# DO MEN AND WOMEN USE A COMMON SEMANTIC FACTOR SPACE TO DESCRIBE VOICE AND PRONUNCIATION?

Leo W.A. van Herpt

## 1.0 INTRODUCTION

When people give judgments about the interrelationships of voice characteristics their inferences are apparently guided by a more or less implicit theory with relation to the human voice. In an earlier paper (van Herpt, 1986:23) data are given which assert the assumption that raters generally use the same underlying theoretical structure when judging both the female and the male Voice and Pronunciation.
Between the raters too there is a reasonable conformity. Especially within social groups there is a great deal of agreement as to what inferences can be made concerning the structure of a concept. The development of these agreed-upon assumptions is promoted by common experiences. Yet, clearly, individuals do not always agree. For many women a voice must be soft to be cultured, whereas for many men a soft voice is associated with (physical) weakness. This suggests that, although raters belonging to the same social group do describe female and male speakers in a common semantic space, different social groups might use different spaces.
In this paper we inquire as to whether difference in sex of rater affects their perceptions and leads to different meanings of voice traits for female and male listeners. The methodological procedures are emphasized below because several techniques have been used unorthodoxly.

## 2.0 CONSIDERATIONS ON DESIGN

In the present experiment we use the comprised form of a test which we developed for the perceptual description of voice and pronunciation of normal Dutch speakers by naive listeners (Fagel, van Herpt & Boves, 1983). The instrument is based on a multivariate description of the object in terms of a limited number of semantic scales which describe Voice and Pronunciation (V&P) in a reasonable stable perceptual space of five orthogonal dimensions. The comprised form of the V&P-test and its dimensions are given in table 1.
In perceptual rating procedures of this type the subjects (listeners) are part of the instrument for assessing differences between speakers on semantic scales. This implies that listeners contribute to error of the instrument. Usually this error is assumed to be negligible, so it is common practice to take mean scores over subject replications to eliminate their variance. It is open to question if this procedure is justified. In an earlier experiment (van Herpt, 1986) we found indications that systematic distortions in the scores of the raters occur. As suggested above, one obvious although insufficiently investigated cause of this bias might be that female and male raters differ in their semantic factor structure.
We performed an analysis of the factor structure of the scales in a design which itself eliminates speaker variance. The concepts to be rated are the qualifier terms themselves, so no acoustical signals are presented. This method eliminates speaker differences entirely since merely the qualifiers with relation to the general notion 'average voice' are given for judgment. By taking the mean scores over listeners the rater variance is eliminated too.

So the resulting factorial structure is solely brought about by the scales and can neither be attributed to a particular sample of voices nor to rater variance.In fact we performed two independent experiments: one with relation to the 'average female voice', the other one concerning the 'average male voice'. This enables us to verify whether male and female speakers are described in a common factor space.

## 3.0 AIMS

The principal aim of the experiment is to inquire whether our qualifiers of voice and pronunciation have the same meaning for men and women or, in other words, whether male and female raters use a common semantic framework in the description of voice and pronunciation.
Three auxiliary issues are considered.
(1) Do raters use the the same underlying theoretical structure when judging respectively the female and the male voice and pronunciation? More specifically we ask whether factor analyses will show the same factor structure when speakers of different sex are judged.
(2) Earlier analyses generally showed a perceptual space of five significant dimensions. There is a possibility that Dimension I (Voice Appreciation) and III (Voice Quality) can further be broken down. Is sex of rater or speaker a variable in this connection?
(3) Are particular scales identifiable which influence negatively the consistency of the factor structure under varying subject samples?

Table 1 - Scales and dimensions of the shortened V & P rating form.

| Scale nr. | Scale terms | | Dimension | |
|---|---|---|---|---|
| Sc01<br>Sc02 | monotonous<br>expressionless | - melodious<br>- expressive | Ia. | Voice Appreciation: Melodiousness |
| Sc13<br>Sc14 | ugly<br>unpleasant | - beautiful<br>- pleasant | Ib. | Voice Appreciation: Evaluation |
| Sc03<br>Sc04 | broad<br>slovenly | - cultured<br>- polished | II. | Articulation Quality |
| Sc05<br>Sc06 | dull<br>husky | - clear<br>- not husky | IIIa | Voice Quality: Clarity |
| Sc07<br>Sc08 | weak<br>soft | - powerful<br>- loud | IIIb. | Voice Quality: Subjective Strength |
| Sc09<br>Sc10 | shrill<br>high | - deep<br>- low | IV. | Pitch |
| Sc11<br>Sc12 | dragging<br>slow | - brisk<br>- quick | V. | Tempo |

