



Article scientifique

Article

1978

Published version

Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

---

## Personality inference from voice quality: The loud voice of extroversion

---

Scherer, Klaus R.

### How to cite

SCHERER, Klaus R. Personality inference from voice quality: The loud voice of extroversion. In: European Journal of Social Psychology, 1978, vol. 8, n° 4, p. 467–487. doi: 10.1002/ejsp.2420080405

This publication URL: <https://archive-ouverte.unige.ch/unige:101858>

Publication DOI: [10.1002/ejsp.2420080405](https://doi.org/10.1002/ejsp.2420080405)

## **Personality inference from voice quality: the loud voice of extroversion\***

KLAUS R. SCHERER

*Universitaet Giessen*

### *Abstract*

*Investigated the process of personality inference from voice quality using 24 male American stimulus persons who served as subjects in simulated jury discussions. Applying a Brunswikian lens model of the inference process, criteria, distal cues, proximal cues and attributions were measured by independent groups of judges: personality criteria by three peers of each stimulus person and, on the basis of content-masked voice samples, distal voice quality indicator cues by six phoneticians, proximal voice percept by ten naïve judges, personality attributions by nine naïve judges. Only extroversion attributions correlate significantly with the criterion, replicating earlier findings. For the inference of extroversion, contrary to other traits which apparently cannot be inferred accurately from voice quality, the following conditions are met: (a) the criterion is associated with ecologically valid voice energy cues (vocal effort and dynamic range), (b) these indicator cues are adequately represented as proximal voice percepts (particularly loudness and sharpness), and (c) percept utilization in the judges' inferential strategy corresponds to the association between criterion and distal indicator cues. Path-analytic procedures are used to test empirically the adequacy of the inference model to (a) account for the variance in the attributions, and (b) explain significant correlations between criteria and attributions in terms of mediating variables.*

### **INTRODUCTION**

The inference of personality from voice seems to be too intriguing a topic to accept wholeheartedly the disappointing conclusion, suggested by many early research efforts, that voice-based personality judgments are mostly stereotyped and inaccurate

\*This paper is based in part on research conducted by the author while at Harvard University and the University of Pennsylvania. Some studies and parts of the analyses have been supported by NIMH grant MH-19-569-01 and DFG grant Sche 156/1. The author gratefully acknowledges contributions by Roger Brown, Barton Jones, Serge Moscovici, Patricia Nève, Robert Rosenthal, and Ursula Scherer.

(Kramer, 1963). Even though this area, along with many others, has gone out of fashion, a small but steady trickle of research reports (cf. Markel *et al.*, 1964; Hunt and Lin, 1967) keeps the area alive, if not well.

To improve on some of the severe methodological shortcomings from which most of the relevant studies suffer, Scherer (1972) has used interactive speech rather than monologues and peer ratings rather than self-ratings as criteria in a cross-cultural study of personality judgment from voice. It was found that both American and German judges could predict dominance for German speakers and extroversion for American speakers with better than chance accuracy. Following Tagiuri (1969), the need for a comprehensive process analysis of person perception was stressed. 'Accuracy' of judgment was seen to depend (a) on the existence of stable voice-personality relationships, (b) on the listener's ability to isolate and to perceive accurately the relevant vocal cues that communicate specific personality traits, and (c) on a large degree of correspondence between actual and inferred voice-personality covariations (Scherer, 1972, p. 208).

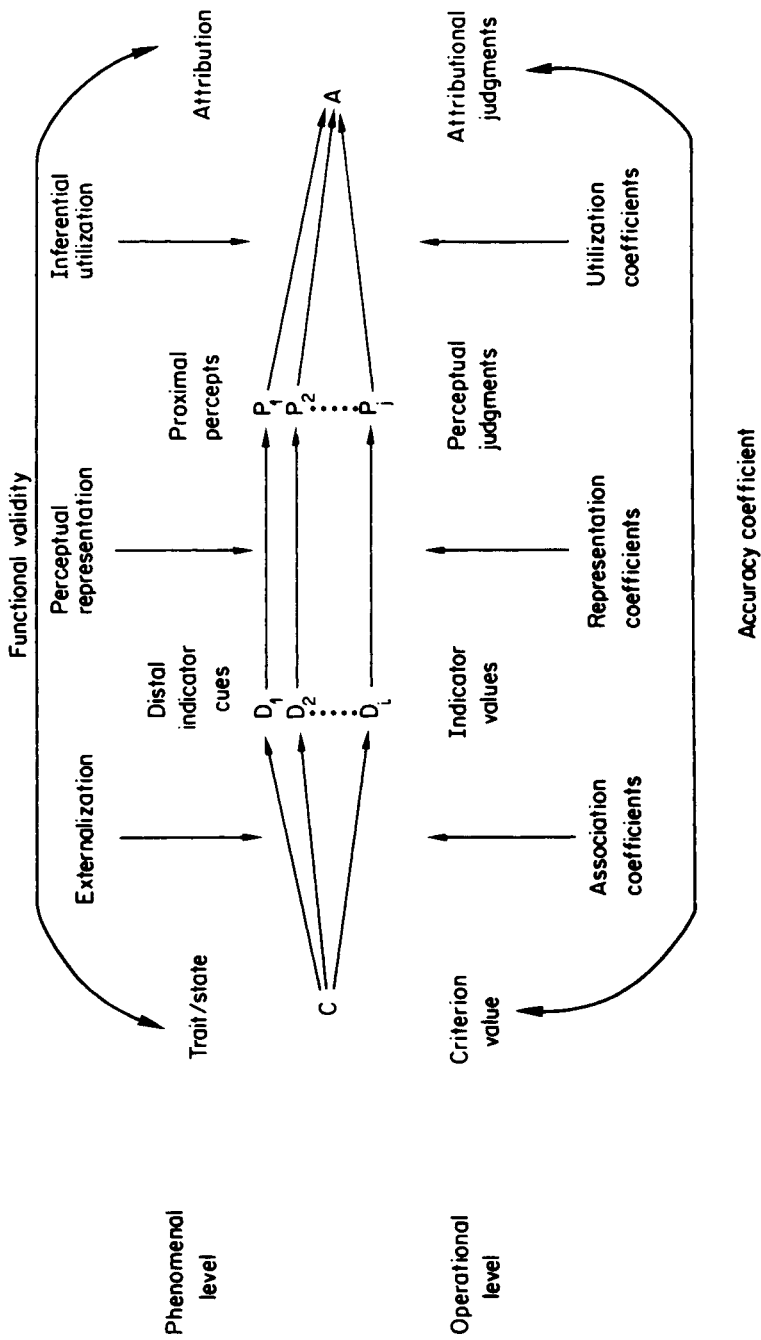
The purpose of the studies reported in this paper is twofold. The first aim is to replicate the earlier finding that extroversion can be judged accurately from the voice of American speakers, using a larger number of speakers as stimulus persons and a different group of listener judges. The second aim is an attempt to test empirically the adequacy of a model of the inference process, based on the Brunswikian (1956) lens model, with respect to its power to account for (1) the variance in the personality attributions and (2) the empirically determined correlation, or degree of 'accuracy', between criterion and attribution.

Since most of the research on clinical judgment provides judges with relatively unambiguous, previously quantified cues, such as test profiles, demographic data, etc., the role of the perceptual representation of the distal stimuli in terms of proximal cues within the organism is generally neglected in these models and is sometimes combined under 'stimulus dimensions' (cf. Dudycha and Naylor, 1966). The author believes that the proximal representation of distal cues of a stimulus person's behaviour needs to be included in a realistic model of the personality inference process.

