

Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception

SCHERER, Klaus R.

SCHERER, Klaus R. Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception. *Journal of Personality*, 1972, vol. 40, no. 2, p. 191-210

DOI : 10.1111/j.1467-6494.1972.tb00998.x

Available at:

<http://archive-ouverte.unige.ch/unige:101791>

Disclaimer: layout of this document may differ from the published version.

# Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception<sup>1</sup>

Klaus R. Scherer,<sup>2</sup> *Harvard University*

---

## INTRODUCTION

The social psychologist interested in the role of the human voice in interpersonal perception processes will find the stacks of public libraries a richly rewarding source of resource material. Specifically, he will find a flood of popular books on how to improve one's voice in order to win friends and influence people. Interestingly enough, such efforts for voice improvement are seen as not only having a favorable effect on other people's impression of the speaker's personality but are also credited with a direct positive influence on the latter (Major, 1920, p. 70, Nelson & Atkinson, 1955, pp. 156-157).

It is not surprising, then, that the question of whether personality can be accurately inferred from voice has led to a large number of empirical studies. Growing out of the psychological approaches to the study of expressive behavior (Buhler, Wolff, G. W. Allport), the pioneering studies by Pear (1931), Allport and Cantril (1932), and Herzog (1933) received their impetus from the development of radio broadcasting. Later studies, well summarized by Diehl (1960) and Kramer (1963), were facilitated by improved methods for voice recording such as the phonograph and the taperecorder. The basic research design, used in all of

<sup>1</sup> This paper is based in part on a doctoral dissertation in the Department of Social Relations at Harvard University. A small part of the data has been reported at the 1971 APA Meeting in Washington, D.C. The research project has been supported by grants to the author from the program on Technology and Society and the Comparative International Program, both of the Department of Social Relations, Harvard University. The author acknowledges with sincere gratitude the contributions of Robert Rosenthal, Roger Brown, Norman Watt, Donald Olivier, and Ursula Scherer, as well as the extensive support by the Institute of Social Psychology, University of Cologne.

<sup>2</sup> Requests for reprints should be sent to the author, who is now at the Department of Psychology, University of Pennsylvania, Philadelphia, Pa. 19104

these studies, consisted in the presentation of standard speech samples of a number of speakers, whose personalities had been assessed by self ratings or inventories, to untrained listener-judges. The latter were required to judge physical attributes and personality traits of these invisible speakers

Although in some studies isolated personality traits have been identified with slightly better-than-chance accuracy, the most persistent finding has been that the inter-judge agreement (reliability) exceeded the agreement of the mean personality ratings with the external criteria, e.g. self ratings or Bernreuter Inventory scores (validity or accuracy). The lack of accurate judgment has frequently been explained by invoking the notion of "vocal stereotypes" (Pear, 1931, p. 30, Diehl, 1960, Kramer, 1963).

In a recent reconsideration of these results, Kramer (1964) has pointed out a number of severe methodological shortcomings of the early studies, e.g. the use of personality tests of doubtful validity (such as the Bernreuter Inventory), the lack of attention to listener differences, and the exclusive use of monologues on the part of the speakers (usually the reading of a standard text passage), rather than dialogue which seems better suited to the expression of personality traits associated with interactive behavior, (e.g. dominance). Most early studies failed to differentiate habitual voice quality from transitory speech variables such as intonation, rate of speech, pauses, nonfluencies, and articulation (cf Scherer, 1971). Furthermore, the speakers used in the early studies differed greatly in age, body type, geographical origin, occupation, education, etc., they were selectively recruited, and usually knew for what purpose their voices were recorded. These factors make confounding of the listener-judges' personality ratings rather likely, quite apart from the possibility that too many sources of variance may wash out any effects of personality variance. More recent studies in this area (Hunt & Lin, 1967; Markel, Meisels, & Houck, 1964) do not systematically eliminate these methodological concerns.

The present study represents an attempt to avoid some of the methodological pitfalls that seem to invalidate many of the earlier conclusions in this area. Furthermore, the study was placed in a cross-cultural context in an attempt to isolate the sociocultural

determinants of personality inferences from voice, as suggested by Sapir's (1927) early warning that "in deducing fundamental traits of personality from the voice we must try to disentangle the social element from the purely personal one. If we are not careful to do this, we may make a serious error of judgment" (p. 895).

### METHODS

All of the research procedures described below were carried out consecutively in Cambridge, Massachusetts and in Cologne, Germany. The procedures were identical in both cases except for very minor variations due to the respective setting. All stimulus and test materials were originally written in English and were translated into German with several checks on the adequacy of the translation in terms of intelligibility and connotations.

*Speaker recruitment.* Approximately 200 addresses were drawn from the address files of the Adult Education Centers in Cambridge and Cologne with the selection criterion (to keep the speaker sample homogeneous) that the respective person had to be male, between 25 and 50 years of age, and holding a white collar job. Each of the persons selected received a letter on university stationery inviting him to take part in a study designed to investigate the effects of personality traits of jurors on the verdict reached in a mock jury trial<sup>3</sup>. Anonymity was assured and reimbursement of expenses, but no honorarium, was promised. Twenty-eight Americans (mean age 34.3 years) and 31 Germans (mean age 35.4 years) who had volunteered to take part in the study, were contacted by telephone and scheduled in groups of six<sup>4</sup> for one of five different evening sessions.

