Voice Quality Analysis of American and German Speakers

Scherer, Klaus R.

How to cite

SCHERER, Klaus R. Voice Quality Analysis of American and German Speakers. In: Journal of Psycholinguistic Research, 1974, vol. 3, n° 3, p. 281–298. doi: 10.1007/BF01069244

This publication URL: https://archive-ouverte.unige.ch/unige:101850
Publication DOI: 10.1007/BF01069244

Journal of Psycholinguistic Research, Vol. 3, No. 3, 1974

Voice Quality Analysis of American and German Speakers

Klaus R. Scherer[1,2]

Received July 20, 1973

Six phoneticians rated the voices of 26 American and 22 German speakers on nine voice quality parameters which were discussed and illustrated by tape-recorded examples before the rating sessions. A reliability analysis showed highly significant interrater agreement on most parameters. Intercorrelations of the expert ratings and correlations with lay ratings of voice are reported and discussed. In concluding, empirical voice-personality relationships are reported and the role of sociocultural and attributional factors in this area is discussed.

INTRODUCTION

Recent attempts to systematize paralinguistic features of speech and to develop transcription systems (Trager, 1958; Crystal and Quirk, 1964) have included habitual voice quality or voice set, i.e., the relatively stable characteristics of the speaking voice as determined by the idiosyncratic nature of an individual speaker's vocal speech organs. In most attempts to use these

This research was supported by NIMH grant I-R03 MH 19, 569-01.

[1]Fachbereich Psychologie der Justus Liebig-Universität, University of Giessen, Giessen, West Germany.

[2]Reprint requests should be addressed to Klaus R. Scherer, Fachbereich Psychologie, Universität Giessen, 63 Giessen, Rathenaustr. 17, W. Germany. Much of the work was accomplished while the author was at the Department of Social Relations, Harvard University, Cambridge, Massachusetts, and at the Psychology Department, University of Pennsylvania, Philadelphia.

systems (e.g., Pittenger *et al.*, 1960; McQuown, in press), voice quality has been given very little attention, partly because in most paralinguistic transcriptions published to date very few different speakers were used and partly because of the problem of definition and assessment of voice quality. Abercrombie (1969) has pointed out the lack of an adequate category system accompanied by relevant acoustic criteria for voice quality assessment. Little has been done in recent years to remedy that situation.

Most attempts to define and measure voice quality stem from an interest in the relationship between voice and personality, particularly with respect to psychopathological symptoms. Unfortunately, most of the relevant studies have been the work of individual experts who have often approached the problem from a phenomenological angle (especially in the German *Ausdrucks-psychologie*; e.g., Rudert, 1965). Even those attempts at definition and measurement that have been accompanied by specification of relevant criteria such as disorders of the vocal organs (Moses, 1954) or electroacoustic measures (Fährmann, 1967) have remained unreplicated and must consequently be considered as instruments of doubtful reliability (Kramer, 1963). This is particularly salient since these researchers worked with relatively few, carefully selected speakers.

Research methods that have used electroacoustic measurements such as energy concentration in various frequency bands in the voice spectrum (Ostwald, 1963; Hargreaves and Starkweather, 1964) have fewer reliability problems, since an exact criterion is used and computerized electronic equipment rather than human judges assesses the variable under measure. However, this type of approach has been similarly restricted to cases of speakers with psychopathological symptoms, mainly depression. In addition, only one aspect of voice quality, a frequency-related one, is measured rather than the range of different characteristics of voice quality, the existence of which is implied by the large number of commonly used voice quality attributes.

The present study attempted to assess the possibility of empirically studying voice-personality relationships by defining a number of voice qualities in such a way as to make ratings by a *group* of expert judges possible. This procedure allows for the computation of reliability coefficients, hence providing some estimate of the replicability of the results. In addition, an attempt was made to avoid the problem of interference by speech variations and language content that arises when unmasked speech is used as basis for judgments of voice quality. In the present case, "randomized splicing" of the audio tape was adopted, a novel masking technique developed by Scherer (1971) which randomly rearranges arbitrary segments of the speech flow. This

technique preserves voice quality but masks both content and most major sequential speech characteristics.

A third feature of the present study was a cross-cultural approach using both American and German speakers to assess possible influences of language differences on judgments of voice quality. The importance of this cultural factor was early recognized by the anthropologist Edward Sapir (1927).

In order to assess the factors involved in self and other perception of voice, the expert voice quality ratings were compared to voice ratings by the speakers themselves, ratings of their voices by three peers, and ratings by various groups of lay judges. Parts of the data were also compared to computerized electroacoustic measurements.

METHODS

Detailed descriptions of the collection and recording methods used to obtain the voice samples utilized in this project are presented elsewhere (Scherer, 1972*a*). In the following, only the most important procedures will be described.

