Journal of Phonetics (1985) 13, 39–52

Familiar voice recognition: patterns and parameters
Part II: Recognition of rate-altered voices

Diana Van Lancker and Jody Kreiman
Phonetics Laboratory, Department of Linguistics, University of California at Los Angeles, Los Angeles, California 90024, U.S.A.

and

Thomas D. Wickens
Department of Psychology, University of California at Los Angeles, Los Angeles, California 90024, U.S.A.

Received 7th September 1984

Abstract: In a further attempt to determine whether the same acoustic parameters contribute equally to the recognition of familiar voices, or whether different cues are important for different voices, samples of 30 famous male voices were altered in speaking rate, a parameter known to be important in the voice recognition process. Voice samples were presented in normal, rate-expanded, and rate-compressed conditions. Seventy-two listeners selected one of six possible responses for each item. Some voices were easily recognized, while others were not, both in the expanded and the compressed conditions. This finding suggested that information about speaking rate is not equally important for all voices. Furthermore, an attempt to regress data from this and from a previous experiment (in which recognition scores were obtained on famous voices presented backward) against nine vocal characteristics (including speaking rate) resulted in a maximum R^2 value of 0.479, further supporting the conclusion that different cues are important for recognizing different voices.

Introduction

In a previous study of the recognition of famous voices (Van Lancker, Kreiman & Emmorey, 1985), we hypothesized that voice recognition, analogously to face recognition (although in a different sense modality), involves perception of a "pattern" from a set of potential auditory cues, different cues being utilized for different voices. Our hypothesis was supported by the results of an experiment in which subjects were asked to identify famous voices from short samples played forward and backward. We found that individual voices differed substantially in the degree to which backward presentation affected recognizability. Some voices were as well recognized backward as forward, while others were nearly unrecognizable backward even though they were easily identified forward.

As described in the previous paper, the perceptual parameter most affected in backward

0095–4470/85/010039 + 14 $03.00/0

presentation is articulatory patterning. Apparently, then, this parameter was essential to the correct identification of some voices but unnecessary for others. The range of variation in the data was clearly apparent in a scatterplot of the percentage of correct recognition of voices heard forward and backward; the correlation between scores obtained on voices presented backward and forward was only 0.55. This observation led us to suggest that a given parameter (in this case, specific articulatory patterns) is central to the characterization or "quality" of some voices, but irrelevant to others, and that the primary or indispensible cue(s) to voice identity vary from voice to voice. The correct view would seem to be that voices are unique patterns, made up of elements taken from a set of possible characteristics; and voice recognition in general–like other pattern recognition processes–is best viewed as operating on this loosely structured constellation of cues, any of which, or any combination of which, can evince recognition.

Further, although the dimensions involved in recognizing a familiar voice are most certainly finite in number, the role a particular dimension plays in familiar voice recognition–even if the voice has an extreme value on that dimension–depends on values present in the other dimensions for that voice. That is to say, values on single acoustic parameters do not reflect perceptual cues operating in the actual recognition process for a given voice. Thus, for example, a voice with a strong foreign accent (e.g. Maurice Chevalier's) may be well-recognized backward if it is also very distinctive in some other dimension(s); whereas another voice with an equally salient accent (e.g. Lawrence Welk) may be poorly recognized backward. The perceptual *context* a cue occurs in must be considered in evaluating its usefulness. This view is in contrast to the spatial metaphor of a multidimensional psychological space or the type of summated differences represented by a linear prediction equation.

As a further test of this view of voice recognition, we designed an experiment in which 30 famous voices were altered in speaking rate. Several studies have shown that listeners are consistently sensitive to differences in speaking rate, particularly in evaluating personality factors in unfamiliar voices (Apple, Streeter & Krauss, 1979; Brown, Strong & Rencher, 1973, 1974). Furthermore, a factor implicating speaking rate in voice identification has emerged from a multidimensional scaling study (Murry & Singh, 1980). Although no research on the importance of rate as a perceptual cue in the recognition of familiar voices can be cited, we feel justified in assuming that speaking rate qualifies as a perceptual parameter of possible importance to familiar voice recognition as well. We assert this to be likely, even though the rate may vary within an individual speaker; despite such intraspeaker variation, the notion of characteristic differences between speakers seems an intuitively valid one. To systematically alter speaking rate in our famous voice samples, each voice was presented at three rates: normal, compressed (i.e. faster than normal), and expanded (slower than normal). We hypothesized that the effect would *not* be uniform across voices, but, instead, voices would be differentially affected by the rate alterations.