*Personality assessment.* In several pretests two personality rating forms had been developed for use with self ratings and peer ratings in a multi-trait multi-method design. The Five Dimensions personality rating form (DIM)<sup>5</sup> consists of scales for five major personality dimensions which are described in detail by providing examples of how persons high or low on this dimension typically feel, act, and relate to other people, thereby providing anchors for judging a particular person. Based on the findings of studies on the factorial structure of personality

<sup>3</sup> This was in fact a secondary purpose of the research project.

<sup>4</sup> One American session was run with four "jurors" because of two no-shows, one German session was run with seven "jurors" because one expected no-show did show up.

<sup>5</sup> This form was modeled very closely after an example provided by Dr. Norman Watt.

judgment (Leary, 1957, D'Andrade, 1965, Passini & Norman, 1966), the following five dimensions were included in the DIM form. Conscientiousness (CON), Emotional Stability (EMO), Extraversion (EXT), Assertiveness (ASS), and Agreeableness (AGR). Raters were to place the person to be rated in one of seven categories in relation to all people of their acquaintance of the same sex and similar age and social status.

The Personality Attribute rating form (PAF) consisting of 35 personality adjectives was designed to cover personality traits that are not tapped by DIM such as achievement, impulsivity, aggressiveness, as well as to provide checks on the internal consistency or reliability of the personality ratings. Ten-point scales with three verbal anchors were provided for each adjective

Upon arrival in the laboratory, the subjects were given envelopes containing all materials<sup>6</sup> for the self ratings of personality and asked to fill out the forms in the prearranged sequence. After the completion of the experiment, shortly before leaving, the participants were each given three sets of the peer rating forms consisting of a letter of explanation and the DIM and PAF forms with appropriate instructions. Subjects were asked to give one set of these rating forms to each one of three acquaintances of the same sex, similar age, and comparable social status. The peers were to return the ratings directly to the investigator. Subjects for whom peer ratings were still missing three weeks later received a letter asking them to remind their acquaintances to return the forms. For almost all subjects at least two peer rating sets were received, for 23 American and 25 German subjects all three sets were returned.

*Voice recordings* After completion of the personality ratings, the six subjects in each session were ushered into the "jury room" which was specially sound-proofed and equipped with a one-way mirror. Subjects sat in a half-circle around a large oval table, facing a "legal expert," a law student who introduced the criminal case<sup>7</sup> to be discussed and opened the discussion. After that, he acted as observer without ever taking part in the discussion. All of the discussions which lasted be-

6 In addition to the forms mentioned above, the Adjective Check List (Cough & Heilbrun, 1965) and short forms of the Personality Research Form (Jackson, 1967) and the Maudsley Personality Inventory (Eysenck, 1959) were administered. The scores on these tests correlated highly with self ratings on DIM and PAF. Subjects also completed voice and speech attribute rating forms.

7 The case, involving a murderer pleading not guilty by reason of insanity, had been chosen to produce maximal disagreement between jurors. In all groups a discussion of at least one hour was needed to arrive at consensus.

tween 1 and 1½ hours were taperecorded on high quality sound recording equipment.<sup>8</sup> Three calibrated microphones were used, arranged in such a way that each speaker sat at roughly the same angle and distance to a microphone.

The jury discussion method had been chosen to create an atmosphere of emotional involvement where interpersonal aspects of personality could be revealed in interactive speech while at the same time keeping the topic of the discussion and the content of the speech somewhat constant. On both counts the situation worked out well: the jurors got very involved, and seemed completely to forget the microphones and that they were only role-playing a jury. Group participation was fairly even in most groups although there were a few jurors who would rarely take the floor and some others who would rarely yield it.

After the "verdict" had been passed, the jurors discussed the case as well as the purpose of the research project as a whole with the legal expert and the main investigator. Some time after the completion of the project the participants received a letter with a summary of the results of the study.

*Voice sample selection.* The master tapes with the recording of the total interaction were edited by preparing smaller tapes on which all the contributions of individual jurors were copied in chronological order, taking one at a time, thus creating a continuous population of speech acts per juror from which small samples could be drawn. It was decided to prepare speech samples of about one minute duration for every juror by taking sentences from various parts of his total contribution to the discussion to provide speech samples from the beginning, middle, and end of the discussion for most jurors except for those who had participated only very rarely or only at one period in the discussion. The actual editing of the one-minute speech samples was carried out by a German research assistant with a reasonable command of the English language who had only a limited understanding of the research aim in order to prevent selection bias. He had not been present during the actual discussion. His criterion was to take 20-second chunks from the beginning, middle, and end, broadly defined, of a speaker's total contribution, in which the respective speaker uttered a full sentence or part of a sentence without being interrupted and without noise from other speakers such as laughing or coughing.