Voice Recording

In Cambridge, Massachusetts, and Cologne, Germany, adult male middle-class speakers (mean age, American 34.3 and German 35.1 years), recruited from adult education center files, were invited to come to the laboratory in groups of six to take part in mock jury discussions. After an initial personality testing session, the six "jurors" in each group discussed a criminal case for about 1 hr, sitting around one side of an oval-shaped table. The discussion was recorded in full, using three Electro-Voice microphones and a Uher 8000 Royal stereo tape recorder in Cambridge and 3 AKG microphones and a B & O stereo tape deck in Cologne. In each case, two speakers shared one microphone.

Sample Selection

From the complete transcript, a 20-sec speech sample consisting of segments taken from the beginning, middle, and end of the session was selected for each "juror." This sample was content-masked by removing all silent pauses, cutting the remainder of the tape into approximately 1-inch pieces (at 7 3/4 inches/sec), randomly rearranging the pieces, and splicing

them back together in random order, following the randomized-splicing procedure outlined in Scherer (1971). Samples from 26 American and 22 German speakers were used in the present study.

Self and Peer Voice Ratings

All speakers rated their own voice on a 35-item voice quality attribute rating form at the end of the personality testing session. In addition, they were asked to give envelopes containing the same voice and personality rating forms to three acquaintances ("peers") of the same age, sex, and social class. These peers were to return their ratings directly to the principal investigator. Using factor-analytic methods on both sets of ratings, the 35 voice attributes could be reduced to the following 15 voice scales: pleasantness, resonance, depth, breathiness, warmth, thinness, high pitch, sharpness, loudness, harshness, gloom, hoarseness, flatness, nasality, and dryness.

Lay Judges' Voice Ratings

Both the American and German voice samples were rated by German and American lay judges, adult females between 20 and 50 years of age recruited from adult education center files in Cambridge, Massachusetts, and Cologne, Germany. The raters were asked to come to the laboratory for a rating study on voice and personality. Voice samples for 12 selected speakers in each case were played back on hi-fi equipment, and the judges rated the voice of each speaker on the 35-item voice quality attribute form. In addition, they rated the personality of each speaker. The 12 American speakers were rated by ten American judges (AS-AJ1) and ten German judges (AS-GJ), the 12 German speakers were rated by eight American judges (GS-AJ) and seven German judges (GS-GJ). In addition, a group of ten American female undergraduate college students (AS-AJ2) rated 24 American speakers on 12 of the voice quality scales mentioned above.

Expert Voice Ratings

Six phoneticians, all Ph.D.s or advanced graduate students in linguistics or speech communication (including one of the authors, B. J.), served as expert voice raters. In a preliminary meeting, the following voice quality parameters to be listened for and the poles of the respective judgmental scales (in parentheses) were discussed and agreed on: (1) absolute pitch height (low, high), (2) pitch range (narrow, wide), (3) loudness or vocal effort (soft, loud),

(4) loudness range or dynamic contrast (little, much), (5) preciseness of articulation (loose, precise), (6) breathiness (none, breathy), (7) creak (none, creaky), (8) glottal tension (open, tight), (9) nasality (none, nasal). These parameters, described in detail below, were rated on 7-point scales.

Before the actual rating sessions took place, the second author (B. J.) recorded a tape with speech samples to illustrate parameters 5-9, which was played back to the raters at various times before and during the rating sessions. Before starting the actual rating procedure, the raters listened to short samples of seven American and five German speakers to get an idea of the range of voices to expect and of the nature of the content-masking technique. The ratings were collected during two separate group rating sessions during each of which both German and American speakers were judged. The raters listened on the average to approximately 2 min worth of voice sample per speaker (continuous repeat of the random-spliced 20-sec speech samples) and then independently recorded their judgments for the nine parameters on a scoring sheet. The voices of 26 American and 22 German speakers were rated in this way.

Voice Quality Parameters

Pitch height (1), or fundamental frequency, was supposed to be judged on an absolute scale for male voices. Pitch range (2) refers to range of pitch variation observable within the voice sample. In listening to a tape-recorded voice, the judgment of absolute loudness (3) would depend on the playback volume and on the distance from the speaker's lips to the microphone as well as on the loudness of his voice. By asking for "vocal effort," it was hoped to discount the first, exogenous factors and isolate the original absolute loudness. One study partially justifies this hope. Brandt *et al.*, (1969) recorded a sentence at seven degrees of vocal effort, "from almost a whisper to a shout," and then amplified or attenuated the recorded signals to the same level of average intensity. Listeners perceived differences almost as great in loudness as in vocal effort. Brandt *et al.* suggest that the cue for loudness was probably width of band, since the sentences produced with greater effort had more power at high frequencies. In any case, vocal effort seems to produce acoustic qualities that enable it to be perceived independently from absolute loudness. In the present study, with intensity varying randomly on top of vocal effort, one would of course expect less accuracy in judgments of effort than Brandt *et al.* found when they artificially held the intensity constant. Loudness range or Dynamic contrast (4) refers to the range of loudness variation that can be observed in the voice sample.