Consequently, the full model as introduced by Brunswik was used as the basis for Figure 1. This version of the lens model is an attempt to provide some degree of standardization of the terminology for the factors and relationships on both the phenomenal and operational level. Since a theoretical discussion is beyond the scope of this paper, only a brief description will be provided.

It is assumed that stable traits and transient states are externalized in the form of distal cues in an organism's appearance and behaviour. The notion of 'externalization' includes both intentional communication of traits or states via verbal or nonverbal behaviour and 'unintentional' behavioural or physiological reactions observable in appearance or behaviour. On the operational level, traits and states are represented by 'criterion values', distal cues by 'indicator values', both of which are to be measured by reliable and valid objective methods rather than by subjective reports. The empirical covariation between criteria and indicators, representing the 'ecological validity' of the cues according to Brunswik (1956), is expressed via 'association (correlation) coefficients'.

The distal cues are proximally represented as 'percepts', being the result of a perceptual process. Operationally, percepts can be assessed via judgments or ratings



using psychophysical dimensions. The correlations between indicator values and perceptual judgments are called 'representation coefficients'. They indicate the precision with which distal cues are mapped in the perceptual space of an organism.

Attributions of traits or states are the results of inference processes based on the percepts of distal cues. These attributions are again assessed via judgments or ratings by observers, using psychological dimensions. The correlations between perceptual and attributional ratings are expressed as 'utilization coefficients' (Brunswik, 1956), indicating the extent of the utilization or weighting of a cue in inferring a trait or state.

The accuracy of the attributions in representing the actual trait or state of the observed organism is operationally defined as the correlation between the criterion value and the attributional judgment. Instead of using the somewhat ambiguous term 'achievement', originally proposed by Brunswik (1956), these correlations are here called 'accuracy coefficients'. It should be noted, however, that accuracy in this sense refers to the power of an attribution to predict a specific, operationally defined criterion rather than some 'real' state of nature.

The methodological strategy adopted in this study is to measure independently trait criteria and distal cue indicator values, as well as perceptual and attributional judgments. This allows a determination of the association, representation, utilization, and accuracy coefficients in personality inference from voice as well as to test the adequacy of the general model to represent the inference process. Consequently, the following questions will be examined:

- (1) Which dimensions of voice quality are 'ecologically valid' distal cues of major personality traits? Operationally, the association coefficients between personality criteria and objectively measured indicator values are determined. The environmental 'predictability' of the criterion is assessed by a multiple regression of the criterion on the most valid voice indicator cues.
- (2) How are those distal voice quality cues represented proximally as percepts? Here, representation coefficients, correlations between indicator values and perceptual judgments need to be computed.
- (3) What is the outcome of the inference process, i.e. which personality attributions are based on the perceptual representation of the distal voice quality cues? This relationship is expressed operationally by utilization coefficients, correlations between perceptual and attributional judgments. The judgmental policy of a judge, i.e. the joint utilization of several percepts, will be assessed in terms of a linear model using a multiple regression approach.
- (4) What are the 'functional validities' (or 'perceptual achievements' in the Brunswikian terminology) of the personality attributions from voice as measured by the 'accuracy coefficients', the correlations between the criteria and the attributional judgments? Are the accuracy coefficients reported by Scherer (1972) confirmed with a new sample of judges and a larger number of stimulus persons?
- (5) How well can the lens model account for the observed inference processes? In the case of the attribution of a specific trait based on a specific dimension or modality of distal cues, the general model represented in Figure 1 is applied by entering indicator values and the corresponding perceptual judgments for all the distal cues that are available to the observer. The more of the variance in the attributional judgments can be accounted for, the more satisfactory becomes the modelling of the inference process by this specific operational version of the lens model.

Similarly, the success of the model in explaining the accuracy of the prediction of

the criterion by the attributional judgment can be assessed by the degree to which the correlation between criterion and attributional judgment is accounted for by the relationship postulated in the model. In this paper, a path-analytic procedure is suggested as a test of the adequacy of the model to account for the empirically found variance in the attributional judgment and the covariance between criterion and attribution.

## METHOD

### **Speaker selection and preparation of stimulus material**

Thirty adult American males took part in the simulated jury discussions in groups of six persons each. The hour-long sessions were recorded on audiotape. Speech samples of 20 s duration were obtained for each speaker by editing out utterances or parts of utterances from his contribution to the discussion. Time sampling was introduced by selecting utterances from three different periods: the beginning, the middle and end of the discussion. Random-sampling could not be used since utterances free from interruptions by other speakers had to be found. These speech samples were content-masked, using the randomized-splicing procedure developed by Scherer (1971) to obtain pure voice quality samples in which content and most sequential speech cues such as rhythm, pauses and intonation are eliminated or masked. These voice samples serve as the stimulus material in all rating studies discussed below. More detailed descriptions of the speaker recruitment, the nature of the jury discussions, and the speech recordings, as well as of voice sample selection and content-masking can be found in Scherer (1972). In the present study 24 speakers for whom sufficient personality criteria ratings and tape-recorded utterances of a duration which allowed sampling had been obtained, were used as stimulus persons.

### **Assessment of personality criterion values**

The alleged purpose of the study was to examine the influence of jurors' personality traits on the outcome of jury deliberations. Before starting the discussion, all participants completed a large number of self-rating personality questionnaires. In addition, peer ratings of personality were obtained by asking the participants to give a set of peer rating forms to each one of three acquaintances of the same sex, similar age and comparable social status. These rating forms were accompanied by a detailed letter of explanation, and peers were asked to return the completed questionnaires by mail directly to the investigator.

The two major rating forms used were a 35-item personality adjective check list and a five personality dimension rating form. Development, pretesting and reliability assessment for these scales are described in detail elsewhere (Scherer, 1970). Since both instruments correlated highly with each other, only the five personality dimension ratings are used in this study. This form consists of five 7-point scales to rate the degree of a person's conscientiousness, emotional stability, extroversion, assertiveness and agreeableness in relation to all people of the rater's acquaintance of the same sex, similar age and social status. To provide anchors for the judgment of a particular person, detailed descriptions of these dimensions, including examples of

how persons high or low on a particular dimension typically feel, act and relate to other people, were provided.

Since there seems to be little agreement on the general validity of self-ratings of personality, and since the earlier study (Scherer, 1972) found that the voice-based personality attribution of extroversion seemed to predict the peer rating criterion value more accurately than the self-rating criterion, only the former is used as trait criterion value in this paper.

### **Assessment of distal cues via indicator values**

Since the stimulus material presented to observers in this study is restricted to the auditory channel, and since the content-masking technique eliminates content and most sequential speech cues, only indicators for voice frequency-related distal cues need to be considered. In the present case, 'voice experts' were used to measure objectively the indicator values for these distal cues.