Twelve American and 12 German speakers were selected, the voice

<sup>8</sup> In Cambridge, a UHER 8000 Royal stereo recorder and in Cologne, a B&O stereo recorder were used. In addition, the sessions were videotaped from behind the one-way mirror. Subjects had full knowledge of all recording procedures.

samples of whom were to serve as stimuli for personality ratings by naive listener-judges. The 12 speakers were selected in such a way that for every personality dimension there were at least two speakers who were clearly seen as either very high or very low on this dimension by their peers. A further criterion was that at least two peer ratings were available for that particular speaker and that these peer ratings showed a high degree of agreement.

*Content masking.* As none of the content-masking techniques used in earlier studies (Kramer, 1963, Rogers, Scherer & Rosenthal, 1971) succeeds in isolating pure voice quality, a new method, consisting of simply cutting a stretch of recording tape into little pieces and splicing them back together in random order, was developed (Scherer, 1971). The resulting voice samples are completely unintelligible and largely free from suprasegmental speech variables such as intonation contours, pauses, rhythm, etc.

Because of the rather tedious splicing procedure two 10-second excerpts from the first and second half of the one-minute speech sample per speaker were treated by the randomized splicing technique. The resulting 20-second voice samples were used as stimuli in the rating procedure.

*Rater recruitment.* Adopting the same procedure used in speaker recruitment, adult females,<sup>9</sup> between 25 and 50 years of age, working at home or on a white-collar job were invited to take part in a cross-cultural research project on personality impression formation. A honorarium was promised to every participant and a bonus to the "best judge" in each group. Volunteers were scheduled for four different rating sessions as follows. In Cambridge, ten American raters (AR) rated American speakers (AS) and ten further American raters (AR) rated German speakers (GS). In Cologne, eight German raters (GR) rated American speakers (AS) and seven further German raters (GR) rated German speakers (GS). Thus, there are four groups of raters AS/AR, AS/GR, GS/GR, GS/AR.

*Rating procedure.* The listener-judges were first acquainted with the rating forms which were the same as those used for the self and peer ratings, i.e. DIM and PAF, and listened to a warm-up randomized spliced voice sample to get used to the sound. Then, the 12 voice samples were played back on high fidelity equipment in a standard sequence, the

<sup>9</sup> A relatively homogenous group of all female raters was chosen as female subjects seemed more available for the rating task. In addition, women are reputed to be slightly better judges in person perception tasks (Tagiuri, 1969, p. 406)

same for each group of raters. Judges were told the nationalities of the speakers. Each voice sample was played back from a tape-loop and was heard until all judges had completed their ratings for the respective speaker. After all 12 speakers had been rated, the "best judge" was determined by spot checks with the peer ratings, judges were paid, and the procedures and results were discussed.

## RESULTS AND DISCUSSION

*Scaling.* A reduction of the 35 attributes in the PAF form to a smaller set of scales was required to eliminate the redundancies that had been introduced as internal consistency checks and to keep the statistics manageable. These scales and their items had to be the same for all eight groups of raters (self, peers, judges) in both cultures to be comparable. A combination of factor analysis and face criteria cluster analysis was used for this purpose.

In eight independent factor analyses<sup>10</sup> one factor, usually the strongest, showed high loadings for: 1) conscientious and dependable, the two others showing high loadings for 2) sociable, dominant, authoritative, and 3) tense, moody, anxious. Clusters of attributes in close vicinity to each other in the same region of the coordinate system were thus identified. The preliminary clusters were then compared and standardized between groups, yielding eight final scales. Attributes which did not cluster consistently were excluded from further statistical analyses. The following list contains the labels for the eight PAF scales with the respective number of items and average item-to-scale correlation coefficient (mean  $r$  for all groups of raters) in parentheses: Dependability (DEP, 3, .93), Task Ability (TAB, 4, .76), Neuroticism (NEU, 5, .79), Stability (STA, 3, .86), Sociability (SOC, 2, .90), Dominance (DOM, 2, .92), Likeability (LIK, 4, .87), and Aggressiveness (AGG, 3, .86). The scales showed equally high internal consistency for all groups of raters.

*Rater reliability.* Two kinds of inter-judge agreement coefficients were computed  $\bar{r}$ , the average intercorrelation between the ratings of all possible pairs of judges, and the analysis-of-variance

<sup>10</sup> The results of the factor analyses have to be interpreted with caution as the number of variables exceeds the number of observations. Factor analysis is used in the present paper as a heuristic device rather than a hypothesis-testing tool. All conclusions are based on the original correlation matrices.



Table 1 Median values\* for average inter-rater correlation coefficients ( $\bar{r}$ , reliability coefficients) and correlation ratios ( $\eta$ , eta).

	American speakers		German speakers	
	$\bar{r}$	$\eta$	$\bar{r}$	$\eta$
AR	.20	.47	.09	.41
GR	.19	.50	.05	.40

\*Median of the 13 PAF- and DIM-Scale values

based correlation ratio  $\eta$  (eta) (Peters & Van Voorhis, 1940, Friedman, 1968), which can be interpreted as a measure of how well the judges agree with each others' ratings for one speaker as compared with their own ratings for other speakers.

Table 1 shows the median values for both types of reliability coefficients. Like most earlier studies conducted with American speakers, the present data show a fair degree of inter-judge agreement for both American and German judges. This is not true, however, for the ratings of the German speakers where there is very little agreement between the judges' ratings.