Experiment 1

Stimuli

In order that many voices may be played to many subjects, we have used famous voices in our studies of familiar voice recognition, rather than the voices of colleagues or friends of subjects, as in previous studies (e.g. Bricker & Pruzansky, 1966; Pollack, Pickett & Sumby, 1954; Compton, 1963; see Van Lancker *et al.*, 1984, for a review). For the present experiment, we selected the 30 voices best recognized in our previous study from the original set of 50 four-second samples of voices of male personalities (entertainers, politicians).

The voice samples had been edited using a computer wave-form editor to eliminate pauses and revealing background noises. Five foils (wrong answers), carefully selected to be plausible choices based on rhetorical style and other characteristics of the sample, were chosen for each target voice. Written pre-tests confirmed that no voice sample was identifiable on the basis of semantic content alone. A complete list of stimulus voices is given in the Appendix; fuller details of our sample selection, preparation, and pre-testing are given in Van Lancker *et al.* (1984).

Three versions of each voice sample were prepared—normal, rate-expanded and rate-compressed. Rate alterations were accomplished using a Lexicon II Varispeech device, which alters the rate of a speech sample without changing the fundamental frequency. Compression of stimuli increased the rate of the original sample by 33% and expansion slowed the rate by 33%. The duration of the compressed stimuli was thus 2680 ms, and that of the expanded stimuli was 5320 ms. These duration differences alone should have no effect on recognizability of the samples, since voice recognition scores have previously been shown to depend more on the number of phonemes in the sample than on sample duration *per se*; further, recognition performance generally peaks at sample durations of 500–1000 ms (Bricker & Pruzansky, 1966; Pollack *et al.*, 1954). Our voice samples are sufficiently long that induced differences in duration between our expanded and compressed stimuli should have little effect on recognition scores; and in any case, the amount of phonological and prosodic information present is the same in all three rate conditions.

Subjects and experimental design

All subjects were adults with no reported hearing abnormalities in either ear. To help assure familiarity with a sufficient number of the voices in our set, only persons educated from grammar school onwards in the United States were included. Subjects were assigned at random to three experimental groups.

Three experimental tapes were prepared, each containing 10 normal, 10 expanded and 10 compressed voice stimuli in random order. Each tape was copied onto a cassette and played on a Marantz PMD-360 tape recorder to a different group of 24 listeners (for a total of 72 subjects). To eliminate strategies based on sample identification rather than voice recognition, no subject heard the same voice sample twice. All voices were heard in all three conditions across the three groups of subjects. For example, David Frost was heard by Group I in the expanded condition, by Group II in the normal condition and by Group III in the compressed condition. The result would be a Latin Square, except for the fact that a mistake in construction of the voice tapes left the square slightly unbalanced.

As in a Latin Square, the interaction between mean differences in recognition abilities and groups of subjects is partially confounded with the effects of rate alterations on individual voices. Thus, we assume that subject–group interactions contribute comparatively little to the variance, and that all observed effects are due to the other experimental factors (rate condition and voice). Several indications support this assumption. First, the size of the groups and the random assignment of individuals to each group should minimize these effects. Second, the three groups were homogeneous with respect to factors previously associated with differences in voice recognition scores. *Post-hoc* analyses showed the groups to be comparable in mean age, in mean numbers of voices familiar to them (see below), and in numbers of males and females (all of which affected recognition scores in our previous study; see Table I). Our previous research showed no relationship between other variables (handedness, television and movie exposure, etc.) and performance in our tasks. We thus feel that differences in voice recognition must be attributed to the voices rather than the group effects.

Table I Composition of the three subject groups: mean number of voices familiar *a priori* in the three experimental conditions, mean age and age ranges, and distribution of sex (Standard deviations are given parenthetically.)