## 4.0 PROCEDURE

The method, after an idea used by Osgood and Suci (1955:332) aims at the extraction of factors from which speaker variance is eliminated. The subjects are simply asked for their opinion concerning the relations between the scales by having them judge each of the scales against the thirteen remaining attribute scales without presenting any speech. E.g. subjects were to rate the concept "loud" male voice on the other 13 seven-point scales. (Another group of subjects rated the qualifiers in relation to the female voice and pronunciation.) So the concepts in this analysis are the qualifier terms themselves. Since only one of the qualifiers of a scale is rated the stimulus list consists of $(13*14)/2 = 91$ items. (A description in detail of the testing procedure is given in van Herpt 1986:21 ff.)
Since the subjects were required to indicate in what direction and measure the 'concept qualifier' is related to the scale terms, the method directly yields results which reflect the degree of similarity in the pattern of interrelatedness. E.g. if 100% of the subjects

Table 2 - Interscale relation coefficients*) equal to or greater than 0.60 for the complete sample and its four cross-sections of female (♀) and male (♂) Voice and Pronunciation and female (F) and male (M) raters.

| Sample | Scale | Sc01 | Sc02 | Sc03 | Sc04 | Sc05 | Sc06 | Sc07 | Sc09 | Sc11 | Sc13 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| MF♂♀ | Sc02 | 77 | | | | | | | | | |
| MF♂ | | 77 | | | | | | | | | |
| MF♀ | | 78 | | | | | | | | | |
| M ♂♀ | | 65 | | | | | | | | | |
| F ♂♀ | | 89 | | | | | | | | | |
| MF♂♀ | Sc03 | | - | | | | | | | | |
| MF♂ | | | - | | | | | | | | |
| MF♀ | | | - | | | | | | | | |
| M ♂♀ | | | - | | | | | | | | |
| F ♂♀ | | | 69 | | | | | | | | |
| MF♂♀ | Sc04 | | | 63 | | | | | | | |
| MF♂ | | | | 60 | | | | | | | |
| MF♀ | | | | 67 | | | | | | | |
| M ♂♀ | | | | - | | | | | | | |
| F ♂♀ | | | | 79 | | | | | | | |
| MF♂♀ | Sc06 | | | | | 77 | | | | | |
| MF♂ | | | | | | 74 | | | | | |
| MF♀ | | | | | | 80 | | | | | |
| M ♂♀ | | | | | | 60 | | | | | |
| F ♂♀ | | | | | | 91 | | | | | |
| MF♂♀ | Sc07 | | - | | | 61 | | | | | |
| MF♂ | | | 62 | | | 60 | | | | | |
| MF♀ | | | - | | | 62 | | | | | |
| M ♂♀ | | | - | | | 60 | | | | | |
| F ♂♀ | | | 62 | | | 78 | | | | | |
| MF♂♀ | Sc08 | | | | | | - | 69 | | | |
| MF♂ | | | | | | | - | 73 | | | |
| MF♀ | | | | | | | - | 64 | | | |
| M ♂♀ | | | | | | | - | - | | | |
| F ♂♀ | | | | | | | 70 | 78 | | | |
| MF♂♀ | Sc10 | | | | | | | | 77 | | |
| MF♂ | | | | | | | | | 82 | | |
| MF♀ | | | | | | | | | 73 | | |
| M ♂♀ | | | | | | | | | 73 | | |
| F ♂♀ | | | | | | | | | 81 | | |
| MF♂♀ | Sc12 | | | | | | | | | 64 | |
| MF♂ | | | | | | | | | | 73 | |
| MF♀ | | | | | | | | | | - | |
| M ♂♀ | | | | | | | | | | 73 | |
| F ♂♀ | | | | | | | | | | - | |
| MF♂♀ | Sc13 | 79 | 77 | 74 | | 67 | | | | 74 | |
| MF♂ | | 78 | 72 | 73 | | 70 | | | | 74 | |
| MF♀ | | 80 | 82 | 76 | | 64 | | | | 73 | |
| M ♂♀ | | 74 | 62 | 65 | | (59) | | | | 66 | |
| F ♂♀ | | 85 | 90 | 86 | 74 | 75 | | | | 83 | |
| MF♂♀ | Sc14 | 82 | 74 | 76 | | 75 | | | | | 90 |
| MF♂ | | 80 | 75 | 72 | | 75 | | | | | 91 |
| MF♀ | | 84 | 73 | 79 | | 74 | | | | | 89 |
| M ♂♀ | | 74 | 64 | 61 | | 64 | | | | | 84 |
| F ♂♀ | | 90 | 82 | 90 | 60 | 86 | | 64 | | | 95 |