Preciseness of articulation (5) seems to involve speed and muscle tonus of the articulators. Precise speech sounds energetic and staccato; its vowels are tenser, consonants are more fortis, and assimilation is slighter than in loose speech. The sample sentence "Good phoneticians always listen extremely carefully" was said three ways on the demonstration tape and was subsequently analyzed by spectrographic methods. The clearest example of assimilation was the [w] in *always*; it had F_2 at 900, 800, and 620 Hz, respectively, in the loose, medium, and precise versions. In the loose version, evidently the lips were least rounded and the tongue was farthest front, in anticipation of the following [ē]. In the precise version, the tongue might have even moved back from [l] before forming the [ē].

For tense vowels, the walls of the vocal tract are stiff; for lax vowels, they are flaccid. Spectrums of four reasonably steady-state vocoids—[o] and [i] in *phoneticians,* [l] in *always*, and [r] in *carefully*—revealed differently shaped formant bands in the three degrees of preciseness. In the more precise versions, bandwidths tended to be narrower, with greater amplitude differences between peaks (formants) and troughs (unresonated areas) than in the looser versions. Even for [l], which showed this effect most strongly, the acoustic difference between lax and tense was slight.

Although fortisness of consonants has no direct acoustic correlates, it tends to make for greater duration; if the articulators touch more tightly, they are apt to stay together longer. In the sample sentence, sonagrams showed the following length ratios between the most precise version and the loosest: whole sentence, 1.36; its three isolable vowels, 0.88; its nine isolable consonants and clusters, 1.53. In the precise version, [d] (in *good*) even had an audible release consisting of a very short vowel before the following [f]. Apparently, precise articulation has little effect on vowel length but greatly prolongs consonants.

Since German has a consonant-and-vowel structure much like that of English, listeners would probably interpret "preciseness of articulation" to mean the same for both languages. In German as in English, one may assume that more precise speech as here defined and rated would have less assimilation, tenser vowels, and longer, more fortis consonants than loose speech.

Breathiness (6) and creak (7) are discussed by Catford (1964) and Crystal and Quirk (1964, pp. 38-40). Breathy vowels have random noise along with voice harmonics; in what Catford calls "whispery voice" and Crystal and Quirk call "huskiness," the noise is louder. Creaky vowels have glottal trill at the same time as voice. Pike (1943, p. 127) calls this feature "laryngealization"; Pittenger *et al.* (1960, pp. 202-203) call it "squeeze." The phonetic descriptions are almost the same despite the multiplicity of labels.

Glottal tension (8) is described most fully by Chiba and Kajiyama (1958, pp. 10-39); what they call "sharp" and "soft" voice, we call "tight" and "open." Tight vowels have more energy in the higher harmonics; impressionistically, they sound "strained or rasping" (Pittenger and Smith, 1957), "pinched, narrow, squeezed," "thin," and "tense" (Lomax 1968, p. 71). In producing them, the arytenoid cartilages are closed; the head tends to be tucked back, the larynx high, and the whole throat constricted; the glottis remains closed for as much as half its cycle (Chiba and Kajiyama). Pittenger (1957; 1960, pp. 204-220) calls tightness "rasp." Catford (1964) calls it "ligamental voice," referring to the closed arytenoids. At the other extreme, open vowels have more energy in the lower harmonics. The auditory impression is "sonorous" (Pittenger *et al.*, 1960, p. 204), "hollow or booming" (Pittenger and Smith, 1957), and "resonant" (Lomax 1968, p. 72). The arytenoids are open, the larynx tends to be low, the throat open; the closing phase of the glottis is short or even absent (Chiba and Kajiyama). Crystal and Quirk (1964, pp. 40-41) call openness "resonance," at least when it is also loud.

Nasality (9) is produced by lowering the velum away from the back wall of the nasopharynx, thereby coupling the nasal cavities to the rest of the vocal tract. The acoustic effect varies; commonly everything above 1000 Hz or so is attenuated. Perceptually, nasality can be quite hard to identify. It is often confused with tightness even though acoustically the two are clearly different.