	Group 1	Group 2	Group 3
Mean number of voices known:			
Normal condition	x = 7.79 (1.9)	x = 7.67 (1.7)	x = 8.58 (1.2)
Expanded condition	8.71 (1.3)	7.92 (2.1)	7.42 (2.2)
Compressed condition	7.08 (1.7)	8.50 (1.6)	7.96 (1.9)
Mean total known	23.58 (3.6)	24.08 (4.7)	23.96 (4.8)
Mean age (years)	35.96 (12.0)	38.42 (13.6)	34.79 (7.4)
Age range (years)	16–56	22–65	24–50
Number of males/females	11/13	10/14	14/10

Procedure

Subjects were told that they would hear short samples of voices of famous men, and that some of the samples had been "acoustically altered." They chose their responses from six possible answers (the target and five foils) which were listed on an answer sheet. They also indicated their confidence in each response and the familiarity of each voice, both on scales of 1–5.

At the end of the experiment, subjects completed a questionnaire indicating, for each item in the voice set, if they knew of the person and if they felt they would normally recognize his voice. Only those subjects who indicated they would normally recognize at least half of the voices were included in the analyses. (Eleven subjects were rejected and replaced in the course of obtaining the 72 required subjects.) Mean numbers of voices deemed recognizable in this survey (out of 30 possible) were 24.04 for subject Group I, 24.05 for Group II and 23.54 for Group III.

Results

Unless otherwise stated, all results reported below are significant with descriptive levels of 5% or less.

Since subject self-reports do not necessarily accurately reflect what subjects actually do or do not "know", *two* sets of analyses of performance averaged across voices were performed: one including, for a given subject, only the voices which the subject reported that he/she would normally recognize (known voices), and one including all voices for all subjects, regardless of their familiarity (all voices).

Mean percentages correct for the three groups, for both "known" and "all" voices, are given in Table II; overall means are presented in Fig. 1. Scores calculated using all voices paralleled those including known voices; scores on known voices were significantly higher, however ($F(1, 71) = 45.91$; MSE = 67.77). Subjects performed well above chance on both known and all voices in all three conditions, with decrements in performance of 16.9% (known) and 11.8% (all) in the rate-expanded condition, and 12.3% (known) and 10.1% (all) in the rate-compressed condition.

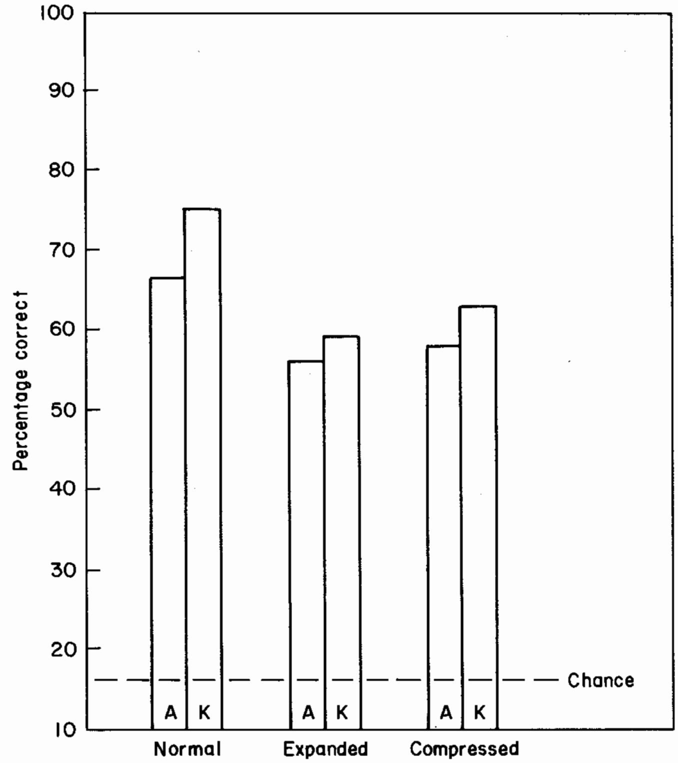

Figure 1 Percentage of correct recognition in the three experimental conditions, averaged over voices and subject groups. A represents all voices, regardless of *a priori* familiarity; K represents known voices (those subjects reported they would normally recognize) only.