*) All relationscores are positive, and truncated after the second decimal. Decimal points and leading zeroes omitted.

rate "expressionless" on one of the extremes of the scale continuum "monotonous-melodious" this would indicate that "expressive-expressionless" and "monotonous-melodious" generally covariate. (In an earlier experiment it is established that judgment of one term of the contrastive scale terms is sufficient to determine the rating of the other one too. (Van Herpt, 1986:22)). If subjects divide randomly (e.g. if the score of the concept "deep" on the scale "soft-loud" is around the scale mean) it would appear that there is no relation.
So the raw scores already reflect the degree of interrelatedness of the fourteen scales and the measure of relation in this analysis is simply the mean score of the subjects. The mean score of each rating is rescaled on an interval from -1 to +1 and entered in a symmetrical 14x14 matrix. Factor analyses generally require product-moment correlations as 'measures of association', we considered our measure of interrelation also appropriate because the system of calculation is logically sound which is confirmed by plausible results of the analyses. This interrelation matrix is factorized. In fact several matrices are constructed; viz. female (♀) and male (♂) voice and pronunciation, and male (M) and female (F) raters are distinguished. By factoring these matrices separately, differences in underlying structure of qualifier terms for different samples (i.c. sex-groups) become discernible.

In this experiment the comprised version of the Voice and Pronunciation test is used and another sample of female and male subjects is drawn from our population of students. Fifteen male and fifteen female students gave their ratings on the 91 item list in relation to the male voice and pronunciation, and 2x15 others did so for the female voice and pronunciation.

## 5.0 SOME REMARKS ON THE FACTOR MODEL AND ANALYSIS

In this study the factor model is not primarely used to summarize data but rather to confirm hypotheses about the structuring of variables in terms of expected distributions of factor loadings over significant factors.
We would have preferred to apply common factor analysis (FcA) which assumes that the scale under consideration is influenced by various underlying determinants, some of which are shared by other scales of the test. The part that is shared with other scales is called common and the analysis is concerned with defining the pattern of common variation among the set of scales. The amount of common variance of a scale is represented by the communality ($h^2$). In factor analysis the main diagonal elements of the correlation matrix are replaced with estimates of communality which causes variation unique to a scale to be ignored. This uniqueness can be divided into two portions of variance - specificity, due to the particular selection of scales, and error variance, due to unreliability of the measurement.
Although this was the preferred model, we had to resort to principal component analysis (PCA) because (Nie et al. 1975:479) 'principal components may be extracted from certain sets of highly correlated variables which cannot be processed by other factoring methods if the correlation matrix cannot be inverted'. (The correlations - which will be examined later - are given in table 2.) Principal component analysis, the solution of factoring with unities in the diagonal of the correlation matrix, patterns <u>all</u> the variation in the set of scales, whether common or unique. That is not a realistic solution because it is not likely that there is no uniqueness in a limited set of scales like ours. And, even for scales that would describe voice and pronunciation very completely and with great precision, we have to hypothesize some error and specificity.
SPSS (Nie et al. 1975:480) allows a modification of the principal component solution 'by replacing the main diagonal of the correlation matrix with some estimates of communality'. In the present analyses we chose this solution that enabled us to take out the presumed unique variance of each scale. This technique will henceforth be referred to as principal factor analysis (PFA). Estimates of communalities are derived as follows. A principal component analysis is performed and the first few components that

have at least two (=common) significant loadings are retained for further rotation. In all analyses this happened to be components with eigenvalues greater than 1.0. The communalities that resulted from these principal component analyses are entered on the diagonal and from this reduced correlation matrix principal factors with eigenvalues greater than 1.0 are extracted. The initial factor solutions then are rotated orthogonally to the conceptually simpler Varimax solutions (which center on simplifying the columns of a factor matrix).
A second problem we encountered was the occurrence of one or more variables with loadings or communalities greater than 1.0. In literature this situation is referred to as a generalized Heywood case. Harman (1967:232) describes a way to solve the problem, but it is a complicated way. In our case the phenomenon appears to have to do too with the high intercorrelations seeing that no scales with uniqueness zero appeared when correlations higher than 0.80 were eliminated.
So we solved our problems rather drastically and eliminated from the analyses scales 13 and 14 that showed the most and the highest correlations with the other scales. Table 2 shows all interscale coefficients greater or equal than 0.60 for the complete sample and its four cross-sections.