Apart from the presence of larger rater differences in the case of the groups rating German speakers,<sup>11</sup> it is possible that the judges, both German and American, rated less uniformly because of stereotyped individual expectations concerning German national character. Or, it could be that the American speakers were "easier" to judge because of less complex personalities, or less diffuse voice qualities, or because they convey more expressive cues by their voices. It could be that the German language, in a similar way as tone languages, makes more use of paralinguistic features such as intonation and stress for structural use (e.g. semantic differentiation). These features may consequently be irrelevant or even misleading as expressive cues or even as indices of stable voice quality. Herzog (1933), for example, argued that intensity variations have structural use in German and are seen as stress rather than expressive cues.

11. For example, in the case of GS/GR and GS/AR, the median  $\bar{r}$ s tend to be much higher than the mean  $\bar{r}$ s indicating that the mean is reduced because of large deviations in the ratings of a few judges.

*Agreement correlations.* Table 2 shows the correlation coefficients between the DIM and PAF ratings of all combinations of rater groups indicating the degree of agreement between self, peer, and judges' mean ratings for both American and German speakers. A first question concerns the degree to which the present findings replicate earlier results which showed inter-judge agreement to exceed the agreement of the personality judgments with external criteria of personality, usually self ratings. Table 3 shows a comparison between the mean inter-rater reliability coefficients and the mean agreement correlations with self and peer ratings for all four groups of judges. The results seem to replicate the earlier findings with respect to self ratings but not with respect to peer ratings. The degree of agreement between the peer ratings and the mean ratings of the judges tends to exceed by far the degree of inter-judge agreement. This finding supports Kramer's (1964) contention that self ratings may not be valid criteria of personality in this type of research. It could be argued that the kind of personality traits that are likely to be communicated in the voice are stable dispositions for specific patterns of interactions with others in which the voice plays an important part as a medium in verbal exchanges. These interpersonal personality dispositions may be more easily recognized by peers who interact frequently with an individual than by the latter who himself might be ignorant of these dispositions or distort his perception and/or his self-report because of ego-defense or social desirability factors.

However, some of the lack of agreement between self ratings and judge ratings may be due to differences in implicit personality theory and in the number of dimensions used for personality ratings. The self ratings of the speakers are distributed in a multi-dimensional space where three factors explain less than 50% of the variance. Thus, it is not surprising that the listener-judges do not agree with the speakers on most of the traits as the judges' ratings are located in a space where three factors explain more than 80% of the variance. To use a somewhat imprecise but quite illustrative picture: the judges' ratings are compressed into a subspace of the personality space used by the speakers themselves and consequently are separated by large distances from the ratings of the

Table 2. Agreement correlation coefficients\* between all groups of raters on PAF- and DIM-Scales for American and German speakers.

Traits	Rater groups											
	American speakers						German speakers					
	SF/PE	SF/GR	SF/AR	PE/GR	PE/AR	GR/AR	SF/PE	SF/GR	SF/AR	PE/GR	PE/AR	GR/AR
Dep	.64**	.15	.16	.64**	.45	.45	-.06	-.45	.09	.36	.25	.24
Tab	.48*	.10	.18	.55*	.35	.68**	.26	.31	.38	.63**	.56*	.80***
Neu	.46	-.20	-.05	.43	.15	.53*	.56*	.16	-.05	.34	.08	.39
Sta	.26	-.32	.27	.55*	.48*	.57*	.25	.30	.04	.21	-.09	.62**
Soc	.46	.64**	.55*	.55*	.55*	.74***	.68**	.01	-.01	.10	.24	-.26
Dom	.80***	.09	.28	.20	.27	.79***	.08	-.10	-.21	.85***	.27	.40
Lik	.22	.10	.12	.54*	.26	.32	.12	-.11	.01	.01	.48*	.00
Agg	.74***	-.06	-.30	.42	-.01	.70**	.70**	-.25	-.47	.00	-.11	.43
CON	.56*	.26	.14	.23	.33	.54*	.42	-.08	-.14	.58*	.04	.52*
EMO	.16	-.11	.20	.43	.45	.30	.07	.20	.28	.79***	-.12	.08
EXT	.75***	.28	.46	.17	.61**	.74***	.27	-.39	.14	.25	.23	.15
ASS	.76***	-.20	-.07	-.09	.01	.74***	.27	.21	.27	.63**	.66**	.13
AGR	.80***	-.05	-.26	.18	-.10	.06	.37	.09	.15	.16	.16	-.12

\*N = 12

\*p < .10

\*\*p < .05

\*\*\*p < .01

Table 3. Comparison between mean<sup>a</sup> inter-rater reliability coefficients ( $\bar{r}$ ) and agreement correlation coefficients between mean<sup>a</sup> listener-judge ratings and mean<sup>a</sup> self and peer ratings

Rater group	Correlations					
	American speakers			German speakers		
	$r/SF^b$	$r/PE^c$	$\bar{r}^d$	$r/SF$	$r/PE$	$\bar{r}$
AR	129	292	196	037	204	102
GR	060	369	170	009	377	086

<sup>a</sup>Mean of 13 PAF- and DIM-Scale values

<sup>b</sup>Mean correlation coefficient over 13 scales between listener-judges ratings and self-ratings of the speakers

<sup>c</sup>Mean correlation coefficient over 13 scales between listener-judges ratings and peer ratings for the speakers

<sup>d</sup>Mean of average inter-rater correlation coefficients (reliability coefficients) over 13 scales

latter. As the peers use somewhat fewer dimensions for their ratings three factors explain about 60% of the variance, there is greater likelihood that more of their ratings will occupy the subspace taken by the judges' ratings which would improve chances for agreement on several personality traits.