Table II Percentage of voices correctly recognized, by group and rate condition. "Known Voices" results include only voices subjects reported they would normally recognize; "All Voices" includes all voices for all subjects, regardless of *a priori* familiarity

Group	Normal condition (%)	Expanded condition (%)	Compressed condition (%)
		Voices listeners "knew"	
1	67.24	44.99	64.50
2	82.62	75.62	56.11
3	76.28	57.59	68.76
Mean	75.38	59.40	63.12
Range	33.3–100	12.5–100	14.3–100
		All voices	
1	60.83	44.58	57.50
2	73.33	68.33	51.67
3	69.58	55.42	64.17
Mean	67.91	56.11	57.78
Range	20.0–100	20.0–90.0	10.0–100

A two-way ANOVA on "known" voice scores (group by rate, with repeated measures on rate) revealed main effects of group ($F(2,69) = 7.12$; MSE = 416.42) and rate ($F(2,138) = 21.16$; MSE = 237.96), and a group by rate interaction ($F(4,138) = 10.83$; MSE = 237.96). An identical ANOVA on "all" voice scores produced parallel results: significant effects of group ($F(2,69) = 5.20$; MSE = 418.42) and rate ($F(2,138) = 15.35$; MSE = 191.55) and a significant group by rate interaction ($F(4,138) = 8.21$; MSE = 191.55).

Figures 2, 3 and 4 show scatterplots pairing recognition rates in the three conditions with each other. (A list of all scores by voice is included in the Appendix.) These scatterplots show clearly that voices varied considerably in how much rate alterations affected recognition; recognition of some voices was very little affected by expansion and/or compression, while other voices were nearly unrecognizable when played in one or the other rate-altered condition. Figure 4, which matches the two rate-altered conditions, shows that these manipulations do not have identical effects on voice recognition scores; some voices are poorly recognized when expanded, but well recognized when compressed, while for others the pattern of scores is reversed. Scores in the normal and expanded conditions are correlated at $r = 0.53$; normal and compressed scores are correlated at $r = 0.66$; and expanded and compressed scores at $r = 0.56$. Thus "speaking rate" did not emerge as a single perceptual factor across the voice set.

Finally, the voices which showed large decrements in recognizability in the expanded and/or compressed conditions differed from those that were severely affected by backward presentation in our previous study. Table III lists the 10 voices in this and the previous

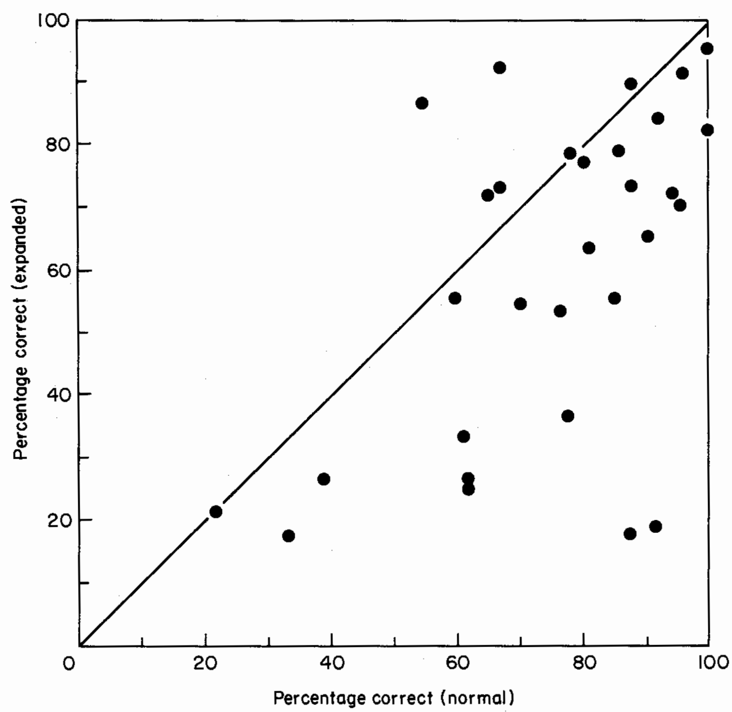

Figure 2 Percentage of correct recognition in the normal rate condition, vs the percentage correct in the expanded condition. Each point represents a single voice; data are included only for subjects who reported they were familiar with a given voice. The diagonal represents equal scores in the two conditions.