RESULTS

Expert Rater Agreement

The relevant data on interrater agreement, or reliability, for all expert voice rating scales are presented in Table I. $\bar{r}$ refers to the average Pearson correlation coefficient between all possible pairs of raters. The rest of the table contains information derived from an analysis-of-variance approach to agreement, using the intraclass correlation coefficient η, ranging from 0 to 1, which can be interpreted as a measure of the extent to which the ratings of the different raters are grouped compactly together for each speaker, i.e., how well the judges agreed with each other's ratings for one speaker as compared with their own ratings for other speakers. η is derived from the F ratio (*cf.* Friedman, 1968), which, of course, is the ratio of the variation between speakers (mean square between) to the variation within speakers among raters (mean square within). The analysis-of-variance approach is most helpful because it allows not only specification of the significance of interrater

Table I. Reliability Coefficients—Expert Voice Ratings

	American speakers					German speakers				
	$\bar{r}$	η	F	Mean square between speakers	Mean square within speakers (among raters)	$\bar{r}$	η	F	Mean square between speakers	Mean square within speakers (among raters)
Pitch height	0.485	0.486	4.872[a]	5.643	1.158	0.689	0.635	9.109[a]	8.088	0.888
Pitch range	0.298	0.391	3.318[a]	3.456	1.041	0.365	0.435	3.990[a]	3.844	0.963
Loudness	0.589	0.602	7.757[a]	4.141	0.534	0.646	0.692	11.757[a]	4.596	0.391
Contrast	0.353	0.372	3.082[a]	3.980	1.291	0.181	0.253	1.776[b]	1.800	1.014
Articulation	0.247	0.338	2.655[a]	3.364	1.267	0.308	0.324	2.512[c]	3.524	1.403
Breathiness	0.249	0.213	1.407	2.666	1.895	0.396	0.306	2.312[c]	4.905	2.121
Creak	0.514	0.431	3.937[a]	8.278	2.103	0.313	0.273	1.966[b]	3.152	1.603
Glottal tension	0.064	0.173	1.084	1.883	1.737	0.015	0.141	0.860	1.419	1.650
Nasality	0.169	0.249	1.720[a]	3.031	1.763	0.015	0.084	0.481	0.419	0.871
		Mean		4.05	1.42		Mean		3.53	1.21
		Standard deviation		1.79	0.47		Standard deviation		1.04	0.49

[a] $p < 0.001$.
[b] $p < 0.05$.
[c] $p < 0.01$.

agreement but also separation of the underlying determinants of interrater agreement. In order to obtain reliable judgment, it is necessary that the raters agree among each other and that there is some spread or range between the stimuli judged. The former is expressed in the present case by the variation between speakers (MS between), the latter by the variation within speakers (among judges–MS within). The lack of significant agreement on any one variable, represented by a low, insignificant F ratio, can result because of either too little agreement between raters, i.e., low MS within, or too little range between the stimuli, i.e., a low MS between, or both. A restricted range or spread between stimuli can lead to low correlations between judgments which may be statistical artifacts (McNemar, 1962, pp. 144-155). Table I shows that for the rating scales pitch height, pitch range, loudness, contrast, articulation, and creak for both American and German speakers, as well as for breathiness for German speakers and nasality for American speakers, there is significant agreement due to a low within-speaker variance and a reasonable range (high between speaker variance) which leads to significant F ratios. Those scales falling short of significance seem to be primarily lacking a sufficient range, i.e., a large enough difference between speakers. Although interrater variation (MS within) is somewhat higher for some of these scales, it is rarely more than one standard deviation above the mean for all nine scales, whereas the MS between for these scales is generally more than one standard deviation below the mean for all nine scales. It is very likely that differences in nasality, glottal tension, and breathiness are far less pronounced among normal speakers than the other voice quality parameters. If the former are conspicuous characteristics of relatively few speakers, it is unlikely that any of these speakers would happen to be included by chance in the relatively small number of volunteer speakers used in the present study.[3] Since the present reliability data show that the agreement between the expert judges is quite satisfactory, the individual expert ratings were combined to means for the further analyses. There is very little difference between the two languages in terms of the interrater reliability with which voice quality was assessed in each case. The pattern of the present findings seems to encourage further usage of a rating approach to voice quality measurement using trained judges and basing judgments on the criteria outlined above.

[3]It is possible that such speakers have interpersonal behavior dispositions, possibly because of their specific voice quality, that render them unlikely to volunteer for an experimental group discussion.

Voice Quality Differences Between Languages

Table I reveals interesting differences between German and American speakers in terms of how much speakers of one language differ from each other on some of the rating scales (size of the mean square between speakers). German speakers seem to differ more on pitch height and breathiness, whereas American speakers seem to differ more among each other on contrast, creak, and nasality.

Another cross-cultural comparison concerns possible stable differences in terms of the absolute levels of specific voice quality parameters between the two groups of speakers. T tests between group differences were computed to check this possibility. The German speakers tend to have somewhat higher-pitched voices ($t = 1.78$, $p < 0.08$) with less contrast ($t = 1.69$, $p < 0.10$), whereas the American voices were judged as much more nasal ($t = 5.01$, $p < 0.001$).