Thus, the evaluation of the patterns of agreement between the self, peer, and judge ratings in Table 2 is hampered by the fact that many of the correlation coefficients are not independent of each other.

Factor analysis was chosen as a technique to provide a multivariate analysis of the basic patterns of agreement between the different groups of raters and to visualize some of the results.<sup>12</sup> In order to avoid dealing with two different kinds of rating scales, the following analyses are based only on the eight PAF-Scales. As the DIM-Scales overlapped to a large extent with the PAF-Scales, there is little information lost. The  $32 \times 32$  matrix of inter-correlations of all PAF-Scale ratings (four groups rating on eight PAF-Scales) was subjected to a principal components factor analysis. The first three factors were subsequently rotated according to the VARIMAX criterion. The factor loadings were then plotted. Figure 1 shows the results for the American speakers,

12. Again, because of the small number of observations, factor analysis was used only as a heuristic tool to facilitate the description of the results. All inferences are drawn from the original  $32 \times 32$  matrix of intercorrelations.

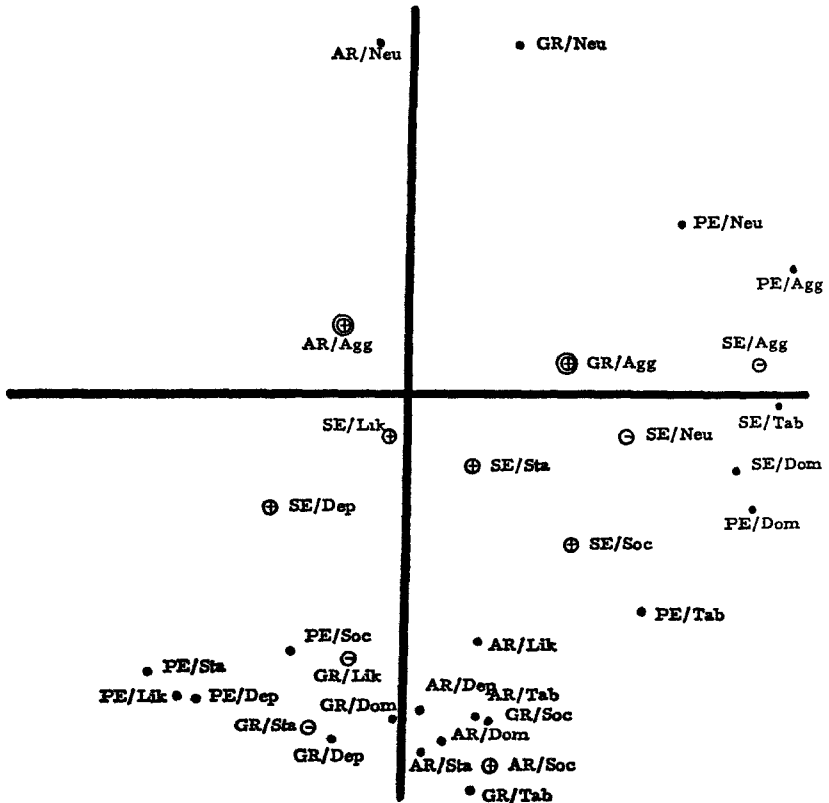


Figure 1. American speakers. Plot of loadings on 3 rotated factors<sup>a</sup> extracted by factor analysis of the agreement correlation matrix.

<sup>a</sup> Loadings on the third factor are represented by the following symbols in the two-dimensional drawing ⊕ 1.00 to .70, ⊕ .69 to .40, ● .39 to -.40, ⊖ -.41 to -.70, ⊖ -71 to -1.00

Figure 2 those for the German speakers. The following discussion is based on these data and to a very large extent, on the original  $32 \times 32$  intercorrelation matrix

Distinctive pairs or clusters of points reflect, generally,<sup>13</sup> high

13. It should be pointed out that although all highly correlated variables will be within a short distance from each other in the graphs of Figures 1 and 2, it does not hold that all variables that are plotted in close vicinity are highly correlated. The three factors plotted in Figure 1 explain only 70.3% of the total variance (Figure 2: 61%), thus some points are projected from a multidimensional space into the three-dimensional space plotted and can consequently be located next to points with which they are not highly correlated. The graphs are used only for greater transparency of the observable patterns.