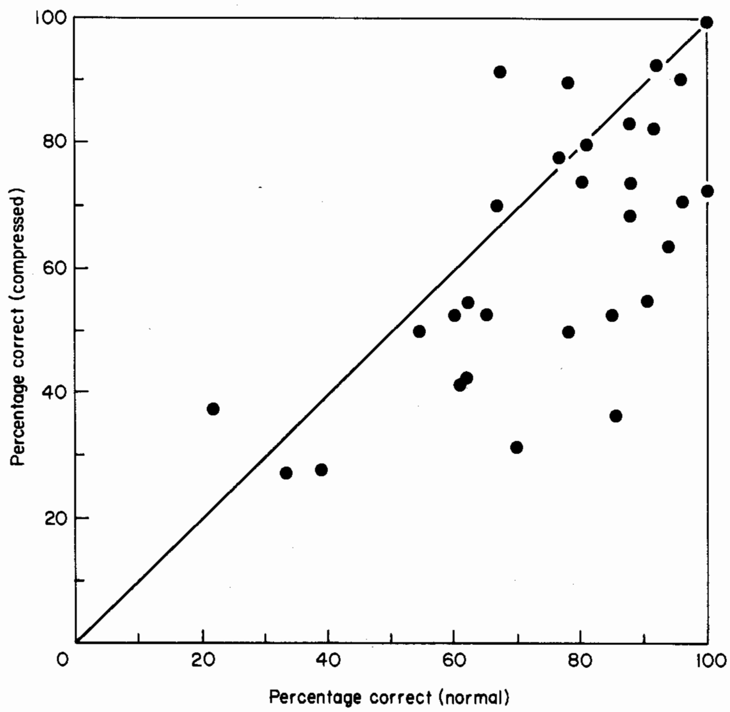

Figure 3 Percentage correct in the normal condition, vs the percentage correct in the compressed condition. Each point represents a single voice, as in Fig. 2.

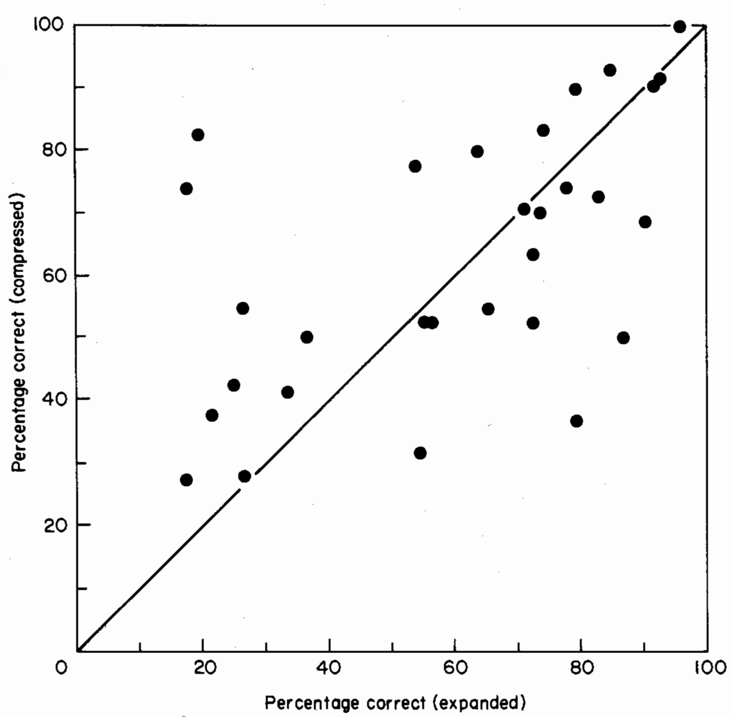

Figure 4 Percentage correct in the expanded condition, vs the percentage correct in the compressed condition (as in Figs 2 & 3).

Table III Mean scores more than one standard deviation away from the mean difference between normal score and score in altered condition

	Expanded	Compressed	Backward
David Frost			X
John F. Kennedy	X		
Edgar Bergen		X	
Vincent Price		X	
Johnny Carson	X		
Bob Hope			X
Maurice Chevalier		X	
Lawrence Welk		X	X
Tony Randall		X	
Martin Mull	X		

experiment for which the difference between the score in the normal condition and that in an altered condition (compressed, expanded or backward) was one standard deviation or more greater than the mean difference in scores. As can be seen, different voices were affected in nearly every case. Besides indicating that our results are not due to peculiarities of the voice samples used, these results support our view of voice recognition; in general, the way that an acoustic alteration affects a given voice depends on the other characteristics of the voice, such that different information is relevant to the recognition of different voices.

