Karolinska Institutet
Technical Audiology
KTH, 100 44 Stockholm 70

Report TA No. 75
June 1974

ADJECTIVE RATINGS AND DIMENSION ANALYSES OF PERCEIVED SOUND QUALITY OF HEARING AIDS. I

Alf Gabrielsson and Håkan Sjögren

ABSTRACT

In order to study the perceived sound quality of certain hearing aid systems twenty subjects rated the sound reproduction of the systems on a number of adjective scales. The systems included one broad band system, one headworn aid, one bodyworn aid, and two systems representing pronounced lowpass and highpass filtering, respectively. The ratings were subjected to various forms of factor analysis. The resulting factors were interpreted as "sharpness/hardness - softness", "clearness/distinctness", "loudness", "brightness - darkness", and "room impression". The factors were tentatively related to the physical characteristics of the systems.

INTRODUCTION

Investigations on the perceived sound quality of sound-reproducing systems, mainly loudspeakers, were described in a series of papers (Gabrielsson, Rosenberg & Sjögren, 1971, 1972, 1973, 1974; Gabrielsson 1974). A common denominator of these studies was the application of multivariate methods like multidimensional scaling and factor analysis (FA) in an attempt to reveal some perceptual dimensions underlying the judgments of the listening subjects. The resulting dimensions could be interpreted in fairly unambiguous ways (for instance, perceived "distinctness/clearness", "loudness", "brightness-darkness", "hardness-softness", "volume") and were tentatively related to various physical characteristics of the sound-reproducing systems in question.

In this and some forthcoming studies similar methods are used to study the dimensionality of perceived sound quality as regards hearing aids. Hearing aids are usually designed to give good intelligibility of speech. However, there are other considerations, too. Hard-of-hearing people using hearing aids often complain about the un-naturalness of the reproduced sounds. Patients often describe sounds from their aids as "hard", "metallic", "strange", "painful", etc. It is obvious that, besides the intelligibility, the perceived sound quality of the hearing aid is of primary importance and may indeed affect the willingness to use, and the way of using, a certain hearing aid. It is not sure, of course, that a hearing aid designed for good intelligibility is also a good one as regards the perceived sound quality.

Depending on the type and extent of hearing loss the judgments about sound quality may be highly specific for each individual. This in turn may make it more difficult to understand the relations between the physical characteristics of the hearing aids and perceived sound quality. To get a reference basis in these questions and to try out the usefulness of multivariate methods in this context we therefore decided to use normal hearing subjects in the introducing experiments.

The method used in the present report is a variant of the "semantic differential" (Osgood et al., 1957; Snider & Osgood, 1969), essentially meaning that the stimuli (here the stimuli are the reproductions given by certain hearing aid systems) are rated by the listening subjects on a number of "adjective scales". These ratings are then subjected to factor analysis (FA) in order to find out some underlying (perceptual) dimensions. The same technique was successfully applied to judgments of loudspeaker reproductions (Gabrielsson et al., 1973).

METHOD

Adjectives descriptive of perceived sound quality of hearing aids were collected by means of questionnaires to people suffering from hearing loss. The adjectives were then used for ratings of the sound quality of five hearing aid systems. The ratings were subjected to factor analysis, factors (dimensions) were interpreted and tentatively related to the physical characteristics of the systems.

Rating scales (adjectives)

Some 200 adjectives compiled from various sources and originally used for judgments on loudspeaker reproductions (Gabrielsson et al., 1973) were taken as a starting point. In a first step these adjectives were rated for suitability or not to characterize hearing aid reproductions by some 30 persons, all of them professionally working in audiological research and audiological treatment. They were also asked to supply the list with other adjectives. As a result of this about 80 adjectives were left for further consideration. After eliminating some words, which were nearly synonymous or represented opposites to other adjectives, a list containing the remaining ones was distributed by mail to about 250 persons suffering from hearing loss. Briefly described, they were asked to judge the adjectives as to "how good you think these words are to describe how a hearing aid may sound". A four point scale was used, 4 = Very good word, 3 = Good word, 2 = Possibly good word, and 1 = Bad word. They were also asked to propose other adjectives of their own choice.

The questionnaire was answered by 172 persons, 83 males and 89 females, 16-70 years old and with highly varying types and degrees of hearing loss. The loss was said to be congenital for 31 persons, acquired for 68 persons while no information was given concerning the remaining 73 persons. Unfortunately, but not unexpectedly, many subjects misunderstood the instruction and believed that they should rate how well they could hear the listed adjective words when they were spoken by another person or the like. Eliminating answers of that type the mean rating and standard deviation for each adjective (on the 1-4 scale above) was computed over the remaining 105 persons. The result was that practically all adjectives got mean ratings between 2.2 and 3.4 (the standard deviation was about 1.0 for each of them). On the whole, then, the selection of adjectives made by the audiologists was confirmed.

Eliminating some of the adjectives with the lowest mean ratings and considering some other proposed adjectives resulted in a final list of 62 adjectives, see Table I (the adjectives are there also translated into English). A certain influence of the authors' subjective opinion was inevitable in the selection. Selection problems like this have to be considered anew for each separate investigation.

Stimuli and listening conditions

Because of the exploratory nature of this experiment, it was decided to select a few systems with highly different characteristics.

System 1 (S_1): Broad-band system without deliberate distortion. This was realized by a tape copy of the master stimulus tape (100-15000 Hz ± 2 dB, electrically measured).