## 6.0 RESULTS

### 6.1 Description of female versus male voice and pronunciation

As mentioned before a sample of thirty raters of both sexes gave their ratings bearing in mind the average female voice; another sample of the same size and composition did so for the average male voice.
In order to verify whether the same factor space is used in the description of female and male voice and pronunciaton both datasets are factorized. Using Kaiser's citerion for the number of factors to be extracted, both principal factor analyses yield four factors, which account with relation to the female and the male voice for respectively 70.1 and 73.6 per cent of explained variance.
Table 3 contains the principal factor analysis matrices of voice characteristics of the average female and the average male voice and pronunciation. In the matrices factor loadings are given who satisfy the one per cent level of significance according to the Burt-Banks formula (Burt, 1952). (Numbers between parentheses are insignificant but are given for comparison.)

Table 3 - PFA matrices of voice characteristics of the average male and the average female V&P.

| Scale | Average male V&P | | | | | Average female V&P | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | F1 | F2 | F3 | F4 | $h^2$ | F1 | F2 | F3 | F4 | $h^2$ |
| Sc01 | 57 | | | 50 | 61 | | 59 | (34) | | 53 |
| Sc02 | 84 | | | | 84 | | 86 | | | 87 |
| Sc03 | | | | 79 | 75 | | | 73 | | 68 |
| Sc04 | | | | 75 | 57 | | | 86 | | 75 |
| Sc05 | | 67 | | | 75 | 73 | | | | 81 |
| Sc06 | | 83 | | | 80 | 81 | | | | 74 |
| Sc07 | 68 | 48 | | | 74 | 69 | (45) | | | 73 |
| Sc08 | | 81 | | | 89 | 73 | | | | 74 |
| Sc09 | | | 82 | | 71 | | | | 81 | 71 |
| Sc10 | | | 97 | | 96 | | | | 89 | 82 |
| Sc11 | 77 | | | | 72 | | 67 | | | 52 |
| Sc12 | 51 | | | | 44 | | 54 | | | 46 |
| % expl.var. | 22.2 | 18.4 | 16.7 | 16.2 | 73.6 | 20.2 | 18.4 | 16.2 | 15.4 | 70.1 |
| Crit. sign. | 44 | 46 | 48 | 51 | | 44 | 46 | 49 | 51 | |

Leading zeroes and decimal points are omitted.
All numbers are truncated.

Comparison of the matrices shows:
1. The obvious similarity of the two factor matrices. The two loadings that are not significant in both the two matrices show a strong tendency in the expected direction.
2. The sum of the variances of each of the scales on the four extracted factors, as indicated by the communalities ($h^2$), too is very similar in both analyses.
3. Factor 1 and 2 changed place in the second analysis and so did factor 3 and 4, but since the differences in percentage of explained variance are small, also in this respect both solutions are similar.

We also verified the similarity on the male and female voice and pronunciation pattern by way of discriminant analyses. This technique takes a matrix of n cases (i.c. judgments) of different groups and m variables for input and calculates an optimal linear combination of discriminating variables for each of the groups. In our experiment the interrelation scores of the scales are the discriminating variables and male versus female voice and pronunciation are the distinguished groups. If - in this case - the analysis would be able to distinguish between the scale patterns of female and male voice and pronunciation, this would indicate that the semantic structure depends on sex of speaker. Furthermore the analysis would identify the 'best' discriminating scales.
Analyses have been performed for female versus male voice and pronunciation according to the ratings of men and separately according to those of women. The voice and pronunciation groups proved statistically indistinguishable: in neither of the analyses any variable reached the criterion (partial F-ratio > 1.0) of statistical significance of centroid separation.

Conclusion
Our samples with raters of both sexes used the same underlying theoretical framework when judging respectively the female and the male voice and pronunciation. This confirms our earlier finding that speakers of different sexes are judged in a common factor space. The discriminant analyses show that this statement also holds for rater samples of men or women only.

