ordinarily cause such a change.

The observation that rate manipulations have greater effect than mean $F\emptyset$ or variance of $F\emptyset$ manipulations depends, of course, on the equivalence of the extremity of the manipulations for these three parameters. The extremity of manipulations would have to be equivalenced on some kind of psychological dimension, since it is *perceived* pitch, rate, or variance of intonation rather than actual that evokes the adjective ratings. In this study, the level at which each manipulation was made was determined by the experimenters listening to a variety of levels of manipulation on each parameter and selecting levels that were maximally extreme within the constraint of still sounding like "real" voices. In order to have confidence in the relative values of the figures for the proportion of variance accounted for, it would be good to obtain realism ratings of voices that cover a broad range of extremity of manipulation on these three parameters and then equivalence the extremity of manipulation for each of the three parameters. It may also be good to produce the synthesized voices at fixed values of each parameter rather than at a given proportion of the speakers' natural values for the parameters. (The latter approach was used in this study.)

It should be remembered that the method used in this study for scaling judges' evaluations of the synthesized voice was only one of a multitude of possible approaches. For example, factor analysis could have been performed on the ratings given by a single judge rather than the averaged ratings of 37 male and female judges. An earlier study[10] indicated that if this had been done, the two factors of competence and benevolence would have accounted for less of the variance in adjective ratings. (In this study they accounted for 91.9% of the variance in the fifteen adjectives with the competence factor having an eigenvalue of 7.086 and the benevolence factor an eigenvalue of 6.702.) Such a procedure would be useful in contrasting the meanings of manipulations of rate, mean $F\emptyset$, and variance of $F\emptyset$ to individual judges. The averaging procedure used in this study drops out individual differences between judges in their impressions of the voices and abstracts common elements of a group of judges. Now that the effects of the manipulations are well established for groups of university students undifferentiated by age, sex, etc., it may be useful to contrast the rating patterns of judge groups that differ in sex, regional dialect, age, etc.

All of the studies in this series are aimed more at understanding the implicit personality theories of judges through their judgments of various contrived voices rather than establishing relationships between speech characteristics and personality attributes of the speaker. They are unique among studies of personality perception in that, as in the old psychophysical methods, they attempt to map out relationships between precise, quantitative physical dimensions and psychological judgments. As in the old psychophysics, the tough problem is finding psychological scales that are reasonably precise and nonarbitrary. The findings of this study must now be expanded to other kinds of rating scales. Although averaged judgments on 15 adjectives combined together into two factors seem to give fairly stable results when these same adjectives are used, the adjectives rating scales are only a small subset of all the possible subjective response methods that could be used. Changing the method may alter the factors. It may be productive, now that the effects of acoustic manipulation on the competence-benevolence factors are well known and replicated (on rate and variance-of-*FØ* at least) over three studies, to determine the effects of these same acoustic manipulations on other psychological response measures. The massive amounts of work in the area of psychological scaling will be useful. Multidimensional scaling would be one of the more productive response techniques to use, since it is extremely general and nonarbitrary, the judges being required (in the method of triads[11]) only to make relative judgments as to which of two or three pairs is most similar. The factored sum of products matrix (derived from the similarities matrix) then gives the coordinates of the voices in a reduced space, the axes of which could be labeled by having judges listen to voices in groups of three, two from one end of an axis and one from the other and then telling which two are most similar, how they are similar to one another (naming one end of the dimension) and how they differ from the third (naming the other end). Another important extension would be to have these voices rated on Osgood's semantic differential, in order to tie these results in to the massive amounts of research centered around that instrument.[12]

There could be productive tie-ins between work on the relationship of acoustic characteristics to emotion (such as that done by Williams and Stevens[13,14]) and studies of acoustic indicators of personality. One way of looking at personality is as the characteristic emotional tone of a person over time. In view of the great consistency in personality judgments of a given voice and the ability of human judges to identify emotions accurately from vocal qualities,[2] it appears that there are reliable acoustic indicators of personality and emotion. Working out the details of the relationships among personality, emotions and acoustic properties could become a very exciting and important area of research in the coming years.

[1]This research was supported by a grant from the National Institute of Mental Health.

[2]E. Kramer, "Judgment of Personal Characteristics and Emotions from Nonverbal Properties of Speech," Psychol. Bull. 60(4), 408–442 (1963).

[3]B. L. Brown, W. J. Strong, and A. C. Rencher, "Manipulation of Vocal Qualities by Speech Synthesis: A New Way to Study Person Perception," Proc. 80th Annual Convention of the American Psychological Association, pp. 197–198.

[4]B. L. Brown, W. J. Strong, and A. C. Rencher, "Perceptions of Personality from Speech: Effects of Manipulations of Acoustical Parameters," J. Acoust. Soc. Am. **54**, 29–35 (1973).

[5]B. L. Smith, B. L. Brown, W. J. Strong, and A. C. Rencher, "Effects of Speech Rate on Personality Perception," Language and Speech (in press).

[6]W. J. Strong, "Machine-Aided Formant Determination for Speech Synthesis," J. Acoust. Soc. Am. **41**, 1434–1442 (1967).

[7]W. J. Strong and R. B. Purves, "Scheme for Automatic Formant Analysis of Speech," J. Acoust. Soc. Am. **51**, 110(A) (1972).

[8]In the study by Smith *et al.* (see Footnote number 5), the effects of nine rate manipulations on six voices were studies. The two voices of this study were rather high in competence–benevolence ratings, whereas the six voices of that study covered a wide range of competence-benevolence ratings. In that more representative situation, rate increases were found on the average to increase competence ratings as well as decrease benevolence ratings, and rate decreases were found to decrease both competence and benevolence ratings.

[9]G. M. Vaughan and M. C. Corballis, "Beyond Tests of Significance: Estimating Strength of Effects in Selected ANOVA Designs," Psychol. Bull. **72**, 204-213 (1969).

[10]B. L. Brown, "A Social Psychology of Variations in French Canadian Speech Styles," unpublished doctoral dissertation, McGill Univ. (1969).

[11]W. S. Torgerson, *Theory and Methods of Scaling* (Wiley, New York, 1959), pp. 259–290.

[12]C. E. Osgood, G. J. Suci, and P. H. Tannenbaum, *The Measurement of Meaning* (Univ. Illinois Press, Urbana, Ill., 1957).

[13]C. E. Williams, K. N. Stevens, and M. H. L. Hecker, "Acoustical Manifestations of Emotional Speech," J. Acoust. Soc. Am. **47**, 66(A) (1970).

[14]C. E. Williams and K. N. Stevens, "Emotions and Speech: Some Acoustical Correlates," J. Acoust. Soc. Am. **52**, 1238–1250 (1972).