Six phoneticians rated the voices of the speakers from the 20 s content-masked voice sample, using the following indicators (7-point scales): (a) absolute pitch height (low, high), (b) pitch range (narrow, wide), (c) vocal effect or loudness (soft, loud), (d) dynamic (loudness) range (little, much), (e) preciseness of articulation (loose, precise), (f) breathiness (none, breathy), (g) creak (none, creaky), (h) glottal tension (open, tight), and (i) nasality (none, nasal). Detailed phonetic descriptions of these voice quality parameters, a description of the rating procedure, and detailed reports on the reliability of the expert judgments and the interrelationships between scales can be found in Scherer (1974a). The mean ratings of the six phoneticians were used as distal indicator values. The following effective reliabilities based on mean inter-rater correlations (Rosenthal, 1973) were found for these values: pitch height .86, pitch range .72, loudness .90, dynamic range .76, articulation .67, breathiness .67, creak .86, glottal tension .24, nasality .54. Since the reliability for the glottal tension rating is clearly unsatisfactory, this variable is not used in any of the further analyses.

### **Assessment of percepts via perceptual judgments**

Percepts are most difficult to measure via judgments or ratings, since the subjective report of the outcome of the perceptual mapping of a distal cue is by necessity subjected to categorization and linguistic coding and is thus most likely the result of an inference process. For example, a naïve listener is generally not able to specify his perceptual representation of the distal cue of creak. If pressed to describe a creaky voice he will most likely try to assign the sound of this voice to one of the voice quality categories for which his language provides a label, possibly calling it a 'harsh'-sounding voice. This problem of categorization and labelling is aggravated by the fact that for the purpose of standardization, labelled categories are provided by the researcher forcing the subject to sort his percepts into a prearranged array of categories. However, if there is some degree of correspondence between the objective, physical space of distal cue differentiation and gradation and the culturally and linguistically imposed category system of reporting the corresponding perceptual dimensions, a 'translation' in the form of representation coefficients becomes possible.

Ratings of voice quality percepts were obtained from a group of ten American undergraduate female volunteers. The judges were told that the study was concerned

with how well naïve listeners could identify different voice qualities. The ratings were made on 9-point scales for the following 12 voice qualities (effective reliabilities for the mean of the judges' ratings in parentheses) which had been found to represent strong clusters in a 35-item voice quality adjective rating form (cf. Scherer, 1970): resonance (.82), breathiness (.73), gloom (.81), thinness (.87), loudness (.86), sharpness (.80), warmth (.73), harshness (.77), high pitch (.91). The voice quality scales were explained, and two examples of the random-spliced voice samples were given to acquaint judges with the nature of the content-masked voice sample. The 24 stimuli were played back using high fidelity sound reproduction equipment. Judges heard each of the 20 s voice samples repeated in a loop until they had completed the respective ratings.

### **Assessment of the personality attributions**

The judges in the personality attribution rating group used the same five personality dimensions rating form described above for the peer ratings. The rating session was conducted in essentially the same way as the voice rating session, except that the judges were told that the purpose of the study was to find out how accurately personality could be judged from exposure to voice quality alone. The personality rating group consisted of nine American undergraduate female volunteers who were paid for their services. A money prize was promised and paid to the most accurate judge in the group. The following effective reliabilities for the mean of the nine judges' ratings for the 24 speakers were found: conscientiousness .53, emotional stability .76, extroversion .71, assertiveness .72, agreeableness .71.

## **RESULTS AND DISCUSSION**

### **Ecological validity of voice quality cues**

Association coefficients between personality criteria and voice quality indicator values were obtained by correlating peer ratings of personality and phoneticians' ratings of voice quality over 24 speakers. If a  $p < .05$  (two-tailed) significance level is set as the limit of acceptable ecological validity, only vocal effort, dynamic range, and nasality attain association coefficients of sufficient strength. This holds only for the traits of emotional stability (vocal effort  $r = .41$ , dynamic range  $r = .42$ ) and extraversion (vocal effort  $r = .43$ , dynamic range  $r = .37$ ,  $p = .07$ , nasality  $r = .43$ ). Thus, American speakers rated as emotionally stable and extroverted by their peers seem to speak with a louder and possibly more nasal voice, using a greater range of loudness variation. This may indicate a more forceful use of the voice in social interaction, and, possibly, a more frequent and skillful use of the vocal organs generally.

The environmental predictability of personality traits on the basis of the voice quality indicators used in this study can be assessed by multiple regression techniques regressing the criterion on the ecologically valid indicators. Predictability is fairly low with  $R^2 = .22$  ( $p < .10$ ) for emotional stability (two predictors) and  $R^2 = .35$  ( $p < .05$ ) for extroversion (three predictors). This does not exclude the possibility that the potential predictability, using additional vocal indicators and/or different assessment methods, may be much higher.



### **Perceptual representation of distal cues**

Representation coefficients were computed by correlating phoneticians' and naïve listener-judges' voice ratings. As mentioned above, the proximal representation of the acoustic voice quality cues cannot be assessed in terms of the same acoustical or articulatory parameters as the phoneticians' ratings. As for other social stimuli, the dimensionality of voice quality perception needs to be assessed by carefully designed multi-dimensional scaling methods. Since studies of this sort have yet to be conducted, popular voice quality labels culled from the literature (Scherer, 1970) were used in this study. Therefore, the judges' voice ratings may have consisted mainly in a categorization of their voice quality percepts according to the terminology imposed by the voice quality rating scales. Thus, the representation coefficients reported here indicate the probability with which certain labels are assigned to voices with particular acoustic characteristics.

The voice quality indicators that seem to have the strongest impact on proximal voice quality percepts are vocal effort and pitch height, the two most basic acoustic dimensions. The distal variable pitch height correlates with high pitch ( $r = .80$ ,  $p < .001$ ), sharpness ( $r = .82$ ,  $p < .001$ ), and gloom ( $r = -.60$ ,  $p < .01$ ). Vocal effort correlates with high pitch ( $r = .59$ ,  $p < .01$ ), sharpness ( $r = .62$ ,  $p < .01$ ), gloom ( $r = -.47$ ,  $p < .05$ ) and breathiness ( $r = -.44$ ,  $p < .05$ ). Thus, pitch and vocal effort seem to be represented by similar proximal percepts: loud voices and voices with relatively high pitch are seen as sharp, high pitched, not gloomy and not breathy. Of the remaining indicators, only preciseness of articulation and creak are proximally represented as percepts although the relationships are rather weak. Unprecise articulation leads to the perception of a gloomy voice ( $r = -.41$ ,  $p < .05$ ), a creaky voice is seen as harsh ( $r = .44$ ,  $p < .05$ ).

Even though it is difficult to evaluate the accuracy or adequacy of the proximal representation of the voice quality indicators due to the differences in categories and labels, pitch or fundamental frequency of the voice seem to be represented more adequately on a frequency related dimension than vocal effort is on an energy dimension. The effects of a louder voice seem to be perceived by naïve judges mostly in terms of the energy distribution in the voice spectrum (timbre) rather than the absolute amount of energy.

Thus, pitch of voice seems to be the most powerful distal determinant of the proximal representation of a speaker's voice by a listener. However, before a final conclusion on the relative power of distal voice cues to influence proximal percepts can be drawn, the differential salience of cues due to the exposure conditions needs to be assessed. It is possible that pitch was such a powerful cue in the present study because the random-splicing technique has masked or reduced the salience of other distal cues.

### **Attribution of personality from voice quality percepts**

Utilization coefficients are computed by correlating voice percepts and personality attribution ratings. There are two distinct patterns of voice cue utilization in the inference of different traits as shown in Table 1. Conscientiousness and emotional stability tend to be attributed when the voice is perceived as resonant, warm, and (for emotional stability only) low pitched. Extroversion and assertiveness, on the other hand, are inferred when the voice is labelled sharp and, to some extent, loud as well as lacking breathiness or gloom.