System 2 (S_2): Head-worn hearing aid.

System 3 (S_3): Low-pass filtered system, 6 dB/octave in the frequency range 100-10000 Hz.

System 4 (S_4): Body-worn hearing aid.

System 5 (S_5): High-pass filtered system, 12 dB/octave in the frequency range 100-5000 Hz.

System 1 may serve as a reference. The two hearing aids were chosen to represent a broad-band (S_4) and a narrow-band (S_2) type of aid. The high-pass filtered system will give highly intelligible speech for many hard-of-hearing persons, but may also be supposed to introduce a "hard" and "sharp" sound. On the other hand the low-pass filtered system may give less intelligible speech but a "soft" and "calm" sound.

During the experiment stimuli were presented over headphones TDH39, cushion MX41AR. Listening levels were set by the experimenters to correspond to a natural level for the respective program sections (see below). Stimuli presented over the two hearing aids were recorded in an anechoic chamber using the 2 cc coupler (IEC).

Six programs were chosen to represent speech, music and "environmental sound".

Program 1 (P_1): Recording from a day nursery, pre-school children engaged in noisy play. Listening level 72 dB SPL.

Program 2 (P_2): Oscar Peterson's jazz trio, piano, bass and drums. Recorded in a grammophone studio. Listening level 73 dB SPL.

Program 3 (P_3): Traffic noise from a city street. Listening level 73 dB SPL.

Program 4 (P_4): Speech. Text read by the experimenter (B.L.) recorded in an anechoic chamber. Listening level 73 dB SPL.

Program 5 (P_5): Organ music. Glazunov: "Prelude and Fugue in D minor", the ending chorale (in fortissimo). Recorded in an empty church. Grammophone record: Proprius PROP 7707 (distribution EMI). Listening level 68 dB SPL.

Program 6 (P_6): Recording from a coffee break at the laboratory. Some fifteen persons chatting. Listening level 73 dB SPL.

Each program section lasted for about 35 seconds. Tape copies of the program material are available on request.

During the experiment the subjects were sitting three at a time in an audiometric booth.

Subjects

There were twenty subjects, 15 males and 5 females, 18-28 years old, all of them with normal hearing (less than 20 dB hearing loss 250-4000 Hz). Only some few of them performed music (as amateurs). Except for three of them they had no special experience of or interest in high-fidelity sound reproduction. All subjects were paid.

Procedure

In total there were 30 different combinations of program sections and hearing aid systems (6 program sections x 5 systems). In the following these are referred to as P x S combinations (P for program, S for system). Since there were 62 adjectives each subject made a total of 30 x 62 = 1860 ratings. This required four experimental sessions of 1.5 - 2 hours each. Each such session comprised all 30 P x S combinations in a randomized order (the randomization was different for different sessions and for the different groups of three subjects sitting together in the audiometric booth).

Each subject had a number of adjective lists, one for each of the 30 P x S combinations. The order of the 62 adjectives on each list was randomized, different for different lists and different individuals. The randomization and printing out of the lists were made by computer. The following instruction was given:

"In this experiment we want to investigate how you perceive different sound reproductions which may occur in connection with hearing aids. You will listen to different sections with music, speech, and 'environmental sounds' which have been recorded on tape from various hearing aids. Each presentation lasts for about 35 seconds and for every such case you shall describe how you perceive the sound reproduction of the section in question by writing a number from 0 to 9 for each of the adjectives on the respective list.

0 means that the reproduction has nothing of the quality denoted by the adjective. 9 means, on the contrary, that the reproduction has a 'maximum' of that quality. For levels between these extremes you use values in between: the more of the quality, the higher number (up to 9); the less of the quality, the lower number (down to 0). Only one number shall be written and decimals are not allowed. Try to make use of the whole scale 0-9. The reproductions are so varying that there are opportunities for using the whole scale.

It is very important to observe that the judgments shall refer to the sound reproduction. We are not interested in your opinion about 'the sounds as such' (for instance, what you think about the music, the voices, etc., as such). Try to control yourself during the experiment that you concentrate on the sound reproduction!"

After the instruction in the first session there were nine preliminary trials, the respective stimuli more or less randomly chosen among the 30 P x S combinations (however, each program section should appear at least once). In the following three sessions the instruction was briefly repeated and two preliminary trials were made. There were two breaks in each session. At the end of the fourth session the subjects answered some questions about judgment principles, if there were some unclear adjectives, etc.

Data treatment

(The reader who is not interested in all details of the data treatment may read this section for orientation only.)

There are four variables to consider in the data analysis: the program sections (P), the hearing aid systems (S), the adjectives (A), and the individuals. The individuals were regarded as a replication factor and the first step was therefore to compute the arithmetic means of the twenty individuals' ratings at each of the 1860 P x S x A combinations. These means form the basic data for the different factor analyses described below and can be obtained on request. (The standard deviation of the individuals' ratings within each P x S x A combination was in general between 1.4 and 2.4 units.)