Intercorrelations of Expert Voice Ratings

In order to assess the degree of covariation between the nine different voice quality parameters, the ratings were intercorrelated. Table II shows the matrix of intercorrelations for both German and American speakers. For American speakers, the rating of pitch height is related to pitch range, articulation, and absence of creak. Loudness furthermore is related to contrast, and glottal tension is correlated with creak. For German speakers,

Table II. Intercorrelations Between Expert Voice Rating Scales[a]

	Pit	Ran	Lou	Con	Art	Bre	Cre	Glo	Nas
Pitch height	–	0.69[b]	0.70[b]	0.40	0.26	−0.24	−0.41	0.49[c]	−0.14
Pitch range	0.45[c]	–	0.51[d]	0.55[d]	0.04	−0.47[c]	−0.21	0.42	−0.15
Loudness	0.38	0.22	–	0.43[d]	0.43[d]	−0.42	−0.28	0.67[b]	−0.01
Contrast	−0.06	0.00	0.58[d]	–	0.13	−0.26	−0.25	0.31	−0.08
Articulation	0.40[c]	0.21	0.06	−0.05	–	0.40	−0.13	0.48[c]	−0.23
Breathiness	0.22	−0.03	−0.11	−0.44	−0.17	–	−0.03	−0.15	−0.25
Creak	−0.59[d]	−0.38	0.05	0.22	−0.34	−0.21	–	0.03	−0.03
Glottal tension	−0.35	−0.27	0.28	0.25	−0.12	−0.04	0.73[b]	–	−0.02
Nasality	0.37	0.09	0.37	−0.11	−0.11	0.25	−0.24	−0.18	–

[a]Lower left matrix diagonal contains intercorrelations of ratings for American speakers ($N = 26$), upper right diagonal those for German speakers ($N = 22$).
[b]$p < 0.001$.
[c]$p < 0.05$.
[d]$p < 0.01$.

Table III. Rotated Factor Loadings for American Expert Voice Ratings

Expert voice rating	Factor 1	2	3	4	Communality
Pitch height	0.560	0.104	−0.503	0.475	0.804
Pitch range	0.073	0.097	0.000	0.954	0.925
Loudness	0.075	0.845	−0.402	0.137	0.900
Contrast	−0.049	0.900	0.179	−0.034	0.845
Articulation	0.935	0.031	0.139	0.048	0.896
Creak	−0.506	0.298	0.319	−0.506	0.703
Nasality	−0.068	0.034	−0.947	0.032	0.903
Sums of squares	1.462	1.635	1.465	1.415	5.976

Table IV. Rotated Factor Loadings for German Expert Voice Ratings

Expert voice rating	Factor 1	2	3	Communality
Pitch height	0.795	0.166	−0.316	0.761
Pitch range	0.841	−0.166	−0.100	0.744
Loudness	0.849	0.285	−0.013	0.802
Contrast	0.660	−0.025	−0.182	0.470
Articulation	0.223	0.939	0.009	0.932
Creak	−0.598	0.646	−0.228	0.827
Nasality	−0.212	−0.052	0.943	0.938
Sums of squares	2.949	1.439	1.086	5.474

the ratings are more strongly interrelated: pitch height correlates with pitch range, loudness, and gloom, and pitch range correlates with loudness, contrast, and absence of breathiness. Loudness also correlates with contrast and articulation.[4]

The patterns of intercorrelations are summarized by factor analyses of those variables that were rated reliably. Tables III and IV show the factor loadings for the reliable expert rating scales for American and German speakers, respectively. These data are based on rotation of the principal

[4]Since there is little between-speaker variation for glottal tension (and consequently no reliable judgment on this variable), the correlation of glottal tension with other variables could be due to statistical artifacts and is consequently disregarded here.

component factors by variable using the varimax criterion, a statistical process which attempts to assign each rating scale as clearly as possible to one factor only.

For the ratings of the American voices, four factors are extracted. There is a clear amplitude-related dimension (factor 2) on which loudness and contrast load highly. Articulation (factor 1), nasality (factor 3), and pitch range (factor 4) seem to represent one dimension each, whereas pitch height and creak do not load clearly on any single factor.

For the ratings of the German voices, the dimensionality of the ratings is less clear. Only three factors are extracted. Both frequency and amplitude-related ratings as well as lack of breathiness load highly on the first factor (factor 1). Other dimensions are formed by creak (factor 3) and articulation (factor 2).

The major difference between the dimensionality of the German and American voice ratings is the absence of a clear amplitude-related factor in the German data and the absence of a clear frequency-related factor in the American voice ratings. Since the same set of judges rated both types of voices, a difference in the dimensionality of the ratings might be due to actual differences in the acoustic structure of the respective languages and/or to corresponding judgmental stereotypes (since the judges knew the respective speakers' language), or to stable differences of the voice quality of speakers in two different cultural and linguistic habitats. All of these possibilities have interesting linguistic and psychological implications. Obviously, this issue cannot be resolved on the basis of the present data alone. These data do provide intriguing support, however, for Sapir's (1927) claim that voice is a social as well as an individual phenomenon and that it may be necessary to "carve out the social part of the voice and discard it" (p. 897).