## 6.2 Identification of inconsistent scales

Comparison of factor analyses of the rater samples differing in gender did - in contrary to the results concerning the speakers in the preceding section - not show congruous results. This might indicate that scales have different meanings for different groups: an assertion which will be confirmed.

Table 4 - Correlations between scales with pronounced differences between ratings by male (upper triangle) and female (lower triangle) raters.

| Scale | Sc03 | Sc05 | Sc06 | Sc07 | Sc08 | Sc10 | Sc14 |
|---|---|---|---|---|---|---|---|
| Sc03 | -- | 17 | 33 | 14 | 01 | 00 | 61 |
| Sc05 | 41 | -- | 60 | 44 | 27 | -33 | 64 |
| Sc06 | 55 | 92 | -- | 05 | 32 | -36 | 21 |
| Sc07 | 36 | 78 | 54 | -- | 56 | 05 | 38 |
| Sc08 | 16 | 42 | 70 | 78 | -- | -37 | -23 |
| Sc10 | 11 | 05 | -07 | 42 | -08 | -- | 17 |
| Sc14 | 90 | 86 | 43 | 64 | 16 | 46 | -- |

Coefficients are multiplied by 100.

On various cross-sections of the object and subject sample we performed a series of numerical explorations in which successively one or more scales were eliminated from the analyses. Thus we established which scales had to be removed in order to decrease optimally the dissimilarities of factor patterns of different rater samples. Especially

Scale 06: "husky - not husky" seemed to be interpreted differently by female versus male raters. Differences in connotations show clearly in the correlations with other scales. Table 4 contains a partial matrix with correlations differing significantly between both groups. (In the upper triangle correlations based on scores of female and male voice and pronunciation rated by men, and in the lower triangle by female raters.)
Close reading of the data reveals that the inconsistencies in Scale 06 are caused by the male raters. Female judges do not differentiate the meaning of the scale "husky - not husky": independent of sex of speaker a consistent correlation with "dull", "weak" and "soft" appears. Male raters however, do not relate "husky" and "weak" when men are judged, although they associate "husky" with "dull" (and "deep" and "low") when judging the female voice and pronunciation.
A consequence of this behavior of Scale 06 (Sc06) is an undesirable dependency of the semantic factor structure on the composition of the samples, which would justify elimination of the scale in further analyses in order to diminish listener variance. Such a decision however suggests further eliminations - e.g. scales 07, 10 and 14 too seem to show the influence of raters' sex - and that would not leave much of our test. It is however, from the correlation coefficients, not clear to what degree these differences influence the dimensions of factor space itself. So it appeared that women are generally more associative in their ratings (show higher correlations between scales), especially when judging female speakers (see table 5), which, it is true, influences the allocation of the speakers but not the semantic reference frame.

Table 5 - Correlations between semantic twin scales for different cross-sections of the samples i.c. female (F) and male (M) raters female (♀) and male (♂) V&P.

| Scales | MF/♂ | MF/♀ | F/♂♀ | M/♂♀ | F/♂ | M/♂ | F/♀ | M/♀ |
|---|---|---|---|---|---|---|---|---|
| 01-02 | 77 | 79 | 90 | 65 | 87 | 64 | 93 | 67 |
| 03-04 | 60 | 68 | 79 | 46 | 78 | 33 | 81 | 57 |
| 05-06 | 75 | 81 | 92 | 60 | 87 | 55 | 98 | 64 |
| 07-08 | 74 | 65 | 78 | 56 | 78 | 64 | 79 | 50 |
| 09-10 | 83 | 73 | 82 | 73 | 84 | 78 | 79 | 69 |
| 11-12 | 74 | 55 | 58 | 73 | 64 | 81 | 50 | 67 |
| 13-14 | 92 | 89 | 95 | 85 | 96 | 86 | 95 | 83 |

Leading zeroes and decimal point are omitted.