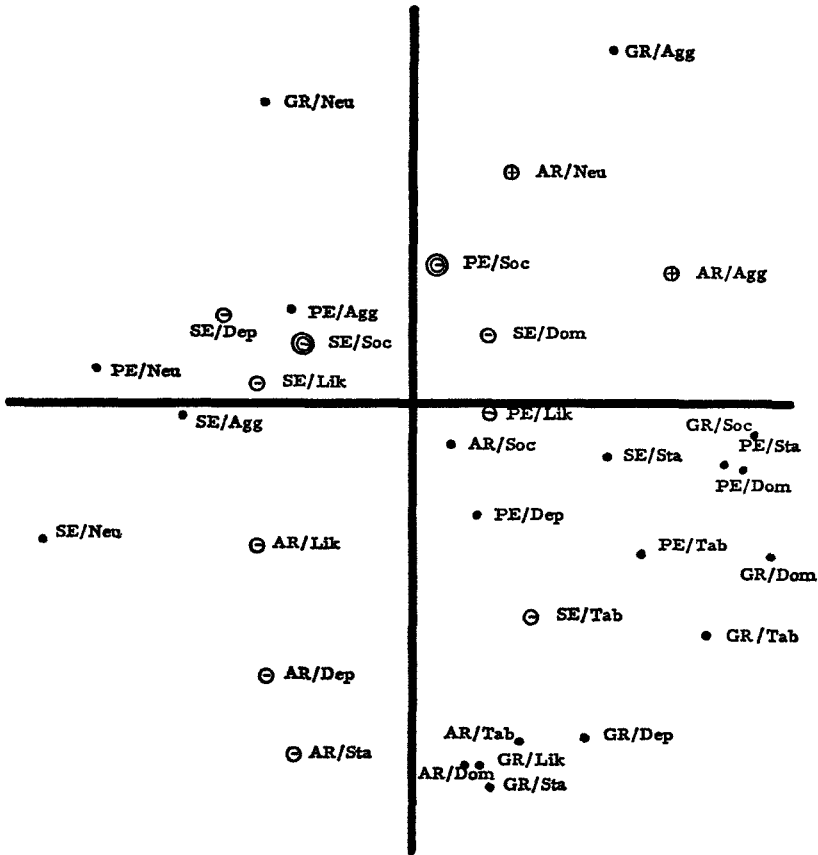


Figure 2. German speakers Plot of loadings on 3 rotated factors<sup>a</sup> extracted by factor analysis of the agreement correlation matrix.

<sup>a</sup> Loadings on the third factor are represented by the following symbols in the two-dimensional drawing ⊕ 1.00 to .70, ⊕ 69 to .40, ● .39 to -.40, ⊖ -.41 to -.70, ⊗ -71 to -1.00.

intercorrelations of the respective trait variables. Clusters formed by trait ratings of only one group, e.g. a bunch of self ratings (SF), reflect implicit personality theories and will be called "halo clusters." Clusters of trait ratings made by two or more different groups of raters will be called pairs or clusters of "corresponding attributions." For example, in the present data for the American speakers there is a significant correlation between the self ratings of Sociability (SF/Soc) and the German judges' ratings of Dom-

inance ( $r = .62, p < .05$ ). Self-attribution and other-attribution correspond as it seems likely that both are based on the same actual characteristic or trait and/or its expressive manifestations but are interpreted somewhat differently. The situation becomes more complex with clusters rather than pairs of corresponding attributions. In the case just referred to above, the self ratings of Sociability also correlate (with  $p < .10$ ) with the German judges' ratings of Task Ability, Sociability, and Aggressiveness, as well as with trait ratings by other groups of raters. Thus, one has to take into account not only corresponding attributions by different groups of raters but also the respective implicit personality theories of these groups ("halo clusters"), in order to account adequately for the patterns of intercorrelations within the larger clusters of corresponding attributions.

In the case of the ratings of the American speakers the following halo clusters (Figure 1) can be easily identified *a*) Dependability/Likeability/Stability/Lack of Neuroticism which will be called the "nice guy"-syndrome, and *b*) Dominance/Task Ability which will be called the "leader"-syndrome. Sociability is often, though not always, associated with the first cluster, Aggressiveness is sometimes, but not always, associated with the second cluster. The patterns of corresponding attributions in Figure 1 (American speakers) can be summarized as follows:

- 1) The attributions of Sociability tend to correspond with each other for all groups, i.e. speakers who see themselves as highly sociable are also seen as sociable by their peers and by judges listening to their voices.
- 2) Listener-judges from two countries tend to agree on most of their attributions, probably largely due to an overlap of the evaluative dimensions. However, those speakers to whom they both attribute the "nice guy"-syndrome *and* the "leader"-syndrome do not attribute either of these syndromes to themselves. Their peers do agree with the judges' attribution of the "nice guy"-syndrome, but not with their simultaneous attribution of the "leader"-syndrome, i.e. the peers see those speakers that they find likeable, dependable, and stable as not necessarily high on Dominance and Task Ability. The peer ratings on the "nice guy"-

syndrome agree better with the German judges' ratings than with the American judges' ratings.

3) Those speakers who see themselves as high on Dominance, Task Ability ("leader-syndrome"), and Aggressiveness are perceived by their peers in exactly the same way. Self and peer ratings also agree on Dependability.