Table 1. Correlations between voice quality percepts and personality attributions (utilization coefficients)

Voice percepts	Personality attribution ratings				Agreeableness
	Conscientiousness	Emotional stability	Extroversion	Assertiveness	
Resonance	.44*	.69‡	.07	.09	.28
Breathiness	-.36	-.01	-.34	-.48*	.34
Gloom	-.22	.10	-.55‡	-.48*	.70‡
Thinness	-.21	-.63‡	.11	.20	-.56‡
Loudness	.09	.28	.28	.30	-.02
Sharpness	-.17	-.47*	.40	.50	-.78‡
Warmth	.48*	.75‡	.08	.02	.29
Harshness	-.58‡	-.25	-.12	-.04	.12
High pitch	-.23	-.61*	.29	.39	-.67‡

$N = 24$ ; \* $p < .05$ ; † $p < .01$ ; ‡ $p < .001$ , two-tailed

If one distinguished between more intrapersonal traits (conscientiousness and emotional stability) and more interpersonal traits of personality (in the sense of directly interaction-relevant behavioural dispositions, cf. Scherer, 1972, p. 207), attributions of intrapersonal traits seem to be based on voice frequency-related percepts, and interpersonal traits on voice energy-related percepts. It could be argued that these utilization patterns are just stereotypical, inaccurate inference habits which do not represent any ecologically valid associations. However, it was shown earlier in this paper that the loudness or vocal effort dimension may well be a valid indicator at least of the interpersonal trait of extroversion.

Other investigations suggest that fundamental frequency of the voice may be a valid indicator of affective disturbance. The author has shown, for example, that voice pitch tends to increase under emotional tension or stress (cf. Scherer, 1978). Thus the utilization of frequency information to attribute intrapersonal traits to a speaker may be seen as a transfer of an inferential strategy that may be functionally valid for the diagnosis of transient affect states across situations but may not be adequate, as seems to be the case from these data, to allow accurate inference of stable intrapersonal dispositions across speakers.

### Inferential policy of the judges

A multiple regression analysis was performed to assess the inferential policy of the group of judges. Using the four percept scales with the highest utilization coefficients for the respective trait as predictors, the following  $R^2$ s were obtained ( $p < .001$  in each case): conscientiousness .64, emotional stability .66, extroversion .46, assertiveness .55, agreeableness .69.

The judges' inferential policy seems to be highly linear, since between approximately 50 per cent and 70 per cent of the variance can be explained by four voice percept scales. This can be interpreted as showing that the personality attributions are strongly determined by the voice percepts and that there is little extraneous influence from other variables. This is probably mainly due to the fact that judges had virtually no access to other than voice cues given the nature of the content-masked stimulus material to which they were exposed.

### Accuracy of the personality attributions

The functional validity or accuracy of the trait inferences is determined by correlating the attribution ratings with the criteria (peer) ratings. As in the earlier study being replicated, a significant accuracy coefficient is found only for the trait of extroversion:  $r = .412$  ( $p = .023$ ,  $N = 24$ , one-tailed). In the earlier study (cf. Scherer, 1972) a coefficient of  $r = .61$  ( $p = .018$ ,  $N = 12$ , one-tailed), had been found. The null hypothesis that these two sample correlations are drawn from a common  $\rho$  (rho) could not be rejected ( $\chi^2 = .4695$ ) and a weighted estimate of  $\rho$  (rho) was computed, yielding  $r = .476$ . If similar correlations can be found in further studies, one may conclude that about a quarter of the variance in peer ratings of extroversion is reflected in extroversion attributions based on voice quality cues only. Since a different, much younger group of female judges and twice the number of stimulus persons were used in this study, the phenomenon seems rather stable.

This conclusion is strengthened by findings in a study with 15 stimulus persons, drawn from the same pool of speakers, using shorter random-spliced voice samples from a different part of the discussion. Even though the accuracy coefficients do not quite reach significance, they point in the same direction (Scherer *et al.*, 1977). Evidence that accurate attribution of extroversion from the voices of American speakers is not restricted to American judges is accumulating: both German and French judges have been shown to achieve significant accuracy coefficients for this trait. In a recent unpublished study with ten French judges conducted in Paris using 12 stimulus speakers and a similar rating procedure, an accuracy correlation of  $r = .71$  ( $p < .005$ ,  $N = 12$ ) was found. As reported in Scherer (1972, p. 200), an accuracy correlation of  $r = .55$  ( $p < .05$ ,  $N = 12$ ) was found for eight German judges rating the same 12 speakers on sociability (using an adjective rating form).

The general conclusion offered by this chain of results is that extroversion can be inferred accurately because (a) there are ecologically valid voice indicator cues associated with the extroversion criterion, (b) these distal cues are adequately represented proximally by voice quality percepts, and (c) the utilization coefficients for these percepts correspond to the ecological association coefficients. Obviously, if not all of these requirements are met, as is the case with all other traits, no functionally valid or accurate attributions can result. One very likely 'missing link' in this case is obviously the lack of ecological associations between trait criteria other than extroversion and specific voice quality indicators. It is possible to postulate a 'value-added' approach with criteria-indicator association as the basis, adequate proximal representation and valid utilization coefficients as added requirements, each building upon the preceding link.

It seems necessary to adopt an inference process paradigm of this kind, with its resulting methodological requirements, if the problem of accuracy in personality judgments from nonverbal cues is to be fruitfully studied. Little is gained by showing that criteria and attributions do or do not correlate, since the cause for a possible lack of accuracy cannot be ascertained. Rather than blaming the judges, one may have to conclude that particular traits cannot possibly be accurately inferred from some types of nonverbal behaviour, since they do not have valid indicators in that particular channel or modality.

# Empirical tests of the general inference model

Since all the components of the general inference model described in the introduction have been operationally defined and empirically measured in the course of the present study, it is now possible to determine the adequacy of the model in explaining the pattern of results.

Figure 2 shows a 'concrete' version of the model applied to the inference of extroversion from voice quality. The choice of the variables to be included in this version of the model was based on the size of the correlations with other variables in 'adjoining' components of the models. The figure shows that the extroversion criterion is associated, even though not very strongly, with vocal effort, dynamic range, and nasality in the voices of the speakers. These distal indicator cues are represented proximally as the percepts of loudness and sharpness and lack of breathiness and gloom. These percepts in turn determine to some extent the extroversion attributions. Since the utilization coefficients mirror the direction of the association coefficients, some degree of accuracy in the inference of extroversion, attenuated by the noise in the transmission system (reflected in the probabilistic relationship between model components) results, as shown above.

In the case of emotional stability, shown in Figure 3, attributions are not accurate, since they seem mostly based on the speaker's pitch height, which is apparently not a valid indicator of the *habitual* personality trait of emotional stability in the peer rating criterion. Vocal effort, which does have some ecological validity, does affect the voice percepts of high pitch and sharpness. However, the utilization coefficients for those percepts point in a direction opposite to the one for the association coefficients, indicating that a valid and properly perceived indicator is misinterpreted in terms of its significance.

It seems quite possible that judges utilized pitch cues in a way which may be appropriate and functionally valid in inferring *situational states* of emotional tension. As shown above, emotional tension associated with a stressful situation in which deception is required of the subjects, leads to a significant increase in voice pitch (cf. Scherer, 1978). Apparently, cue utilization which may be highly functionally valid in inferring transient or situational states may not be helpful, and may even be positively misleading, in inferring stable or habitual traits.