So we decided for further analyses, and first tried to distinguish statistically between the two rater sex-groups on the basis of their own ratings of scale interrelations in connection with the average male and female voice and pronunciation. A discriminant analysis calculated a weighted linear combination of six scales which classified 85.7% of the raters correctly. Applying the same formula to another cross-section of the rater sample produced in over 70% of the cases correct identification of raters' sex. This and other analyses clearly indicate dependency of scale scores on raters' sex.
In order to get a clue concerning the most discriminating scales, we used each scale separately as a discriminating variable. Four scales - 03, 06, 07 and 14 - each classified the cases in over 50% with the correct sex groups. And when stepwise analyses are performed in which either Sc06 or Sc07 is the first variable to be considered for entry in the analysis, Sc03 and Sc14 do not qualify for analysis. When Sc06 and Sc07 are entered both in one analysis, they appear to be about equally important (standardized canonical discriminant function coefficients of resp. 0.60 and 0.63).
To decide whether it is sufficient to eliminate one of these two scales in order to obtain optimally identical factor patterns when using different rater groups, several principal component analyses are performed. The factor patterns of Sc01 to Sc14 for resp. female and male raters are given below. In table 6a Sc06 is left out of consideration, in table 6b Sc07 is.
Comparison of the matrices in table 6 shows that elimination of Sc07 increases the differences between the matrices of men versus women, whereas removal of Sc06

results in smaller dissimilarity of these matrices. This suggests that scores on Sc06 include enough rater variance to disrupt the common factor space of female and male raters. Such behavior would justify final removal of Sc06 from the test. But again the results are too ambiguous to settle the matter. After removal of as well Sc06 as Sc07 differences in factorial structure remain apparent. Men associate 'husky', 'weak' and 'soft' with a cultured articulation whereas women relate these qualifers to a dull voice. Besides, the Tempo factor has an evaluative connotation for women, which it does not have for men. So, we decided for one more attempt.

Table 6a - Principal components of V&P ratings by female resp. male judges after elimination of scale 06: husky - not husky.

| | Female raters | | | | Male raters | | | |
|---|---|---|---|---|---|---|---|---|
| Scale | F1 | F2 | F3 | F4 | F1 | F3 | F4 | F2 |
| Sc01 | 64 | 56 | | | 72 | (25) | | |
| Sc02 | 58 | 65 | | | 66 | (35) | | |
| Sc03 | 88 | | | | 72 | | | |
| Sc04 | 90 | | (-15) | | 60 | | (-49) | |
| Sc05 | 60 | | 70 | | 66 | | (28) | |
| Sc07 | | | 93 | | | | 74 | |
| Sc08 | | | 86 | | | | 85 | |
| Sc09 | | | | 99 | | | | 91 |
| Sc10 | | | | 89 | | | | 85 |
| Sc11 | | 74 | | | | 85 | | |
| Sc12 | | 84 | | | | 86 | | |
| Sc13 | 91 | | | (48) | 91 | | | (35) |
| Sc14 | 79 | 48 | | | 91 | (17) | | |

Table 6b - PCA matrices of V&P ratings by female resp. male judges after elimination of scale 07: weak - powerful.

| | Female raters | | | | Male raters | | | |
|---|---|---|---|---|---|---|---|---|
| Scale | F1 | F2 | F3 | F4 | F1 | F3 | F4 | F2 |
| Sc01 | 47 | 72 | | | 71 | (35) | | |
| Sc02 | (37) | 82 | | | 59 | 58 | | |
| Sc03 | 81 | | | | 73 | | | |
| Sc04 | 95 | | (-12) | | 44 | | -76 | |
| Sc05 | 50 | | 76 | | 70 | | (13) | |
| Sc06 | (26) | | 95 | | 60 | | (49) | |
| Sc08 | | | 85 | | | | 78 | |
| Sc09 | | | | 99 | | | | 89 |
| Sc10 | | | | 91 | | | | 85 |
| Sc11 | | 80 | | | | 84 | | |
| Sc12 | | 72 | | | | 75 | | |
| Sc13 | 78 | | | 54 | | | | (32) |
| Sc14 | 60 | 69 | | | | (29) | | |

All numbers are truncated; decimal points and leading zeroes omitted. Loadings between parentheses are given for comparison but are insignificant at 1% level of significance according to the Burt Banks formula.

Our next approach to get grasp on the relation between the scales and the sex differences of the raters may be debatable but yields interesting and plausible supplementary results. The interrelation matrix of men judging male and female voice and pronunciation is added on the one of women containing the interrelations between the same fourteen scales. These data are factorized (see right half of table 7). Next a fifteenth variable (Sc15) is added which marks the sex of the rater. Sc15: "male-female" is assigned a value of about zero on each row of the 'female' matrix and of about one in the 'male' matrix. The correlations of Sc15 with the other 14 scales are rather low, only the correlations with scales 03, 06, 07 and 14 - which are around 0.30 - reach a 5% level of significance. These data are factorized too (see table 7).