Thus the judges correctly identified Sociability as attributed by both the speakers themselves and their peers, as well as the "nice guy"-syndrome, as attributed by the peers only. As Sociability is correlated with the "nice guy"-syndrome for both the peer ratings and the judges' ratings, these attributions are not independent. This suggests the possibility that only one personality trait can be correctly, i.e. corresponding with either self or peer attribution, judged from the voice quality of the speaker, while a larger number of corresponding attributions are due to the fact that both judges and peers make use of similar halo clusters or personality inference structures

This conclusion is supported by the pattern of correspondence between the attributions of the various groups of raters for the German speakers in Figure 2. As to halo clusters, there is the "leader"-syndrome for the peers (where it is combined with stability), for the American judges, as well as for the German judges who (as in the case of the American speakers) link the "leader"-syndrome with Sociability and the "nice guy"-syndrome.

The attributions reflected in these halo clusters show a large degree of correspondence. Those speakers to whom the peers attribute Dominance and Task Ability (the "leader"-syndrome) as well as Stability are seen by both the German judges and, to a somewhat lesser extent, the American judges<sup>14</sup> as high on the "leader"-syndrome. The German judges also tend to attribute Sociability and Dependability to these speakers. It can be argued that the patterns of corresponding attributions seem to imply that both American and German judges have based their attributions of the "leader"-syndrome, as well as other traits seen as related to it (in the case of the German judges), on some characteristics

14. The AR tend to agree with the peers on Task Ability ( $r = .56, p < .10$ ) but only slightly on Dominance ( $r = .27$ ). However, there is significant correlation for the DIM-Scale Assertiveness ( $r = .66, p < .05$ , cf. Table 2).



of the speakers, apparently communicated in the voice, that had led the peers to rate the latter high on the "leader"-syndrome.

In conclusion, only the "leader"-syndrome (Dominance/Task Ability) seem to have been identified consistently better than chance by the listener-judges in the case of the German speakers.

### CONCLUSION

If peer ratings are accepted as valid external criteria of personality, the present results lead to the conclusion that both American and German listener-judges can correctly identify Extraversion/Sociability in the case of American speakers and Assertiveness/Dominance in the case of German speakers. The cross-cultural difference with respect to the traits that can be judged accurately points to interesting implications in terms of modal personality.

Although there are few empirical studies in the area of cross-cultural differences in modal personality, a number of impressionistic analyses of national character as well as popular stereotypes assert the important role of the dominance-submission dimension for interpersonal relationships in German society (Fromm, 1941, Lewin, 1948; Dicks, 1950) and the high premium placed on outgoing, sociable behavior in American society, where an equalitarian ideology presumably discourages dominance and authoritarianism (Mead, 1942, Gorer, 1948; Riesman, 1950).

It is not unreasonable, therefore, to assume that those traits that are seen as socially desirable in a culture or that, in the case of modal personality, fulfill important functions for institutions and interactions in a specific sociocultural system (Inkeles & Levinson, 1969), can be recognized more accurately on the basis of expressive cues such as voice quality than other, less salient personality traits. This differential accuracy may be based on one or both of the following phenomena: One, a speaker may, wittingly or unwittingly, manipulate his expressive cues in such a way as to project strongly the socially desirable trait leading to a preponderance of cues indicative of that trait which will ease the recognition of the trait in the speaker by observers. Two, as the correct inference of an interaction partners' position on a modal personality dimension may be an important determinant of

interaction strategy, observers may be especially attentive and sensitive to those expressive cues which best communicate the salient trait or dimension. The second hypothesis does not seem to receive much support from the present data as American and German judges agree rather well on most of their ratings and show very similar patterns of corresponding attributions with peer ratings.

Concerning differential speaker expressiveness, the present data provide no clues as to whether the American voices communicated mainly extraversion cues, and the German voices mainly dominance cues. It is possible that the judges differentially perceived predominant cues for each group of speakers, attributed the corresponding modal personality trait, and then used their implicit personality theories to attribute further related traits yielding the halo clusters in the present data. However, the present correlational approach does not allow to test the viability of this explanation.

There is a further possibility to explain the present results. Rather than assuming that extraversion and dominance are communicated by different voice cues in each culture, it could be argued that voice quality reflects a general interpersonal dimension of personality, such as competence and/or ease in interacting with others rather than specific traits. On the basis of such "interpersonal dimension cues" in the voice, which may have been acquired in the course of verbal interactions and which may be a function of the role the speaker played in these interactions, listeners may attribute, using implicit personality theories, all those traits that seem to be related to competence in interpersonal interaction such as sociability, dominance, task ability, emotional stability, etc.

The reason that these attributions tend to correspond with the peer ratings of Sociability for American speakers and Dominance for German speakers may be that the interpersonal dimension of personality is reflected by sociable, outgoing behavior in Americans and by assertive behavior in Germans or that this dimension is perceived or coded by the peers in terms of modal personality. Some support for the latter proposition may be seen in the fact that peer and self ratings, especially for the German

speakers, show little agreement for those traits most closely related to the modal personality (cf. Table 2).