Discussion

As stated above, different groups heard a given voice in only one of the rate conditions (normal, expanded or compressed); and the group effect thus represents the combined effect of mean differences in groups of subjects (assumed to be zero) *and* differences contributed by the interaction of individual voices and the rate conditions. Similarly, the observed group by rate interaction could indicate that subject groups differed in which rate they found more difficult (an unlikely outcome), or that voices are differentially affected by rate expansion and compression (that some voices are more easily recognized expanded than compressed, and vice versa), or both. Since large mean differences between randomly selected groups of subjects are unlikely, and since our subject groups differ little on factors previously associated with differences in recognition scores, we conclude that most of these observed effects are due to differences in the effects of rate alteration on individual voices.

Relationship of performance to voice characteristics

We conducted another study to further test our hypothesis that different voices are identified by different parameters. To do so, we tested the opposite assumption—that a fixed set of vocal characteristics uniformly and sufficiently specifies voice patterns—by designing a rating scale to be used on the voice samples used in the studies described above. If the ease of identifying an altered voice were systematically related to specific vocal characteristics, then it should be possible to predict the identification scores from characteristics of the voices. The extent of this association should be measurable by linear regression. To investigate this possibility, we rated the characteristics of the voices and then attempted to use these ratings to predict the overall performance.

Method

The 30 famous voice samples were independently rated by five phoneticians on nine different scales drawn from factor analysis and multidimensional scaling studies of voice quality (Cartèrette & Barnebey, 1975; Voiers, 1964; Scherer, 1974; Colton & Estill, 1979; Walden *et al.*, 1978; Murry & Singh, 1980; see Table IV). Raters were provided with a tape containing the 30 unaltered 4 s samples. The voices were identified only by number on the tape and on the response form for that voice, although three of the raters were familiar with the voice samples and, of course, all raters recognized various individual voices. The raters listened to each voice segment as many times as they wished and rated the voice on the nine scales by marking points on 153 mm line segments. Endpoints of the scales were identified as indicated in Table IV. The raters were instructed to base the rating on the samples only, and not to use known characteristics of any voices or personalities that could be identified. Ratings were measured to the nearest 0.5 mm; the five ratings were averaged to obtain a single score for each voice on each measure.

Table IV Scales on which phoneticians rated the famous voice samples

Scale	Endpoints
Rate	Fast talker/slow talker
Mean pitch	High-pitched voice/low-pitched voice
Pitch range	Monotone voice/varying or colorful voice
Breathiness	Breathy voice/clear voice
Creakiness	Rough or creaky voice/clear voice
Accent	General American/strongly accented
Weight	Light voice/heavy voice
Thinness	Rich voice/thin voice
Articulation	Crisp articulation/sloppy or loose articulation

Analysis

To assess the reliability of these ratings across raters, the intercorrelation of the individual scores on each scale was measured, as was the correlation between each rater's score on each scale and the sum of ratings on that scale across raters. The correlation of each rater's scores with the totalled ratings for each scale are given in Table V. Note that some correlations are very high, and that no correlation is so poor that it would necessarily be a poor predictor of voice recognition performance in our tasks.

Multiple regression was used to predict the recognition performance on the 30 voices (using scores on "known" voices) in each condition. Since some scales appeared more reliable than others, the variables were entered into the regression in the order of the groups of features that we had judged decreasingly reliable (see Table V). Thus *rate*, which showed the greatest consistency across raters, was entered first, *pitch* and *accent*, which formed a second grouping, were entered next, and so on. The squared multiple correlation coefficients from these regressions are shown in Table VI.