Even though concrete versions of the Brunswikian inference model showing the correlations between the components of the model may be of considerable heuristic value, a more formal test procedure is clearly called for since many relevant factors, such as the intercorrelations among variables within components, are not considered. As mentioned in the introduction, one obvious adequacy criterion is the extent to which the components of the model, represented by the variables measured in a test of a 'concrete' version, can account for the variance in the attribution rating. This can be assessed by regressing the attribution rating on all other variables in the concrete version of the model to obtain  $R^2$ , the proportion of the variance accounted for by those variables. However, the use of multiple regression in this context requires specific procedures for entering predictors in the regression equation if a test of the model is to be feasible.

The regression of the attribution rating ( $y_A$ ) on percept ( $x_P$ ), distal indicator ( $x_D$ ) and criterion ratings ( $x_C$ ) can be represented as follow:

$$y_A = \alpha + \beta_{P1}x_{P1} + \dots \beta_{Pk}x_{Pk} + \beta_{D1}x_{D1} + \dots \beta_{Dk}x_{Dk} + \beta_Cx_C + \epsilon$$

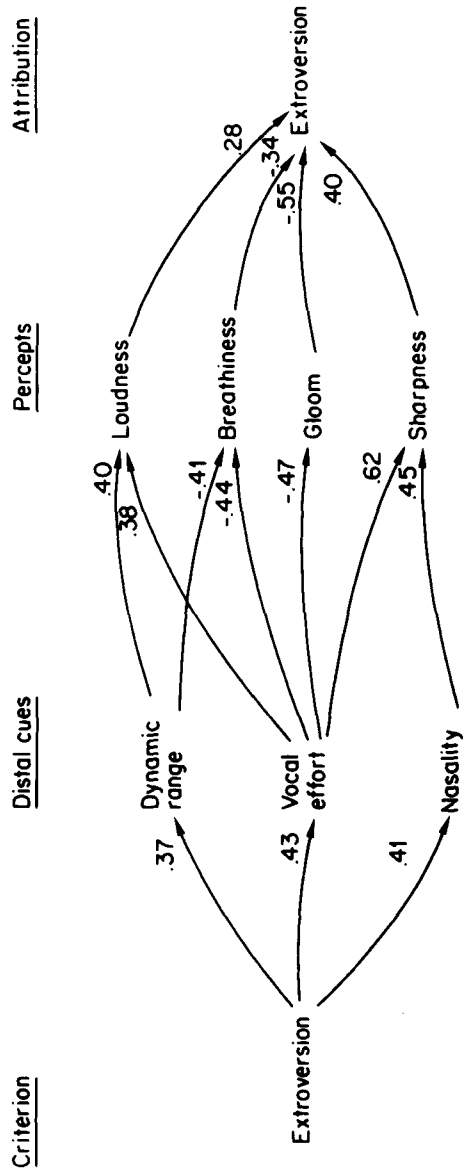


Figure 2. Process model for the inference of extroversion\*

\*Curvilinear arrows indicate hypothetically assumed causal relationships. Strength of the relationships is indicated by Pearson's *r*s. Intercorrelations among variables for each type of measurement have been omitted to simplify the Figure.

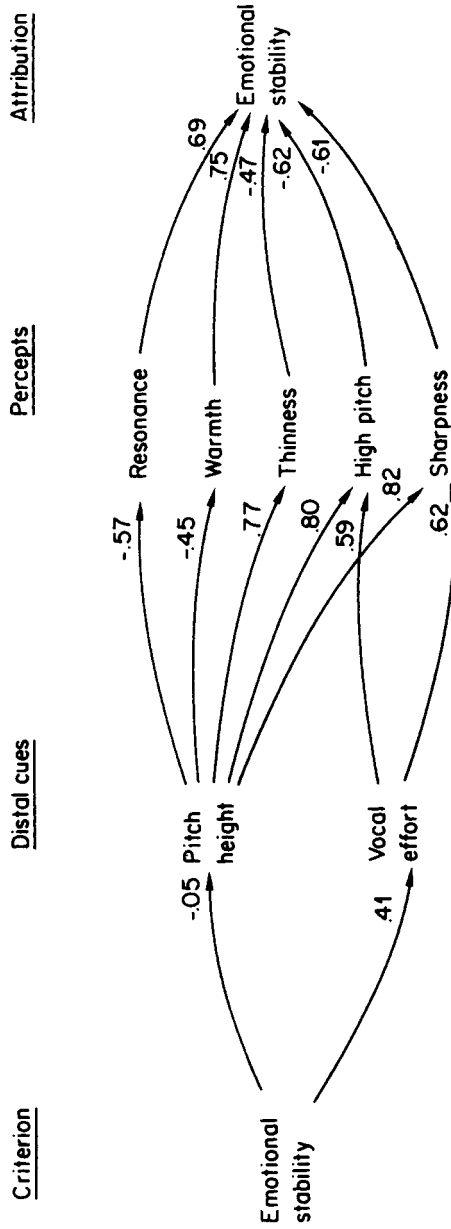


Figure 3. Process model for the inference of emotional stability\*

\*Curvilinear arrows indicate hypothetically assumed causal relationships. Strength of the relationships is indicated by Pearson's  $r$ s. Intercorrelations among variables for each type of measurement have been omitted to simplify the Figure.

This formalization allows the identification of two important criteria for the goodness of the model: (a)  $R$ , the multiple correlation  $r_{y_A \hat{y}_A}$ , should be high compared with the residual or error component  $\varepsilon$ , and (b) the predictor weights  $\beta_D$  and  $\beta_C$  should be low in comparison to the percept weights  $\beta_P$ . The latter criterion is based on the assumption that, barring telepathy or extrasensory perception, traits and distal cues *must* be proximally represented by percepts if they are to systematically influence attributions. Consequently, high  $\beta_D$ 's point to the fact that distal cues affect attribution without their proximal representation having been identified, measured and entered into the model. Similarly, a high  $\beta_C$  indicates that neither distal indicators nor their proximal representations for a trait which is inferred correctly have been identified in the model.

Therefore, a *series* of regression analyses is required in which the attribution is first regressed on the percepts only ( $\hat{y}_{A.P}$ ,  $R^2_{A.P}$ ), then on the percepts and distal cues, in that order ( $\hat{y}_{A.PD}$ ,  $R^2_{A.PD}$ ), and finally on percepts, distal cues, and criterion ( $\hat{y}_{A.PDC}$ ,  $R^2_{A.PDC}$ ). The goodness of the model in terms of accounting for the inferential process is reflected in the slope of the increase in  $R^2$  which should be a negatively accelerated function approaching an asymptote. The results of such a series of regression analyses for the voice and personality ratings are as follows: Extroversion  $R^2_{A.P} = .44$  ( $p < .05$ ),  $R^2_{A.PD} = .51$  (NS),  $R^2_{A.PDC} = .52$  (NS); Emotional stability  $R^2_{A.P} = .59$  ( $p < .01$ ),  $R^2_{A.PD} = .62$  ( $p < .05$ ),  $R^2_{A.PDC} = .63$  ( $p < .05$ ).