To gain more complete understanding of personality impression formation from vocal cues, it is necessary to complement the "accuracy" approach with a "process" approach to person perception (Taguri, 1969). In the present case, a detailed analysis of the personality inference/attribution process requires the assessment of the empirical covariation patterns between specific voice quality cues and valid external criteria of personality and the perception of these cues by listener-judges, as well as the measurement of the cognitive inference structures, or perceived probabilities of association between voice cues and personality traits, that determine the attributions made on the basis of the observed cues. "Accurate" judgment of personality from voice, then, seems to depend on a variety of factors: the existence of stable voice-personality relationships, the listener's ability to isolate and to perceive accurately the relevant vocal cues that communicate specific personality traits, and a large degree of correspondence between actual and inferred voice-personality covariations. The present study seems to indicate that the presence of these factors is not independent of sociocultural variables such as the social desirability or the interactional significance of specific personality traits in terms of national character or modal personality structure.

#### SUMMARY

American and German listener-judges rated personality traits of American and German speakers on the basis of short, content-masked voice samples taken from natural speech in mock jury discussions. As in earlier studies, little agreement with self ratings of personality was found. Better-than-chance agreement was found with peer ratings of Extraversion/Sociability for American speakers and Assertiveness/Dominance and related traits for German speakers. The results are discussed in terms of modal personality structure affecting corresponding attributions through the differential availability of expressive cues in speakers' voices, the sensitivity of listeners to such cues, and the congruence between cognitive inference structures of listeners and empirical voice-personality covariations.

## REFERENCES

- Allport, G W & Cantril, H Judging personality from the voice *Journal of Social Psychology*, 1934, **5**, 37-55
- D'Andrade, R G Trait psychology and componential analysis *American Anthropologist*, 1965, **67**(2), 215-228
- Dicks, H V Personality traits and national social ideology *Human Relations*, 1950, **3**, 111-154
- Diehl, C F Voice and personality An evaluation In D A Barbara (Ed), *Psychological and psychiatric aspects of speech and hearing*. Springfield, Ill C C Thomas, 1960.
- Eysenck, H J *Manual of the Maudsley Personality Inventory* London University of London, 1959
- Fromm, E *Escape from freedom*. New York Farrar and Rinehart, 1941
- Friedman, H Magnitude of experimental effect and a table for its rapid estimation *Psychological Bulletin*, 1968, **70**, 245-251
- Gorer, G *The American people* New York Norton, 1948
- Gough, H G & Heilbrun, A. B. *The Adjective Check List manual* Palo Alto, Cal. Consulting Psychologists Press, 1965
- Herzog, H Stimme und Persönlichkeit. *Zeitschrift für Psychologie*, 1933, **130**, 300-379
- Hunt, R G & Lin, T K Accuracy of judgments of personal attributes from speech *Journal of Personality and Social Psychology*, 1967, **6**, 450-453
- Inkeles, A & Levinson, D J National character The study of modal personality and sociocultural systems In G Lindzey & E Aronson (Eds), *Handbook of social psychology* Reading, Mass Addison-Wesley, 1969, Vol. IV
- Jackson, D N *Manual for the Personality Research Form* Goshen, NY Research Psychologists Press, 1967
- Kramer, E The judgment of personal characteristics and emotions from non-verbal properties of speech *Psychological Bulletin*, 1963, **60**, 408-420
- Kramer, E Personality stereotypes in voice A reconsideration of the data. *Journal of Social Psychology*, 1964, **62**, 247-251
- Leary, T *Interpersonal Diagnosis of Personality* New York Ronald, 1957
- Lewin, K Some social psychological differences between the United States and Germany In G Lewis (Ed), *Resolving social conflicts Selected papers on group dynamics, 1935-1946*. New York Harper, 1948, 3-33.
- Major, C T. *Your personality and your speaking voice* New York Grosset & Dunlap, 1920
- Markel, N N, Meisels, M, & Houck, J E Judging personality from voice quality *Journal of Abnormal and Social Psychology*, 1964, **69**, 458-463
- Mead, M *And Keep your powder dry An anthropologist looks at America*. New York Morrow, 1942.
- Nelson, T F & Atkinson, W K. *Speech and your personality*. Chicago Sanborn, 1955
- Passini, T F. & Norman, W T A universal conception of personality structure? *Journal of Personality and Social Psychology*, 1966, **4**, 44-49
- Fear, T H *Voice and personality* London Chapman & Hall, 1931
- Peters, C C & Van Voorhis, W R *Statistical procedures and their mathematical bases* New York, McGraw Hill, 1940
- Riesman, D *The lonely crowd* New Haven. Yale University Press, 1950
- Rogers, P L, Scherer, K. R & Roenthal, R. Content filtering human speech. A simple electronic system *Behavioral Research Methods and Instrumentation*, 1971, **3**, 16-18
- Sapir, E Speech as a personality trait. *American Journal of Sociology*, 1927, **32**, 892-905.

- Scherer, K R Randomized-splicing. A simple technique for masking speech content. *Journal of Experimental Research in Personality*, 1971, 5, 155-159.
- Taguri, R Person perception In G Lindzey & E Aronson (Eds ), *Handbook of social psychology* Vol 3 Reading, Mass Addison-Wesley, 1969

*Manuscript received April 30, 1971*

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.