The traditional application of factor analysis (FA) as, for instance, in research on intelligence tests and intelligence factors involves two variables: tests and individuals. Usually the correlations between the tests over the individuals form the starting point for FA (Harman, 1967; Sjöberg, 1972; Gorsuch, 1974; these works also serve as general references for all details about factor analytical procedures discussed in the following). In the present case involving three variables (P, S, and A, disregarding the individuals by taking means over individuals as stated above) the situation is more complex, and there is presently no generally accepted strategy for applying FA to such situations. Therefore several alternative approaches have to be tried and compared between themselves. For the present data the following alternatives were used:

a) FA on (product moment) correlations between adjectives over the 30 P x S combinations. This means in essence that the three variables (P,S,A) are reduced to two by combining P and S to a single variable with 30 (=6P x 5S) levels. This is reminiscent of the traditional application of FA in semantic differential research, where FA is usually applied to correlations between adjective scales over concepts (Osgood et al., 1957).

b) FA on correlations between the 30 P x S combinations over the adjectives, that is, the "reverse" of the preceding alternative.

c) FA on correlations between adjectives over systems within each of the six program sections (that is, one such analysis for each of the program sections).

d) FA on correlations between adjectives over systems in average over the program sections.

e) FA on correlations between systems over adjectives in average over the program sections (the "reverse" of the preceding alternative).

(The analysis used in all these cases was in fact component analysis. However, the term factor analysis is retained for convenience.)

The resulting solutions were rotated to "simple structure" according to the varimax principle (for orthogonal rotation) and in most cases also according to the "biquartimin" and the "direct oblimin" principles for oblique rotation. All FA:s and rotations were made by means of the FACTO program in the IBM Scientific Subroutine Package or the FACTOR ANALYSIS program in the X-series Supplement to Biomedical Computer Programs (BMDX 72). Factor scores were computed by means of the FASCORE program by Bäckström (1967) or within the BMDX 72 program.

f) Tucker (1966) recently proposed a model for a three-mode factor analysis (see also Levin, 1965), that is, for data classified in three different ways. Since this is the situation met here (three variables: P, S, and A) three-mode FA was applied to the present data by means of the TREMOD program by Gruvaeus et al. (1971), followed by varimax rotations according to a program by Muthén (1973). This procedure results in a factor loading matrix for each of the three modes (that is, programs, systems, and adjectives) and a "core matrix" representing the interrelations between the factors obtained for the different modes.

g) Different variants of FA were also applied to the data after these were "corrected" in a certain way due to reasons described in the following. The purpose with the individuals' ratings on adjective scales is to get descriptions of the perceived sound reproduction of the different systems. There is an apparent risk, however, that the ratings not only refer to the sound reproduction as such but also to the respective program section. For instance, there is no coincidence that the highest ratings of adjectives as "shrill" and "screaming" occur at program section 1, in which there are indeed many "shrill" and "screaming" voices of children. By the same line of reasoning it is no surprise that the highest ratings for adjectives as "hissing" and "whistling" occur for the traffic noise in program section 3. Although the subjects got the instruction to concentrate on the sound reproduction it is admittedly very difficult to do so (this was also pointed out by some subjects). A more realistic interpretation of data is probably that they somehow reflect the "joint" sound perception resulting both from a certain program section and a certain hearing aid system.

To "compensate" for the influence of the respective program sections and to get the effects of the different systems on perceived sound quality as "pure" as possible one may think of many different analysis methods. An apparent way is to make FA only within each program section (alternative c above) and then compare the results at the different program sections. However, this procedure makes no use of tendencies in the data which are common for all program sections. Furthermore the solutions at different program sections may not be directly comparable because of differences in rotation. Another way would be to make FA only on data taken in average over the program sections (alternatives d and e above). In view of the frequently occurring interactions between program sections and reproductions (Gabrielsson et al. 1971, 1972, 1973) this is not quite satisfactory.

Still another way which was tried here is to "correct" the means of the 1860 P x S x A combinations by subtracting from each of them the mean of the corresponding P x A combination. Consider for instance the adjective "shrill". As noted above the highest ratings of "shrill" appeared at program section 1 (mean value over the five systems = 5.9), the lowest ratings occurring for program section 4 (the male voice; mean value over the five systems = 3.2; the mean values for the other program sections were between these extremes). That would probably mean, for instance, that a system with a rating of 6.0 at program section 1 does not contribute so much to the perceived "shrillness" since the mean in "shrillness" at this program section (over all systems) was 5.9. However, if a system got a rating of 6.0 at program section 4 (where the mean over systems was 3.2) it would probably mean that this system made a substantial contribution to the perceived "shrillness". To study the influence of the systems on the perceived sound quality it may therefore seem reasonable to compensate for the influence of the program sections by subtracting from each separate P x S x A mean the mean of the corresponding P x A combination (for the examples above the new values would be 6.0 - 5.9 = 0.1 and 6.0 - 3.2 = 2.8, respectively; of course the new data may obtain negative values). The 1860 means were corrected according to this principle and FA was applied on these according to alternatives a and b above and also in connection with the three-mode FA.

It is not possible, nor necessary, to present the results of all these different analysis methods. Although there are certain differences in the interpretation of the resulting factors depending on whether one applies FA to the correlations between adjectives or to the correlations between P x S combinations (alternatives a and b, respectively) and/or depending on the type of rotation (orthogonal or oblique) it is quite clear that three or four factors return in practically all analyses with rather small differences in meaning. Some of the most representative and typical results are presented below.

RESULTS

Factor analyses on correlations between adjectives

The results given here refer to the factor analyses described in alternatives a, c, d, and g under Data treatment. However, the results from the analyses according to alternatives c and d (FA within each program section and FA in average over program sections, respectively) showed a good agreement with the results from the more comprehensive analyses according to alternatives a and g and are therefore not given separately.