Clearly, the concrete model for the emotional stability inference fares better in the light of the criteria outlined above since the  $R^2$  for extroversion is generally lower and since there is a discontinuity in the slope of the  $R^2$  increase. The point at which such discontinuities or steps occur helps to pinpoint the weaknesses of the respective model since all components entered into the regression model before the  $R^2$  jump occurs are inadequately represented by definition.

This regression approach can obviously be refined by using the step-wise regression procedures for the variables within components to obtain the smallest set of (hopefully fairly orthogonal) predictor variables producing a value of  $R^2$  close to the one which can be attained if all variables within that component are entered into the regression equation. However, the general multiple regression paradigm is not optimally suited to test this model. A more complex evaluation procedure is needed to assess how well the respective version of the model can explain the functional validity of the attribution, i.e. the accuracy coefficient. The present author has suggested earlier (Scherer, 1974b) to use path-analytic procedures for this purpose since a causal chain of effects of sets of variables is theoretically postulated. Departing from the central theorem of path analysis (Duncan, 1966, p. 5)

$$r_{iy} = \sum p_{iq} r_{qy}$$

the empirically found correlation ( $r_{AC}$ ) between attribution and criterion is split into the contributions of the direct and indirect paths from C to A, where the index  $q$  runs over all intermediate variables in the model (in this case distal cues D and percepts P). This is demonstrated in Figure 4 where a simple one-dimensional model of the extroversion inference is shown.

The path coefficients indicating the strength of the relationships in Figure 4 are obtained by regressing gloom (P) on vocal effort and the extroversion criterion (in that order) and the extroversion attribution on gloom, vocal effort, and the criterion (in that order). Thus, the double arrows represent the paths which are theoretically

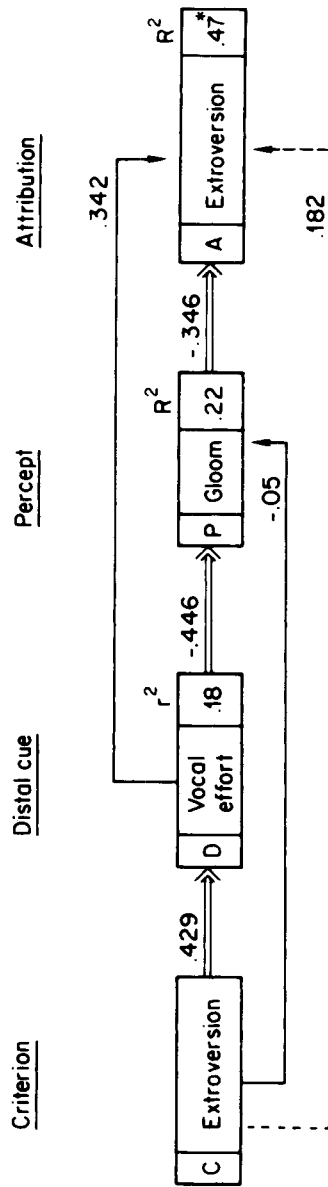


Figure 4. One-dimensional path analysis model for the extroversion inference

\* $p < .05$ .

† The dashed line represents the direct path, the double line the postulated indirect paths, and the single line the indirect paths compatible with the model. Coefficients shown are standardized coefficients except for  $r_{CD}$  which is a Pearson  $r$  (the direction shown is theoretically postulated).  $R^2$ s based on all predictors from which paths lead to the variable.



postulated to be the central mediators of the relationship. The single arrows represent paths which, though compatible with the model, are peripheral in the sense that they 'jump' or bypass the distal cues or percept components. This indicates (if their path coefficients are high) that the respective component has not yet been exhaustively operationalized in terms of empirically assessed variables. The dashed arrow represents the direct effect of the criterion on the attribution after the effect due to the postulated intermediate variables has been removed.

It can be easily shown from Figure 4 that the accuracy coefficient can be split into the contributions of the postulated central indirect paths ( $r_{CD}p_{PD}p_{AP}$ ), the peripheral indirect paths ( $r_{CD}p_{AD}$  — distally based, bypassing percept component, and  $p_{PC}p_{AP}$  — proximally based, bypassing the distal cue component), and the direct path ( $p_{AC}$ ):

Accuracy coefficient		Peripheral effects			Direct effect
		Central effect (postulated)	distally based	proximally based	
$r_{CA}$	=	$r_{CD}p_{PD}p_{AP}$	+ $r_{CD}p_{AD}$	+ $p_{PC}p_{AP}$	+ $p_{AC}$
.412	=	.066	+ .147	+ .017	+ .182
(100%)	=	16%	+ 36%	+ 4%	+ 44%

This conceptualization allows the specification of several requirements to be met if a specific model is to adequately explain an accuracy coefficient.

The coefficient for the direct path,  $p_{AC}$ , should be close to zero or at least negligible compared to the strength of the central and peripheral paths. If it is not, the relationship between criterion and attribution is mediated through variables not entered into the model and thus cannot be adequately accounted for.

The optimal state of affairs would of course be

$$r_{CA} \approx r_{CD}p_{PD}p_{AP}.$$

However, since in practice it is unlikely that both distal cues and percepts can be exhaustively measured it may be a more realistic aim to obtain a reasonable proportion between the central and the peripheral effects,

$$r_{CD}p_{PD}p_{AP} > r_{CD}p_{AD} + p_{PC}p_{AP}$$

being desirable.

The relative strength of the two types of peripheral paths indicate where efforts to improve the model are appropriate: if a strong path 'jumps' or bypasses the percept component, further variables in this domain ought to be measured since a distal cue is clearly affecting the organism's attribution. If the distal cue component is bypassed by a strong path, a search for further distal indicators and ways of measuring them is called for, since the observers must make use of externalized cues of the trait which they are able to infer with better than chance accuracy.

Figure 5 shows an example of an expanded model of the extroversion inference process in which the frequency dimension has been added to the energy dimension. The accuracy coefficient is now split into contributions made up by the sums of the