Several things should be noted in interpreting this table. (1) Successive rows indicate the cumulative effect of the predictors in that row and those listed in previous rows. (2) Individual data came from the separate voices, not from subjects; subject variability has been ignored in this analysis. (3) As a consequence, only 30 observations are available for the analysis, so the number of error degrees of freedom is small and the value of R^2 is considerably inflated due to accommodation to chance factors. The penultimate line in the table

Table V The correlation of each rater's scores with the totalled ratings for each scale. The bottom line of the table shows the order in which scores were entered into the regression equation

	Scale								
Rater	Rate	Mean pitch	Accent	Articulation	Weight	Thinness	Creakiness	Pitch range	Breathiness
1	0.893	0.669	0.902	0.669	0.727	0.746	0.730	0.695	0.797
2	0.929	0.910	0.801	0.829	0.859	0.875	0.861	0.723	0.566
3	0.921	0.872	0.893	0.671	0.685	0.419	0.623	0.768	0.591
4	0.953	0.848	0.706	0.684	0.403	0.383	0.723	0.642	0.437
5	0.940	0.814	0.835	0.918	0.839	0.868	0.688	0.740	0.718
Mean	0.927	0.823	0.827	0.754	0.703	0.658	0.725	0.714	0.622
	Entered 1st	Entered 2nd		Entered 3rd		Entered 4th	Entered last		

Table VI Squared multiple correlation coefficients for variables regressed against recognition scores. Normal, expanded and compressed scores are from the present research; forward and backward scores are from research reported in Van Lancker *et al.* (1984)

Variable	df	Normal	Expanded	Compressed	ΔExpanded	ΔCompressed	Forward	Backward	ΔBackward
Rate	1	0.004	0.064	0.018	0.002	0.059	0.000	0.098	0.162
Mean pitch									
Accent	3	0.052	0.012	0.094	0.057	0.117	0.017	0.253	0.380
Articulation									
Weight	5	0.108	0.041	0.110	0.182	0.154	0.032	0.270	0.386
Thinness	6	0.194	0.102	0.354	0.183	0.231	0.092	0.287	0.386
Pitch range									
Creakiness									
Breathiness	9	0.264	0.227	0.431	0.235	0.349	0.154	0.319	0.479
Adjusted R^2		0.000	0.000	0.176	0.000	0.056	0.000	0.013	0.245
Effect noted		None	Creakiness	Thinness	Creakiness	None	None	None	Rate, accent

gives values of R^2 that have been adjusted for this effect (e.g. Marascuilo & Levin, 1983, p. 98).

Results from eight sets of regressions are presented in this table. Five of these concern the present experiment; the first three use the total recognition score as the to-be-predicted variable, one set of regressions having been done for each condition. The other two sets use the difference between the scores for the manipulated voices and those for the normal voices. For example, 'ΔExpanded' represents the difference in score between normal and expanded voices. The remaining three regressions show a similar analysis based on the results of Van Lancker *et al.* (1984). Results of the regression on scores in the normal condition may be viewed as a replication of the regression on forward scores from this previous experiment.

As an adjunct to the stepwise analysis, we also examined the individual associations at the start of the analysis. Only a few of these would be treated as "significant" by a conventional F-test, even uncorrected for multiple tests (i.e. with critical value $F(1,28) = 4.2$). These are the predictors that would be entered first in an unconstrained stepwise analysis. These predictors are listed in the bottom line of Table VI.

Discussion

The results of these regressions are not impressive. There are few effects that would be judged relevant by any criteria, and what few indications there are do not show any particularly systematic pattern over the conditions, nor are they related to the reliability of the variables; for example, *creaky*, a variable with low reliability, was "significant" for the expanded condition.

On this basis, we are inclined to conclude that linear prediction is not a successful way to understand the recognition effects for these voices. This failing, we feel, does not result from low validity of the predictors, but from the fact that individual characteristics do not predict systematically across the entire population of voices. In fact, that a voice has an extreme value on some dimension (say a strong accent or lisp) does not mean that that parameter is necessarily crucial for recognizing the voice. How central any one parameter is to the quality of a voice can only be determined in the context of the other characteristics of the voice. Thus the assumption of uniform vocal parameters serving to specify voice patterns is not supported by these analyses. Instead, these analyses suggest, in agreement with our previous studies, that characteristics helping one to identify Lawrence Welk's voice–whether or not acoustically altered–may be completely irrelevant to recognizing W. C. Fields'.

Thanks are due to Tom Wuesthoff, who prepared the stimulus tapes, ran subjects and helped tabulate the data; to Cathy Cornelius and Joy Penaloza, who also helped with subjects and data tabulation; to Karen Emmorey, who served as a rater, helped prepare the voice samples, and worked with us on the study which generated the hypotheses tested here; to Daniel Kempler and Peter Ladefoged, who served as raters; and to Emanuel A. Schegloff, who lent us the Varispeech device. This research was supported by NIRA Research Grant 17500-01.
Send requests for reprints to Diana Van Lancker, Department of Linguistics, University of California at Los Angeles, Los Angeles, CA 90024, U.S.A.