In general two or three factors accounted for about 90 % of the total variance. The first factor was the same in all analyses, while there were certain differences between the results from the orthogonal rotation (varimax) and those from oblique rotations (biquartimin and direct oblimin, which were very similar between themselves) as regards the other factors.

The varimax rotated solution from FA on correlations between the adjectives over the 30 P x S combinations (alternative a) is given in Table I. Two factors accounted for 88 % of the total variance. The corresponding factor loadings from an analysis of the same type but based on "corrected" data (alternative g) were very similar to those in Table I and need not be given separately. The rank order in factor scores for the P x S combinations in the two factors is given in Table II for original (uncorrected) data as well as for corrected data (in the factor scores there were differences between these two analyses, see below).

Tables I and II about here

The first factor (F I) may be interpreted as a "hardness-softness" or perhaps "sharpness-softness" dimension. The adjectives with the highest positive loadings (.98-.99) in F I are "sharp", "hard", "shrill", "clashing", "pointed", "harsh", "screaming", "cracked", "pricking" and the like. Three adjectives expressing negative evaluations, "irritating", "painful", and "tiring", also appear here. On the other side of this dimension there are adjectives like "soft" (-.99), "calm", "mild", "soothing", together with adjectives expressing positive evaluations: "pleasant" and "natural".

Looking at the rank order of factor scores for the P x S combinations in this factor (Table II) it seems clear that this dimension is related to the frequency curve. The highpass filtered system (S_5) is, as expected, outstanding on the "sharp" side and the lowpass filtered system (S_3) on the "soft" side. In between there are (from "sharp" to "soft") the head-worn system (S_2), the body-worn system (S_4), and the broad-band system (S_1). This seems to be in accordance with the frequency response of the systems: the head-worn aid being narrow and with resonance peaks, the body-worn aid with a flatter response, and the broad-band system still flatter. The rank order of the systems is somewhat more obvious for the solution on corrected data than for that on original data.

The last-mentioned fact also recurs generally in the following analyses. That is, the solutions on corrected data do not alter the interpretation of the dimensions but as a rule give a more consistent rank order of the systems in the respective factors as it was expected (the data were corrected in an attempt to study the effects of the systems independently of the program sections, see Data treatment).

In earlier reports (Gabrielsson et al. 1971, 1972, 1973) a "brightness-darkness" dimension was found and related to the frequency curve in a way similar to that described above for "sharpness-softness". It is interesting to see that the factor loadings in F I for "bright" and "dark" are .91 and -.86, respectively. Although these loadings are high they fall short of the loadings for "sharp" and "soft", respectively (and for many other adjectives, too). It may be that with such extreme highpass and lowpass filtering as here (see Stimuli) the character of "bright" or "dark" is not quite as adequate as descriptions in terms of "sharp" or "soft". It may be noted, too, that terms like "sharp" and "soft" are very often used by hard-of-hearing people when characterizing the sound quality of their hearing aids.

The <u>second factor</u> (F II) may be interpreted as "clearness/distinctness". There are high (negative) loadings for "brilliant", "clear", "ringing", "open", "distinct", and "clean". On the opposite side there are high loadings for "restrained", "muddy/confused", "monotonous", "blurred", "shut up", "soft/gentle", "faint", and "subdued". It is noticed that adjectives denoting some kind of "room characteristics" appear in this factor with somewhat lower loadings: "open", "airy", "giving impression of room", "wide", and "near" on the "good" side of the dimension, while adjectives as "shut up", "distant", and "narrow" appear on the other side. (In certain analyses these adjectives occupy a factor of their own, see below.) "Pleasant" and "natural" occur on the "clearness" side, the latter one with a fairly high loading (-.57).

The rank order in the factor scores (Table II) for the systems in this "clearness" dimension is in accordance with what could be expected: the broad-band system (S_1) quite outstanding, then (roughly) the highpass filtered system (S_5), the body-worn system (S_4), and, on the "worse" side, the head-worn system (S_2), and the lowpass filtered system (S_3). As earlier the rank order of the systems is more evident from the analysis on the corrected data. Even there, however, there are indications of program x system interactions. For instance, the lowpass filtered system (S_3) is relatively better when reproducing the music sections (P_2 and P_5) than when reproducing the other program sections; the highpass filtered system (S_5) is relatively worse when reproducing the male voice (P_4) and so on.

On the whole "clearness" seems to be related to frequency range and the amount of distortion. The greater the frequency range, the better "clearness" as examplified by the broad-band system or by the body-worn system compared to the head-worn system. The less distortion, the "clearer" reproduction as illustrated by the broad-band system compared to all others or the body-worn system compared to the head-worn. Whether highpass filtering is better or worse for "clearness" than lowpass filtering seems to vary with the program sections (which seems reasonable). It should be emphasized (here as elsewhere in this report), that the detailed psychophysical relations may vary considerably depending on the type of program section and have to be established by follow-up experiments.

There was also a suggestion of a <u>third factor</u> in which only a few adjectives had loadings of some importance: on one side "soft/gentle" (.48), on the other side "noisy" (-.71), "coarse" (-.68), "rumbling" (-.55), "blurred" (-.39), and "loud" (-.34). An outstanding factor score on the former side occurred for the highpass filtered reproduction of the male voice (P_4), which indeed sounds very "soft/gentle" due to the extreme filtering. On the latter side the highest factor scores appeared for some reproductions by the head-worn system.