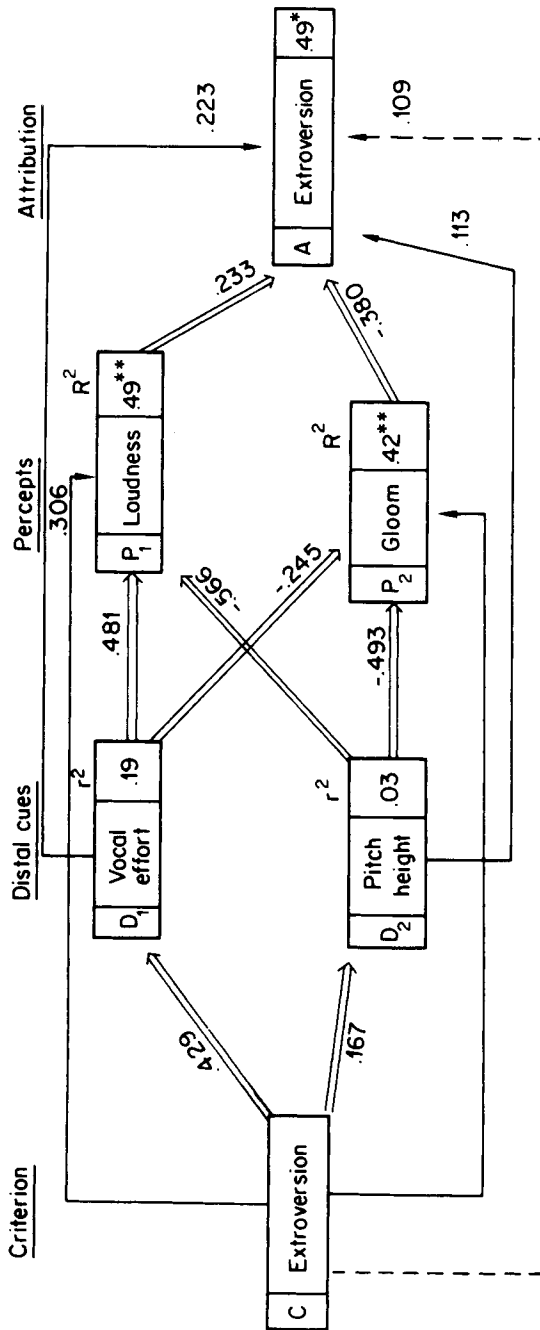


Figure 5. Two-dimensional path analysis model for the extroversion inference

\* $p < .05$  \*\* $p < .01$

†The dashed line represents the direct path, the double line the postulated indirect paths, and the single lines the indirect paths compatible with the model. Coefficients shown are standardized coefficients except for  $r^2_{CD}$  and  $r^2_{CD}$ , which are Pearson  $r$ s (the direction shown is theoretically postulated).  $R^2$ s based on all predictors from which paths lead to the variable.

products of the various path coefficients. There are four terms for the effect of the central paths

$$(r_{CD1}p_{P1D1}p_{AP1}; r_{CD2}p_{P2D2}p_{AP2}; r_{CD1}p_{P2D1}p_{AP2}; r_{CD2}p_{P1D2}p_{AP1}),$$

since each distal cue can affect each of the percepts, and two terms each for the peripheral paths. If the path coefficients are multiplied and summed accordingly, the following separation of effects result:

Accuracy coefficient	Central effects (postulated)	Peripheral effects		Direct effect
		distally based	proximally based	
$r_{CA}$	$= \sum_{i,j=1}^2 r_{CDi}p_{PjDi}p_{APj}$	$+ \sum_{i=1}^2 r_{CDi}p_{ADi}$	$+ \sum_{j=1}^2 p_{PjC}p_{APj}$	$+ p_{AC}$
.412	= .097	+ .091	+ .115	+ .109
(100%	= 24%	+ 22%	+ 28%	+ 26%)

Compared to the one-dimensional version of the model a greater proportion of the accuracy coefficient is now explained through model mechanisms since the size of the direct effect has been reduced. Even though the size of the centrally mediated effect has also increased, there are still rather sizeable peripheral effects indicating the possibility to include still further percept variables and distal cue variables in the model to further increase the strength of the central effect. It should be noted that the size of the  $R^2$  is quite independent of the degree to which the accuracy coefficient can be accounted for by theoretically postulated mechanisms. Thus the  $R^2$  increases only marginally (.47 to .49) from the one- to the two-dimensional model. This is due to the fact that some of the variance directly accounted for by the criterion (and/or the distal cues) in the first version of the model is now accounted for by an intermediate variable which increases the explanatory (in terms of effect mediation), but not the predictive power (in terms of variance accounted for) of the model.

One may ask whether the path-analytic procedure suggested above serves any useful purpose if the accuracy coefficient is insignificant and there is no relationship to explain. Clearly, a path-analytic model is preferable to the heuristic models shown in Figures 2 and 3, since the intercorrelations among the variables are included in the assessment of the strength of the different effects. As in the case of a significant accuracy coefficient, one would want the model to be as accurate a representation of the inference process as possible. Thus one could identify points in the model where there are strong peripheral paths indicating the need to enter further intermediate variables.

A path-analytic model can also be used to pinpoint the region of the model responsible for the lack of accuracy by assessing the congruence of partial path coefficients (both in size and direction) which make up a common postulated effect. Of particular importance is the possibility of finding suppressor effects indicating that a more accurate inference could result if it were not for the misperception or misinterpretation of some cues. Such suppressor effects can be identified in the case where there is a high direct effect ( $p_{AC}$ ) and strongly negative indirect effects. The identification of such dysfunctional suppressor effects may open possibilities to

improve inferential strategies of observers in the sense of increasing the functional validity of their cue utilization policy.

## CONCLUSION

In this study an attempt was made to apply a Brunswikian lens model to the inference of personality from voice. All components of the model were operationalized and empirically assessed to obtain estimates of trait externalization (in the form of indicator cues), of cue perception and cue utilization as well as of the accuracy or functional validity of the resulting attribution. The data point to the conclusion that the extroversion attribution from voice quality attains some degree of functional validity because the trait criterion is externalized in distal cues related to voice energy, mainly vocal effort and dynamic range, and because the proximal representations of these cues are utilized in a corresponding fashion to arrive at the extroversion attribution.

These results replicate the earlier finding that extroversion can be judged from voice quality (at least for male American speakers and female observers regardless of cultural origin). In addition, the inference processes involved have started to emerge more clearly. It has been shown that path-analytic procedures provide a promising methodology to formally test the adequacy of specific versions of the lens model to particular domains in person perception and nonverbal communication.

Similar methods can be used to assess the inference processes involved in attributions of personality traits, emotional states, and interpersonal attitudes from a wide variety of verbal and nonverbal cues. Even though this kind of modelling and its methodological requirements are more cumbersome and time consuming than straightforward studies confined to the correlation between criteria and attributions, the information on the nature of the inference process gained seems to be well worth the effort.

One of the major pay-offs of this approach should be a tendency toward a greater consistency of research results in this area of interpersonal perception and communication. Since the inference process is broken down into its components it should be possible to assess the specific causes of failures to replicate earlier results, such as differences in cue utilization strategies between groups of judges (cf. Wiggins, 1973), lack of proximal representation of ecologically valid indicators (possibly due to methodological factors) etc.

The process modelling approach as advocated in this paper can be easily combined with a 'cue-isolation-and-integration' methodology. Indicator cues associated with traits and states should be isolated by masking or selective exposure techniques (just as random-splicing of the speech sequence has been used here to isolate voice quality cues) in order to assess 'pure' utilization coefficients for such cues unaffected by jointly occurring covariation or independent cues. In a second step, these isolated cues can be systematically paired in different combinations to assess the observers' cue utilization strategies in more complex stimulus situations (cf. Scherer, 1977). As in studies of clinical judgment, the question of configurational or linear cue utilization, the role of subjective composition rules (cf. Hoffman, 1968; Slovic and Lichtenstein, 1973) needs to be explored. The effect of the type of inferential strategy used is obviously not independent of 'objective composition rules' in terms of information specificity or

generality of different types of cues. It has been shown earlier in this paper (cf. also Scherer *et al.*, 1977) that additivity or suppression effects can result depending on the degree of congruence in the information value of jointly utilized cues.

The question of information processing and the integration of discrepant cues has always been a major concern of cognitive social psychology (cf. Anderson, 1971; Hastorf *et al.*, 1970). However, in many of the relevant studies verbal stimuli are used exclusively as information input. Even though the results obtained in these studies may be highly representative for the cognitive processing of semantic information, they may not at all capture the nature of inference processes from nonverbal behavioural cues. The latter may, after all, be more typical of daily life than inferences from lists of adjectives. The scarcity of studies in this area shows that Brunswik's (1956) call for representative designs in psychology has remained unheard so far.