Two results are striking.
1. The analysis in table 7b shows a factor in which the sex indicator goes with Sc07 and Sc08, our dimension of Subjective Strength, and shows a minimal relation of the sex variable with the Clarity dimension (Sc05+Sc06) that appears to go with Sc08:soft-loud.
2. Comparison of the two factor patterns in table 7 shows a very strong correspondence after Varimax rotation.

So, at first sight the relation of Sc07: "weak-powerful" with Sc15: "male-female", as indicated by high loadings in factor 4 only, makes Sc07 suspect again. But, as can be seen from the distribution of explained variance over the components and from the concordance of the matrices in table 7 the influence of Sc15 on the original pattern of loadings is very small. The pattern seems to be too robust to be influenced by the sex variable, which makes, in combination with the small change in its communality, a strong effect of Sc07 on the factorial structure unlikely too.

Table 7 - Principal components of combined ratings by men and women of female and male V&P, with and without a rater sex variable (Sc15).

| | With sex indicator | | | | | Without sex indicator | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Scale | F1 | F2 | F3 | F4 | $h^2$ | F1 | F2 | F4 | F3 | $h^2$ |
| Sc01 | 76 | | | | 86 | 82 | | | | 87 |
| Sc02 | 64 | 55 | | | 88 | 72 | (40) | | | 90 |
| Sc03 | 85 | | | | 83 | 82 | | | | 84 |
| Sc04 | 79 | | | | 75 | 73 | | | | 79 |
| Sc05 | | | 65 | | 86 | | | 62 | | 87 |
| Sc06 | | | 87 | | 86 | | | 87 | | 86 |
| Sc07 | | | | 74 | 76 | | | | 85 | 85 |
| Sc08 | | | 55 | 56 | 93 | | | 50 | 68 | 93 |
| Sc09 | | -77 | | | 85 | | -80 | | | 84 |
| Sc10 | | -82 | | | 91 | | -89 | | | 92 |
| Sc11 | | 89 | | | 87 | | 85 | | | 88 |
| Sc12 | | 88 | | | 84 | | 90 | | | 88 |
| Sc13 | 91 | | | | 88 | 90 | | | | 89 |
| Sc14 | 92 | | | | 92 | 94 | | | | 91 |
| Sc15 | (-18) | (22) | (-09) | -68 | 56 | | | | | |
| % Var. | 31 | 26.2 | 13.4 | 12.9 | 84.2 | 33.1 | 26.2 | 13.2 | 14.8 | 87.8 |

Loadings <.50 for comparison between parentheses.
Leading zeroes and decimal points omitted; numbers truncated.

## 7.0 CONCLUDING REMARKS

The question whether raters use the same underlying theoretical structure when judging respectively the female and the male voice and pronunciation is answered rather unequivocal. Several criteria indicate that this and former analyses are sufficient alike to accept both structures as being equivalent. The same dimensions as before are identified and the analyses show the same factor structure when speakers of different sex are judged, i.e. the qualifier structure is relatively invariant under varying concepts, in this case male and female voice and pronunciation.
But although there is no change in factor structure attributable to concepts there was some change due to subjects. This answers the main question in this study, whether female and male raters use a common semantic framework in the description of voice

and pronunciation. It appears that not all of our qualifiers of voice and pronunciation have the same meaning for men and women. Granted an overall agreement between female and male raters there are a few discernible differences, especially Dimension III: Voice Quality remains recalcitrant. Female speakers keep this dimension intact, but men do not include Sc05: 'dull-clear' in the cluster with 'husky', 'weak' and 'soft' and instead class Sc04: 'broad-cultured' with them. This answers our question whether sex of rater is a variable in the tendency of the Voice Quality dimension to break down. As for the other dimensions only the relation between 'Voice Appreciation' and 'Tempo' is a little conspicuous. Women are on the whole more evaluative in their ratings but only in the Tempo factor the difference with male ratings is enough to show significantly in the pattern of loadings, however it is not large enough to change the factorial structure.
Our last question concerned the identification of scales that influence negatively the consistency of the factor structure under varying subject samples. Considering our above remark concerning the sex effects in Dimension III it was to be expected that removal of Sc05 would diminish the difference between the overall factor patterns of female versus male raters. However the opposite proved to be the case: the differences enlarged. Eliminating by turns the other three scales from the analyses showed that removal of Sc06: 'husky - not husky' lowers most differences between the matrices but does not eliminate all of them, particularly the above mentioned differences in Dimension III remain immovable.
Potentially an improvement of the scale battery can be achieved by removal of Sc06 and by breaking down the Voice Quality factor which can possibly be brought about by adding a scale in relation to sharpness (Bismarck, 1974; Bloothooft, 1985).