The <u>oblique rotations</u> according to the bi-quartimin and the direct oblimin methods gave very similar results between themselves but partly differed from the varimax solution with regard to F II and F III. The direct oblimin solution is seen in Table I.

Three factors accounted for 91 % of the total variance.

F I in this solution is apparently the same as in the varimax solution, that is, "sharpness-softness". The rank order of the P x S combinations in their factor scores was also very similar to that obtained for the varimax solution.

F II may be interpreted as a "clearness/distinctness" dimension as in the varimax solution. "Clear" and "distinct" have the highest factor loadings on one side (-.61), while "noisy", "coarse", "rumbling", "blurred", and "muddy" dominate on the other side (loadings from .97 to .67). The latter adjectives also appeared on the same side in F II at the varimax solution but there together with adjectives as "restrained", "soft/gentle", "faint" and "subdued", that is, together with adjectives mainly denoting a low loudness. In the present solution, then, "clear" and "distinct" are set in opposition to something "noisy", "rumbling" and "blurred" (as was also the case for the varimax solution). The loudness denoting adjectives do not appear in F II here (unlike the situation at varimax solution but are instead a main part of F III (se below). Likewise adjectives denoting various "room characteristics" do not appear in F II here (as they did in the varimax solution) but are also included in F III below.

F III may mainly be interpreted as a "loudness" dimension, however, with a touch of perceived "room characteristics" of the reproductions. The adjective "soft/gentle" is outstanding on one side (.96), followed by "restrained", "faint", "monotonous", "subdued", and "shut up". On the other side the highest loadings are held by "loud", "ringing", "open", "wide, "impression of room", "brilliant", and "airy". The "loud - soft/gentle" opposition is thus combined with an impression of an "open" or "airy" reproduction in contrast to a reproduction which is "shut up" or "narrow".

In the authors' opinion the direct oblimin solution seems somewhat more consistent than the varimax solution. The oblique relation between factors in the direct oblimin solution is mainly limited to F II and F III, which correlate positively +.42 (correlation between F I and F II +.01, between F I and F III

+.14). In this context it is interesting to compare the rank orders of the P x S combinations in their factor scores in F II and F III (Table III). On the "noisy", "coarse", and "blurred" side of F II there are the lowpass filtered system (S_3) and the head-worn system (S_2), on the "clear" and "distinct" side there are the broad-band system (S_1) and the highpass filtered system (S_5); the body-worn system (S_4) lies in the middle. The situation in F III is similar in certain respects: on the "soft/gentle", "restrained" and "shut up" side there is mainly the lowpass filtered system (S_3) and on the "loud" and "open" side the broad-band system (S_1) is outstanding. The head-worn system (S_2) and the body-worn system (S_4) mainly appear in the middle, while the highpass filtered system (S_5) is scattered over most of the range including the reproduction of the male voice (P_4S_5) as the most extreme P x S combination on the "soft/gentle" and "shut up" side (the same thing occurred in F III at the varimax solution). A detailed study of Table III suggests many P x S interactions, that is, the position of a system within the rank orders may depend on which program section is reproduced; compare, for instance, the position of the lowpass filtered system (S_3) in F II when reproducing the music sections (P_2S_3, P_5S_3) and when reproducing the other program sections.

Table III about here

Factor analyses on correlations between P x S combinations

The results given here refer to the factor analyses described in alternatives b, e, and g under Data treatment. Since the results from alternative e agreed with the results from the more comprehensive analyses in b and g they are not given separately here.

In general two or three factors accounted for 85-90 % of the total variance. There were practically no differences between the results from the three different rotation methods (varimax, bi-quartimin, direct oblimin). The analyses on the corrected data resulted in fewer interpretable factors than the analyses on the original data. On the other hand the rank order of the systems in the different factors was more consistent in the solution based on corrected data.

The results may be described rather briefly since they in most respects agree with those obtained from the factor analyses on correlations between adjectives. F I is, as in the earlier analyses, a "sharpness-softness" dimension. The rank order of the P x S combinations in their factor loadings in F I is similar to that given for F I in Table II, that is, with the highpass filtered system (S_5) and the head-worn system (S_2) belonging to the "sharp" side, the lowpass filtered system (S_3) and the broad-band system (S_1) relatively more to the "soft" side, and the body-worn system (S_4) in between. The adjectives "tiring", "sharp", "shrill", "obtrusive", "hard", and "clashing" had the highest factor scores on one side, while the highest factor scores on the opposite side belonged to "soothing", "soft", "pleasant", "mild", "natural", "calm", and "dark".

In F II there were substantial factor loadings for P x S combinations only on one side of the continuum with the highest loadings (-.91 to -.60) occurring for all reproductions by the lowpass filtered system (S_3) and for three reproductions by the body-worn system (S_4), viz. the reproductions of program sections 3, 4, and 6 which have relatively less treble content than the other program sections. The adjectives with the highest factor scores on this side were "dull", "subdued", "muddy", and "dark". This "dullness/darkness" dimension is thus mainly related to the degree of bass content (relative to treble content). It is apparently also related to F I, and at the oblique rotations the correlation between F I and F II was about +.35.