Correlations Between Expert Voice Ratings and Electroacoustic Measures

For 12 American speakers, measures of voice energy, energy variation, pitch, and pitch fluctuation were assessed by using a digital computer laboratory facility designed for on-line speech research at the speech communication unit at MIT. The core of the facility is a 36-channel filter-bank spectrum-analyzer covering 150-7250 Hz. Details on the hardware as well as the computer programs used can be found in Wolf (1972). The voice energy or intensity measure used is a monotonic function of the voice energy in the range of 150-2000 Hz which is measured by averaging the energy level detected in the first 18 filters. The pitch level of a voice or the fundamental

Table V. Correlations of Expert Voice Ratings with Computerized Filter-Bank Measurements (N = 12 Speakers)[a]

	Computer-based electroacoustic measure			
Expert voice rating	Average energy	Energy variation	Average pitch	Pitch variation
Pitch height	0.446	0.046	*-0.285*	-0.201
Pitch range	0.042	-0.141	0.196	*0.446*
Loudness	*0.461*	0.153	-0.215	-0.205
Contrast	0.099	*0.584*[b]	-0.418	-0.305
Articulation	0.337	0.092	0.069	0.204
Breathiness	0.096	-0.318	0.328	0.138
Creak	0.005	-0.046	0.292	0.111
Glottal tension	-0.429	0.401	0.282	0.151
Nasality	0.364	-0.003	0.134	-0.073

[a]The lack of significant positive correlation for pitch height seems to be primarily due to the technical difficulties in the computerized extraction of fundamental frequency. It is possible that the extraction program at times measures the first harmonic rather than the fundamental. These "octave jumps" are difficult to detect in purely statistical analyses of natural speech and may lead to severe statistical artifacts.
[b]$p < 0.05$.

frequency (F_0) is measured by low-pass filtering above the first harmonic and detecting zero crossings. The results are shown in Table V. These data seem to support and validate the respective voice quality ratings of the expert judges, even though the correlations do not always reach significance because of the small number of observations (N = 12).

Correlations Between Expert Voice Ratings and Lay Ratings

Pearson rs between the mean rating for each group of lay judges and the expert voice ratings were computed. Consistent relationships were found, especially for the expert ratings of pitch and loudness. These correlations are shown in Tables VI and VII. The data show that for both American and German speakers, for both types of lay judges as well as self and peer ratings, the higher-pitched voices, as defined by expert ratings, are seen as high and sharp, and not deep, warm, resonant, or gloomy (Table VI). Loud voices as defined by expert ratings are seen by the lay judges (but not by speakers themselves or their peers) as loud, sharp, and lacking warmth (Table VII).

Some additional findings reaching statistical significance can be summarized as follows. High pitch range is related negatively to the German self (SF) and peer (PE) ratings of depth (SF, -0.28; PE, -0.44) and breathiness (SF, -0.51; PE, -0.52). High contrast is negatively related to the German self

Table VI. Correlations of Expert Pitch Height Ratings with Listener Judges' Voice Ratings[a]

Listener judges' ratings	American speakers (N = 26)					German speakers (N = 22)			
	Self	Peer	AJ1	AJ2	GJ	Self	Peer	GJ	AJ
High pitch	0.37	−0.36	0.41	0.80[b]	0.35	0.29	0.05	0.87[b]	0.79[c]
Depth	−0.12	−0.04	−0.33	* [d]	−0.50	−0.42	−0.49[e]	−0.83[b]	−0.84[b]
Sharpness	0.35	−0.14	0.53	0.82[b]	0.67[c]	−0.31	−0.20	0.84[b]	0.52
Resonance	0.03	−0.00	0.15	−0.57[c]	−0.28	0.15	−0.12	−0.66[e]	−0.77[c]
Warmth	0.13	−0.16	0.13	−0.45[e]	−0.59[e]	0.36	−0.29	−0.79[c]	−0.66[e]
Gloom	0.27	−0.07	−0.65[e]	−0.60[c]	−0.26	−0.43	−0.19	−0.32	−0.69[e]

[a]AJ1, judges rated both voice and personality (V + P); AJ2, judges rated voice only (V).
[b]$p < 0.001$.
[c]$p < 0.01$.
[d]*, not rated.
[e]$p < 0.05$.