If we now try to summarize the differences in scoring behavior of men and women two main differences in approach can be distinguished.
The first one shows a suggestive pattern that indicates stereotypical behavior. For men Strength is important: a weak voice is not melodious, although speaking too powerful is broad. Both sexes associate Strength with masculinity. Women in their ratings appear to consider Tempo more important than Strength, which shows in considerably higher evaluative loadings on that factor.
Secondly, women differ from men in that they consistently score more extreme. They would appear to have more and stronger connotations of attributes, to weigh the appreciative connotations of qualifiers they consider relevant, heavier then men do. Men seem less willing to indicate dependence among qualities and are less evaluative. The latter phenomenon is mentioned also by Osgood and Tannenbaum (1955) who observe that more extreme judgments are characteristic of more emotionally oriented individuals such as women are supposed to be. We wonder if either associative or evaluative are more appropriate qualifiers of this behavior.

Finally, three remarks in relation to the analyses.
The first concerns an alternative of the Varimax rotation we used. In this type of problems in which we want to find out whether the factorial composition of matrices based on scores of different samples are identical, another type of rotation should have been used. Especially application of Procrustean transformations that rotate to factors with properties that resemble as much as possible the properties of a number of previously defined factors, would have been an improvement.
Second, the reliability coefficients of the scales would have been better estimates of the communalities in our principal factor analyses. The total variance of a variable is made up of common factor variance (communality) and uniqueness. The latter can be broken down in variance particular to the variable under consideration (specificity) and error variance or unreliability. In other words reliability is communality plus specificity, or reliability is the total variance minus error. Consequently, the communality of a variable is less than or equal to the reliability of that variable and equals the reliability only when the specificity vanishes.
When using the reliability coefficient as (upper limit) estimate of the communality the difference between this coefficient and the communality resulting from the analyses is a measure of the specificity. The advantage of using this procedure is that estimates are

used from which the error is rightly removed and that the amount of specific variance can be calculated. When the specificity is high, this can be a reason to add scales to the battery in order to 'catch' that specificity in another factor. A disadvantage is that the reliability has to be calculated separately. This however, is anyhow useful in order that unreliable scales can be removed prior to analyses.
Our third methodological remark concerns an evaluation of the importance of the findings in the present study. In semantic differential studies variance comes from three sources: the subjects, the objects and the scales. Generally the variance due to subjects is relatively small. The undesirable systematic subject variance - the existence of which is shown in the present study - forms part of that variance. So, to be able to evaluate its importance analyses of variance in order to assess the (relative) contribution of each of the three modes to the total amount of variance, are necessary.

In the present study sex of the perceiver was a considered variable and degree of training was more or less controlled. In a follow up study expert raters and naive listeners will be compared.

## REFERENCES


Bismarck, G. von (1974). Sharpness as an attribute of the timbre of steady sounds. Acustica 30, 159-172.

Bloothooft, G. (1985). Spectrum and timbre of sung vowels. Dissertation, Vrije Universiteit van Amsterdam.

Burt, C. (1952). Tests of significance in factor analysis. Brit. J. Psychol., 5, 109-133.

Fagel, W.P.F., Herpt, L.W.A. van & Boves, L. (1983). Analysis of the perceptual qualities of Dutch speakers' voice and pronunciation. Speech Communication 2, 315-326.

Harman, H.H. (1967). Modern Factor Analysis. The University of Chicago Press, Chicago.

Herpt, L.W.A. van (1986). Influence of rater's sex on voice and pronunciation assessment. Proc. Inst. Phonetic Sciences, Univ. of Amsterdam, 10, 19-39.

Nie, N.H., Hull, C.H., Jenkins, C.B., Steinbrenner, K. & Bent, D.H. (1975). SPSS - Statistical package for the social sciences. McGraw-Hill, New York.

Osgood, C.E. & Suci, G.J. (1955). Factor Analysis of Meaning. J. of Exp. Psychology, 50, 325-338.

Osgood, C.E. & Tannenbaum, P.H. (1955). The principle of congruity in the prediction of attitude change. Psychological Review, 62, 42-55.