However, F II might be somewhat dubious. When the analyses were made on the corrected data it did not appear as a separate factor but was fused into F I, which again was interpretable as a "sharpness-softness" dimension (or possibly "sharpness-softness/dullness/darkness", that is, with a certain modification of one pole of the dimension).

F III finally was interpretable as a "clearness/distinctness" factor with the broad-band system (S_1) outstanding in factor loadings on one side (roughly .90 to .70). In contrast the head-worn system (S_2) and one or two reproductions by the other systems (S_3, S_4, S_5) had slightly negative loadings.

Three mode factor analyses

The results of the three-mode factor analyses will be described only in rather general terms. As pointed out by Tucker (1966) there are many questions to be considered when using three mode FA. The results will, for instance, be dependent on whether the data are scaled (normalized) in different ways. As with ordinary FA the results are further dependent on which type of rotation is applied (only the varimax rotation was tried in the present case). It will also make a difference which three of the four matrices (one for the program sections, one for the systems, one for the adjectives, and the so-called "core matrix" which represents the interrelations between the factors obtained in the three modes) are varimax rotated and, consequently, which one is compensationally rotated to fit the data.

A number of different analyses was tried on the present data, including different normalizations (among them that described in alternative g under Data treatment), different varimax rotations, and different number of factors. Looking at the resulting factor loading matrices for the adjectives (which are the easiest to interpret) one generally finds the same factors which were described for the earlier analyses: "sharpness-softness", "brightness-darkness", "clearness/distinctness", and "loudness". The adjectives denoting "room characteristics" are in most cases fused into the "clearness/distinctness" factor but appear as a separate factor in certain analyses, characterized by the adjectives "shut up", "narrow", and "distant" on one side versus "open", "airy", "wide", and "giving impression of room" on the other side.

The interpretation of factors for the systems was in most cases rather obvious (as, for instance, if the broad-band system was contrasted against all others, or if the lowpass filtered system was contrasted against the highpass filtered system). The interpretation of factors for the program sections was fairly obvious in some cases (for instance, if the two music sections were contrasted against all other sections) but also uncertain in many cases. Finally, the interpretation of the core matrix (which is important since it represents the relations between the factors for the adjectives, the programs, and the systems) was hard to do. Although there were suggestions to meaningful interpretations in some respects no quite convincing overall interpretation was obtained.

On the whole, then, the results from the three mode FA were in many respects similar to those from the earlier analyses. It is necessary to get more experience of three-mode FA in further experiments to make an evaluation of its applicability to the present problems.

DISCUSSION

A well-known problem in experiments of this type is that the resulting factors (dimensions) may depend on the specific selection of program sections, reproduction systems, adjectives, and individuals. It is thus necessary to investigate the generality of the resulting factors over a number of various selections in these respects. However, the dimensions described here agree in many respects with dimensions found for loudspeaker reproductions (Gabrielsson et al., 1971, 1972, 1973, 1974).

Another embarrassing point is the dependency of the results on the specific factor analytical procedure, for instance, whether one chooses to apply the simple structure criterion to the adjectives or to the P x S combinations and which type of rotation is applied to satisfy this criterion. Considering the exploratory purpose with this investigation, however, these problems are not so important. It is quite clear that essentially the same factors emerged from the different analyses tried here. And, above all, the resulting factors are only taken as working hypotheses for follow-up experiments in which their validity and the underlying psychophysical relations have to be investigated separately for each factor.

A problem is also the delicate shades of meaning for many of the adjectives. It is sometimes proposed that the meaning of an adjective may be clearer defined to the subjects if one also gives the opposite of the actual adjective. While this may be possible in certain cases it is not so simple in many other cases. Conversely, it is in some cases possible to think of several opposites. The whole question is still more complicated due to individual differences in the way of interpreting the adjectives. Still the present approach seems to work in a satisfactory way.

A related problem discussed earlier is the difficulty to separate the effects of the reproducing systems from the "effects" of the program sections as such. The correction of data described in alternative g under Data treatment might be of a certain value for the practical purposes at hand but presents no general "solution".

In summary the following perceptual dimensions (and attaching psychophysical relations) were suggested by the present experiment:

a) "Hardness/Sharpness - Softness". In the present context "hardness/sharpness" is related to the highpass filtering system (S_5) and to the head-worn system (S_2), the latter being characterized by a relatively small frequency range and introducing distortion components. "Softness" is related to the lowpass filtering system (S_3) and, to a certain extent, to the broad-band system (S_1) and the body-worn system (S_4), both of them with a broad frequency range and introducing relatively little distortion.

b) ("Brightness) - Darkness". In the present experiment this dimension did not appear quite so clear-cut as in some earlier experiments. However, it formed a separate factor at the FA on correlations between P x S combinations, the perceived "darkness/dullness" being related to the lowpass filtered system (S_3) and to the degree of bass content (relative to treble content) in the program sections.