## REFERENCES

- Anderson, N. H. (1971). 'Integration theory and attitude change', *Psychological Review*, **78**: 171-205.
- Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*, University of California Press, Berkeley.
- Dudycha, L. W. and Naylor, J. C. (1966). 'Characteristics of the human inference process in complex choice behavior situations', *Organizational Behavior in Human Performance*, **1**: 110-128.
- Duncan, O. D. (1966). 'Path analysis: Sociological examples', *American Journal of Sociology*, **72**: 1-16.
- Hastorf, A. H., Polefka, J. and Schneider, D. J. (1970). *Person Perception*, Addison Wesley, Reading, Mass.
- Hoffman, P. J. (1968). 'Cue consistency and configurality in human judgment'. In: Kleinmuntz, B. (Ed.), *Formal Representation of Human Judgment*, Wiley, New York. pp. 53-90.
- Hunt, R. G. and Lin, T. K. (1967). 'Accuracy of judgments of personal attributes from speech', *Journal of Personality and Social Psychology*, **6**: 450-453.
- Kramer, E. (1963). 'The judgment of personal characteristics and emotions from nonverbal properties of speech', *Psychological Bulletin*, **60**: 408-420.
- Markel, N. N., Meisels, M. and Houck, J. E. (1964). 'Judging personality from voice quality', *Journal of Abnormal and Social Psychology*, **69**: 458-463.
- Moses, P. J. (1954). *The Voice of Neurosis*, Grune and Stratton, New York.
- Rosenthal, R. (1973). 'Estimating effective reliabilities in studies that employ judges' ratings', *Journal of Clinical Psychology*, **29**: 342-345.
- Scherer, K. R. (1970). 'Attribution of personality from voice: a cross-cultural study on the dynamics of interpersonal perception', *Unpublished doctoral dissertation*, Harvard University.
- Scherer, K. R. (1971). 'Randomized splicing: A note on a simple technique for masking speech content', *Journal of Experimental Research in Personality*, **5**: 155-159.
- Scherer, K. R. (1972). 'Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception', *Journal of Personality*, **40**: 191-210.
- Scherer, K. R. (1974a). 'Voice quality analysis of American and German speakers', *Journal of Psycholinguistic Research*, **3**: 281-298.
- Scherer, K. R. (1974b). 'Persoenlichkeit, Stimmqualitaet und Persoenlichkeitsattribution: Pfadanalytische Untersuchungen zu nonverbalen Kommunikationsprozessen', *Bericht ueber den 28. Kongress der Deutschen Gesellschaft fuer Psychologie*, Vol. 3, Hogrefe, Göttingen, pp. 61-73.
- Scherer, K. R. (1978). 'Non-linguistic vocal indicators of emotion and psychopathology'. In: Izard, C. E. (Ed.), *Emotions in Personality and Psychopathology*, Plenum Press, New York.
- Scherer, K. R. and Oshinsky, J. (1977). 'Cue utilization in emotion attribution from auditory stimuli', *Motivation and Emotion*, **1**: 331-346.

- Schere, K. R., Scherer, U., Hall, J. A. and Rosenthal, R. (1977). 'Differential attribution of personality based on multi-channel presentation of verbal and nonverbal cues', *Psychological Research*, 39: 221-247.
- Slovic, P. and Lichtenstein, S. (1973). 'Comparison of Bayesian and regression approaches to the study of information processing in judgment'. In: Rappaport, L. and Summers, D. A. (Eds.), *Human Judgment and Social Interaction*, Holt, Rinehart and Winston, New York.
- Wiggins, N. (1973). 'Individual differences in human judgments: a multivariate approach'. In: Rappaport, L. and Summers, D. A. (Eds.), *Human Judgment and Social Interaction*, Holt, Rinehart and Winston, New York.

## RÉSUMÉ

On a étudié les processus d'inférence concernant la personnalité, inférence fondée sur la qualité de la voix de 24 personnes-stimuli (américains de sexe masculin) dont on enregistrerait les discussions dans des simulacres de jury. On a appliqué le modèle de 'lentille' (lens model) de E. Brunswik pour l'étude de l'inférence et on a fait évaluer par des groupes de juges indépendants: les critères (portant sur la personnalité), les indices distaux, les indices proximaux et les imputations (attributions); les critères de personnalité par trois camarades de chaque personne — stimulus; les indices distaux de la qualité de la voix par six phonéticiens qui écoutaient des échantillons de voix sans pouvoir comprendre le contenu, les percepts proximaux de la voix par dix juges naïfs, les imputations quant à la personnalité par neuf juges naïfs. Seules les imputations d'extraversion corrèlent de manière significative avec le critère (jugements par les pairs) ce qui confirme des résultats antérieurs. Les autres traits apparemment ne peuvent être inférés de manière précise à partir de la qualité de la voix; en ce qui concerne l'extraversion les conditions suivantes doivent être réalisées: a) le critère est associé aux indices écologiquement valides d'énergie de la voix (effort vocal et registre de l'énergie de la voix), b) ces indices sont représentés de manière adéquate comme des percepts proximaux de la voix (en particulier la force et la netteté) et c) l'utilisation des percepts dans la stratégie d'inférence des juges correspond à l'association entre le critère et les indices distaux. On a utilisé la méthode de la 'path-analysis' pour tester de manière empirique la validité du modèle d'inférence: a) pour rendre compte de la variance dans les imputations et b) pour expliquer les corrélations significatives entre les critères et les imputations en termes de variables intermédiaires dans le modèle de 'lentille'.

## ZUSAMMENFASSUNG

Untersucht wurde der Prozeß, mit dem aus der Stimmqualität auf die Persönlichkeit geschlossen wird; 24 männliche Amerikaner dienten in simulierten Jurydiskussionen als Stimuluspersonen. Unter Bezug auf das Brunswiksche Linsenmodell des Inferenzprozesses wurden Kriterien, distale und proximale Hinweisreize und Attributionen durch unabhängige Gruppen von Beurteilern gemessen: die Persönlichkeitskriterien durch drei Bekannte jeder Stimulusperson; die distalen Hinweisreize auf der Basis von Stimmbeispielen, bei denen der Inhalt verdeckt blieb, durch sechs Phonetiker; die proximale Stimmwahrnehmung durch zehn naive und die Persönlichkeitsattributionen durch neun naive Beurteiler. Lediglich die Attributionen von Extraversion korrelieren signifikant mit dem Kriterium, womit frühere Befunde repliziert wurden. Wird Extraversion erschlossen, so werden anders als bei anderen Merkmalen, die sich offensichtlich nicht genau aus der Stimmqualität erschließen lassen, folgende Bedingungen erfüllt: (a) das Kriterium ist mit ökologisch validen Hinweisreizen der Stimmenergie (stimmlicher Aufwand und dynamischer Bereich) verknüpft, (b) diese Hinweisreize werden angemessen in proximalen Stimmperezepten (insbesondere Lautstärke und Schärfe) repräsentiert, und (c) der Perzeptgebrauch in der Schlußstrategie der Beurteiler entspricht der Verknüpfung von Kriterium und distalen Hinweisreizen. Pfadanalytische Methoden werden benutzt, um empirisch zu prüfen, wie das Inferenzmodell (a) die Varianz in den Zuschreibungen deutet und (b) signifikante Korrelationen zwischen Kriterien und Attributionen anhand vermittelnder Variablen erklärt.