Table VII. Correlations of Expert Loudness Ratings with Listener Judges' Ratings[a]

Listener judges' ratings	American speakers (N=26)			German speakers (N=22)	
	AJ1	AJ2	GJ	GJ	AJ
Loudness	0.52	0.38	0.83[b]	0.61[c]	0.48
Sharpness	0.49	0.62[d]	0.60[c]	0.54	0.61[c]
Warmth	−0.26	−0.39	−0.61[c]	−0.60[c]	−0.59[c]

[a]AJ1 = V + P, AJ2 = V.
[b]$p < 0.001$.
[c]$p < 0.05$.
[d]$p < 0.01$.

and peer ratings of breathiness (SF, −0.48; PE, −0.40), sharpness (SF , −0.51; PE, −0.50), and hoarseness (SF, −0.56; PE, −0.40). Good articulation is negatively related to the lay judges' ratings of flatness (GS-GJ, −0.54; GS-AJ, −0.46; AS-AJ, −0.78). High expert ratings on creak are positively related to American lay judges' ratings on harshness (GS-AJ, 0.45; AS-AJ1, 0.80; AS-AJ2, 0.44) and hoarseness (GS-AJ, 0.51; AS-AJ, 0.72).

CONCLUSIONS

The results reported above may be interpreted to show the feasibility of an expert rating approach to voice quality analysis. Satisfactory reliability has

been found for most voice rating scales, and the parameters for which no reliability was found do not seem to be differentiated well enough between the speakers used in the present study. Since this fact rather than a lack of interrater agreement seems to be the reason for the lack of reliability, it seems feasible to attempt to reliably assess these scales by using speakers who differ more widely on these parameters. Parts of the present ratings are also supported by electroacoustic analysis.

The interesting relationships between the expert voice ratings and the naive perceptions of voice quality by lay judges seem to demonstrate the importance of pitch and loudness as the most powerful vocal dimensions affecting lay judgments. Additional findings concerning relationships between expert ratings and lay ratings provide interesting suggestions but need to be replicated further. The lack of strong correspondence between self and peer ratings of voice quality and expert ratings raises the interesting question of the validity of self and peer ratings of voice. It is well known that the self-perception of one's own voice is rather distorted (Holzman and Rousey, 1966). It seems possible that social desirability factors and the attempt to bring the perception of one's voice in line with the self-image partly account for these distortions.

This seems to extend to peer perception of voice quality. It is possible that if one has been acquainted with a person for a long period of time and has heard that person's voice in many situations, the resulting wealth of experience interferes with the accuracy of the description of the voice.

A final point concerns the usefulness of the expert voice ratings in terms of their relationships to personality variables of the speakers. These data, which are reported in detail elsewhere,[5] show consistent and significant correlations between expert voice ratings and self and peer ratings of personality traits of the speakers. High pitch range seems to be associated with self-attribution of dominance, emotional stability, and affiliative tendencies, and with peer attribution of sociability and likability in both German and American speakers. Loudness and large dynamic contrast are related to self and peer attributions of emotional stability and sociability in American but not in German speakers, for whom contrast seems related to self-attributions of personal adjustment, orderliness, and achievement as well as to peer ratings of dependability and likability. Good articulation is associated with both self and peer attribution of dominance and task ability in German but not in American speakers. A creaky voice in American speakers seems to be negatively associated with self-attribution of emotional stability.

[5] The author has reported these data in an article not yet published entitled "Voice quality correlates of self, peer, and stranger personality attributions."

If the existence of these relationships can be further demonstrated in replications of the present results, a promising area of research using some of the techniques advocated in this report seems to lie ahead. It will be particularly important to separate the social factors of voice-personality relationships from the purely intraindividual ones, following Sapir's early suggestion, by expanding the research to other culture and language communities. Only if further cross-cultural invariance can be found, such as the relationship between high pitch range and peer perception of sociability and likability in the present case, does it seem worthwhile to speculate on and research the issue of a common underlying physiological substratum of voice production and personality or behavior dispositions that transcend transitory states of arousal. The present data seem to point to a preponderance of social factors, both as far as salience of particular voice qualities is concerned and in regard to specific voice-personality relationships. If personality is seen in terms of self and peer attributions rather than definite and objectively measurable "traits" of a speaker, it is possible to account for the present results in terms of specific attribution rules from voice to personality which may be different for self and peer perception and which may vary widely between cultures (*cf.* Scherer, 1972*a,b*). In the context of attribution theory and self-attribution theory (*cf.* overview in Hastorf *et al.*, 1970), it seems perfectly reasonable to argue that voice quality, which is one of many nonverbal cues which we constantly monitor (Holzman and Rousey, 1966), can serve as the basis of inferences concerning our intrapersonal dispositions and interpersonal behavior intentions. Obviously, these attributions can and do influence behavior. This line of thought can be easily carried further to the argument that voice quality influences self-concept and alters perception, and consequently influences behavior and interaction outcomes. This argument leads to the assumption that voice may, to some extent, shape personality and behavior rather than being the result of an impact of personality factors on the nerve fibers in the vocal apparatus, a view often held (Moses, 1954; Holzman and Rousey, 1966). This view obviously does not exclude the possibility of short-term influences of autonomic arousal on the voice organs affecting voice quality, as for example in highly emotional states such as fear or anger (*cf.* Scherer *et al.*, 1972). Stable or habitual voice quality, however, may well be an independent rather than a dependent variable in personality formation and behavior.