There seem to be interesting relations between the two dimensions hitherto mentioned. As noted under Results the adjectives "bright" and "dark" got rather high factor loadings in the "hardness/sharpness - softness" dimension, for "bright" on the "hard/sharp" side and for "dark" on the "soft" side. Analogous phenomena were also found in Gabrielsson et al. (1973). Which of the factors will appear in the most obvious form might depend on the actual stimuli as suggested under Results. It can be noted, too, that the evaluative adjectives ("pleasant", "natural", "irritating", "painful", "tiring") get rather extreme loadings on the "soft" and "hard/sharp" side, respectively, while they get more neutral loadings in the "brightness-darkness" dimension (Gabrielsson et al., 1973).

c) "Clearness/Distinctness". As appearing here the perceived "clearness/distinctness" seems related to the frequency range (the greater, the better) and the amount of distortion (the less distortion, the better). Its relation to the present highpass and lowpass filtering seems to vary with the program sections. A "clear/distinct" reproduction, as examplified above all by the broad-band system (S_1), is contrasted against reproductions which are "muddy", "blurred", "noisy" and the like, sometimes also "restrained", "faint", and "subdued" (see also again the 1973 report). The specific meaning of "clear/distinct" may thus vary somewhat in different contexts.

In this dimension there also often appear adjectives denoting various "room characteristics" of the reproductions, see below.

d) "Impression of room". This is a rather loose term for a dimension in which there is a contrast between reproductions which are "open", "wide", "airy", give "impression of room" and reproductions which are "narrow", and "shut up". The adjectives "near" and "distant" also often appear on the respective sides of this dimension.

In fact these adjectives most often appear fused into a "clearness/distinctness" dimension or sometimes into a "loudness" dimension. In general the "clear/distinct" reproductions and/or the "loud" reproductions are also those which are being characterized as "open", "wide", "airy", etc. In some analyses, however, the "room dimension" appeared as a separate factor. Its relative importance and independence may reasonably be expected to vary with the stimulus context.

e) "Loudness". A separate loudness dimension was found in at least one analysis, contrasting adjectives as "soft/gentle", "restrained", "faint" against "loud", and "ringing". In most analyses, however, it seems that "loud" and "soft/gentle" are fused into some of the other dimensions.

Once again it should be emphasized that the listed dimensions and the suggested psychophysical relations are considered only as working hypotheses for further research which is in progress. An important thing to notice in this context is the frequently occurring interactions between program sections and systems, some examples of which were mentioned under Results (many more could be given).

In summary it seems that the results provide a good basis for further research on perceived sound quality of hearing aid systems. It is necessary, of course, to gradually introduce people suffering from hearing loss and using various types of hearing aids. Such experiments are now under preparation.

ACKNOWLEDGEMENTS

The authors express their gratitude to Björn Lindström, who served as experimenter and took part in the practical planning of the experiment.

REFERENCES

Bäckström, L. (1967). Datamaskinprogram II. Department of Psychology, University of Uppsala.

Gabrielsson, A. (1974). An empirical comparison between some models for multidimensional scaling. Scand. J. Psychol., 14, 73 - 80.

Gabrielsson, A., Rosenberg, U. & Sjögren, H. (1971, 1972). Judgments and dimension analyses of perceived sound quality of sound-reproducing systems. I and II. Rep. Psychol. Lab. Univ. Uppsala, No. 115 - 116 (also as Rep. Techn. Audiology, Kar. Inst., Stockholm, No. 66 and 68).

Gabrielsson, A., Rosenberg, U. & Sjögren, H. (1973). Adjective ratings and dimension analyses of perceived sound quality of sound-reproducing systems. Rep. Psychol. Lab. Univ. Uppsala, No. 141 (also as Rep. Techn. Audiology, Kar. Inst., Stockholm, No. 72).

Gabrielsson, A., Rosenberg, U. & Sjögren, H. (1974). Judgments and dimension analyses of perceived sound quality of sound-reproducing systems. J. Acoust. Soc. Amer., 55, 854 - 861.

Gorsuch, R.L. (1974). Factor analysis. W.B. Saunders Comp., Philadelphia.

Gruvaeus, G., Wainer, H. & Snyder, F. (1971). TREMOD: A 360/75 FORTRAN program for three-mode factor analysis. Behavioral Science, 16, 421 - 422.

Harman, H.H. (1967). Modern factor analysis (rev. ed.). University of Chicago Press.

Levin, J. (1965). Three-mode factor analysis. Psychol. Bull., 64, 442 - 452.

Muthén, B. (1973). TMFAROT - a computer program for varimax rotation in three-mode factor analysis. Department of Statistics, University of Uppsala.

Osgood, C.E., Suci, G.J. & Tannenbaum, P.H. (1957). The measurement of meaning. University of Illinois Press.

Sjöberg, L. (1972). Factor analysis and multidimensional scaling. Department of Psychology, University of Gothenburg.

Snider, J.G. & Osgood, C.E. (Eds.) (1969). Semantic differential technique. Aldine Publishing Company, Chicago.

Tucker, L.R. (1966). Some mathematical notes on three-mode factor analysis. Psychometrika, 31, 279 - 311.

TABLE I. Factor loadings for adjectives at varimax rotation and at direct oblimin rotation.