Appendix

Percentage correct recognition (by subjects *a priori* familiar with a given voice only), by voice and rate condition

Voice	Normal condition (%)	Expanded condition (%)	Compressed condition (%)
David Frost	76.5	53.3	77.8
Steve Martin	81.0	63.6	80.0
Hubert Humphrey	77.8	78.9	90.0
Will Rogers	87.5	90.0	68.8
John F. Kennedy	87.5	17.4	73.9
Everett Dirkson	91.7	84.6	92.9
Buddy Hackett	80.0	77.3	73.9
Dick Cavett	60.9	33.3	41.2
Leonard Nimoy	60.0	55.5	52.6
Edgar Bergen	94.1	72.2	63.6
Paul Harvey	66.7	73.3	70.0
Vincent Price	85.7	79.2	36.4
William F. Buckley	38.9	26.7	27.8
W. C. Fields	100	95.8	100
Johnny Carson	91.3	19.0	82.6
George Burns	95.8	70.8	70.8
Bob Hope	87.5	73.9	83.3
Ronald Reagan	33.3	17.4	27.3
Walter Cronkite	100	82.6	72.2
Jack Benny	95.7	91.3	90.5
Eddie Canter	66.7	92.3	91.7
Maurice Chevalier	85.0	55.6	52.6
Lawrence Welk	90.5	65.2	55.0
Gerald Ford	65.0	72.2	52.6
Richard Pryor	61.9	25.0	42.1
Nelson Rockefeller	54.5	86.7	50.0
Tony Randall	70.0	54.5	31.6
Jimmy Durante	21.7	21.7	37.5
Richard Burton	61.9	26.3	54.5
Martin Mull	77.8	36.4	50.0

References

Apple, W., Streeter, L. A. & Krauss, R. M. (1979). The effects of pitch and speech rate on personal attributions, *Journal of Personality and Social Psychology*, **37**, 715–727.

Bricker, P. D. & Pruzansky, S. (1966). Effects of stimulus content and duration on talker identification, *Journal of the Acoustical Society of America*, **40**, 1441–1449.

Brown, B. L., Strong, W. J. & 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. & Rencher, A. C. (1974). 54 voices from 2: 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.

Carterette, E. C. & Barnebey, A. (1975). Recognition memory for voices, *Structure and Process in Speech Recognition*, (A. Cohen & S. G. Nooteboom, eds), pp. 246–265. New York: Springer-Verlag.

Colton, R. H. & Estill, J. (1979). Perceptual aspects of some voice qualities, *Journal of the Acoustical Society of America*, **65**, supp. 1, S115.

Compton, A. J. (1963). Effects of filtering and vocal duration upon the identification of speakers, aurally, *Journal of the Acoustical Society of America*, **35**, 1748–1752.

Marascuilo, L. A. & Levin, J. R. (1983). *Multivariate Statistics in the Social Sciences: A Researcher's Guide.* Monterey, California: Brooks/Cole.

Murry, T. & Singh, S. (1980). Multidimensional analysis of male and female voices, *Journal of the Acoustical Society of America*, **68**, 1294–1300.
Pollack, I., Pickett, J. M. & Sumby, W. H. (1954). On the identification of speakers by voice, *Journal of the Acoustical Society of America*, **26**, 403–406.
Scherer, K. R. (1974). Voice quality analysis of American and German speakers, *Journal of Psycholinguistic Research*, **3**, 281–298.
Van Lancker, D., Kreiman, J. & Emmorey, K. (1985). Familiar voice recognition: patterns and parameters. Part I: recognition of backward voices. *Journal of Phonetics,* **13**, 19–38.
Voiers, W. D. (1964). Perceptual bases of speaker identity, *Journal of the Acoustical Society of America*, **36**, 1065–1073.
Walden, B. E., Montgomery, A. A., Gibeily, G. J., Prosek, R. A. & Schwartz, D. M. (1978). Correlates of psychological dimensions in talker similarity, *Journal of Speech and Hearing Research*, **21**, 265–275.