ACKNOWLEDGMENTS

The author would like to thank William A. Coughlin, Barton Jones, Peter Ladefoged, John A. Onufrak, Elaine K. Ristinen, Clodius Willis, and Malcah Yaeger for their interest and willing cooperation.

REFERENCES

Abercrombie, D. (1969). Voice qualities. In Markel, N. N. (ed.), *Psycholinguistics,* Dorsey, Homewood, Ill.

Brandt, J. F., *et al.* (1969). Vocal loudness and effort in continuous speech. *J. Acoust. Soc. of Am.* **46**: 1543-1548.

Catford, J. C. (1964). *Phonation Types: The Classification of Some Laryngeal Components of Speech Production: In Honour of Daniel Jones,* Longmans, London.

Chiba, T., and Kajiyama, M. (1958). *The Vowel: Its Nature and Structure,* Phonetic Society of Japan, Tokyo.

Crystal, D., and Quirk, R. (1964). *Systems of Prosodic and Paralinguistic Features in English,* Mouton, The Hague.

Diehl, C. F. (1960). Voice and personality: An evaluation. In Barbara, D. A. (ed.), *Psychological and Psychiatric Aspects of Speech and Hearing.* Thomas, Springfield, Ill.

Fährmann, R. (1967). *Die Deutung des Sprechausdrucks,* 2nd ed., Bouvier, Bonn.

Friedman, H. (1968). Magnitude of experimental effect and a table for its rapid estimation. *Psychol. Bull.* 70: 245-251.

Hargreaves, W. A., and Starkweather, J. A. (1964). Voice quality changes in depression. *Lang. Speech* 7:84-88.

Hastorf, A. H., Schneider, D. J., and Polefka, J. (1970). *Person Perception,* Addison-Wesley, Reading, Mass.

Holzman, P. S., and Rousey, C. (1966). The voice as a percept. *J. Personal. Soc. Psychol.* 4: 79-86.

Kramer, E. (1963). The judgment of personal characteristics and emotions from non-verbal properties of speech. *Psychol. Bull.* 60: 408-420.

Lomax, A. (1968). *Folk Song Style and Culture,* AAAS, Washington.

McNemar, Q. (1962). *Psychological Statistics,* 3rd ed., Wiley, New York.

McQuown, N. A. (ed.) *Natural History of an Interview,* New York (in press).

Moses, P. J. (1954). *The Voice of Neurosis,* Grune and Stratton, New York.

Ostwald, P. (1963). *Soundmaking: The Acoustic Communication of Emotion,* Thomas, Springfield, Ill.

Pike, L. (1943). *Phonetics: A Critical Analysis of Phonetic Theory and a Technic for the Practical Description of Sounds,* University of Michigan Press, Ann Arbor.

Pittenger, R. E., and Smith, H. L., Jr. (1957). A basis for some contributions of linguistics to psychiatry. *Psychiatry* 20: 61-78.

Pittenger, R. E., Hockett, C. F., and Danehy, J. J. (1960). *The First Five Minutes: A Sample of Microscopic Interview Analysis,* Martineau, Ithaca, N.Y.

Rudert, J. (1965). Vom Ausdruck der Sprechstimme. In Kirchhoff, R. (ed.), *Ausdruckspsychologie,* Vol. V. *Handbuch der Psychologie,* Hogrefe, Göttingen.

Sapir, E. (1927). Speech as a personality trait. *Amer. J. Sociol.* 32: 892-905.

Scherer, K. R. (1970). Attribution of personality from voice: A cross-cultural study on the dynamics of interpersonal perception. Unpublished Ph.D. thesis, Harvard University.

Scherer, K. R. (1971). Randomized splicing: A simple technique for masking speech content. *J. Expt. Res. Personal.* **5**: 155-159.
Scherer, K. R. (1972*a*). Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception. *J. Personal.* **40**: 191-210.
Scherer, K. R. (1972*b*). Persönlichkeit, Stimmqualität und Persönlichkeitsattribution. *Proceedings of the 28th Congress of the Deutsche Gesellschaft für Psychologie, Saarbrücken 1972,* Hogrefe, Göttingen (in press).
Scherer, K. R., Koivumaki, J., and Rosenthal, R. (1972). Minimal cues in the vocal communication of affect: Judging emotions from content-masked speech. *J. Psycholing. Res.* **1**: 269-285.
Trager, G. L. (1958). Paralanguage: A first approximation. *Stud. Linguistics* **13**: 1-12.
Wolf, J. J. (1972). Efficient acoustic parameters for speaker recognition. *J. Acoust. Soc. of Am.* **51**: 2044-2056.