Adjectives	Varimax F I	Varimax F II	Direct oblimin F I	Direct oblimin F II	Direct oblimin F III
Avlägsen ("Distant")	.62	.67	.70	.36	.28
Behaglig ("Pleasant")	-.91	-.38	-.95	-.05	-.15
Briljant ("Brilliant")	-.34	-.88	-.45	-.47	-.48
Brusig ("White noise")	.80	.45	.84	-.06	.33
Bullrig ("Noisy/Rumbling")	-.09	.56	.03	.97	-.14
Dov ("Dull")	-.83	.46	-.76	.47	.32
Dämpad ("Subdued")	-.64	.71	-.55	.46	.55
Frän ("Harsh")	.97	.11	.98	-.05	-.08
Fräsande ("Hissing")	.87	.26	.89	-.12	.16
Fyllig ("Full/-toned/")	-.85	-.44	-.88	.07	-.32
Grov ("Coarse/Gruff")	-.44	.43	-.34	.90	-.15
Grumlig ("Muddy/Confused")	.10	.90	.23	.67	.41
Gäll ("Shrill")	.98	-.02	.98	-.11	-.17
Hård ("Hard")	.98	.06	.99	.06	-.21
Ihålig ("Hollow")	.84	.40	.89	.06	.17
Instängd ("Shut up")	.53	.78	.61	.25	.50
Irriterande ("Irritating")	.94	.32	.97	.04	.07
Jämn ("Uniform/Smooth")	-.89	-.27	-.93	-.12	.02
Kall ("Cold/Chilly")	.95	.18	.97	-.07	.02
Klar ("Clear")	-.35	-.86	-.47	-.61	-.35
Klingande ("Ringing")	.17	-.84	.08	-.36	-.65
Ljus ("Bright/Light")	.91	-.26	.86	-.39	-.18
Luftig ("Airy")	-.67	-.64	-.74	-.08	-.45
Lugn ("Calm/Quiet")	-.97	-.09	-.99	-.05	.18
Matt ("Faint/Feeble")	-.59	.74	-.51	.35	.65
Mild ("Mild")	-.97	-.14	-.99	-.05	.12
Mjuk ("Soft")	-.99	-.08	-.99	.05	.11
Monoton ("Monotonous")	.02	.86	.12	.33	.65
Mullrande ("Rumbling")	-.60	.51	-.51	.80	.05
Mörk ("Dark")	-.86	.35	-.80	.49	.19
Naturlig ("Natural")	-.80	-.57	-.86	-.16	-.29

TABLE I continued

Adjectives	Varimax F I	Varimax F II	Direct oblimin F I	Direct oblimin F II	Direct oblimin F III
Nära ("Near")	-.67	-.56	-.75	-.34	-.17
Orolig ("Agitated")	.93	.27	.97	.16	-.07
Plågsam ("Painful")	.96	.20	.98	-.04	.02
Påträngande ("Obtrusive")	.94	.18	.97	.06	-.08
Ren ("Clean/Pure")	-.65	-.71	-.74	-.39	-.29
Rogivande ("Soothing")	-.94	-.26	-.97	.01	-.05
Rumskänsla ("Impression of room")	-.70	-.61	-.76	.00	-.48
Rå ("Raw/Crude")	.92	.27	.96	.21	-.10
Skarp ("Sharp")	.99	-.05	.98	-.16	-.16
Skrapande ("Scraping")	.93	.23	.94	-.13	.12
Skrikig ("Screaming")	.97	.04	.97	-.09	-.12
Skrovlig ("Rough/Raucous")	.79	.43	.85	.34	.00
Skrällig ("Clashing")	.98	.04	.99	.05	-.23
Skärande ("Shrill")	.98	.06	.99	-.08	-.11
Spetsig ("Pointed")	.98	.00	.98	-.20	-.07
Sprucken ("Cracked")	.97	.20	.99	-.04	.01
Stark ("Loud")	.83	-.28	.82	.19	-.64
Stickande ("Pricking/Sharp")	.98	.05	.97	-.20	-.03
Sträv ("Harsh")	.83	.34	.87	.10	.08
Stötig ("Knocking")	.89	.15	.90	-.03	-.02
Suddig ("Blurred")	.21	.84	.34	.73	.27
Svag ("Soft/Gentle"; opposite to "loud")	-.09	.75	-.04	-.19	.96
Tillbakahållen ("Restrained")	.01	.92	.12	.38	.66
Trång ("Narrow")	.74	.60	.81	.14	.35
Tröttande ("Tiring")	.90	.39	.95	.14	.09
Tunn ("Thin")	.95	.15	.95	-.22	.11
Tydlig ("Distinct/Clear")	-.51	-.74	-.62	-.61	-.18
Vass ("Sharp/Keen")	.99	.03	.99	-.16	-.09
Vid ("Wide")	-.71	-.61	-.76	.10	-.55
Vinande ("Whistling")	.92	.21	.94	-.07	.05
Öppen ("Open")	-.49	-.80	-.57	-.13	-.62

Table II. Rank order of P x S combinations in factor scores at the varimax rotated solution on original data and on corrected data.

Factor I		Factor II	
Original	Corrected	Original	Corrected
P_6S_5 (1.62)	P_6S_5 (1.77)	P_6S_3 (1.50)	P_6S_3 (1.51)
P_5S_5	P_3S_5	P_4S_3	P_1S_3
P_2S_5	P_5S_5	P_4S_5	P_3S_3
P_3S_5	P_4S_5	P_3S_3	P_2S_5
P_1S_5	P_2S_5	P_4S_4	P_4S_3
P_2S_2	P_4S_2	P_6S_2	P_5S_2
P_5S_2	P_1S_5	P_1S_3	P_4S_5
P_4S_5	P_2S_2	P_3S_2	P_6S_2
P_1S_2	P_3S_2	P_3S_4	P_2S_2
P_3S_2	P_6S_2	P_4S_2	P_2S_3
P_6S_2	P_5S_2	P_5S_2	P_5S_3
P_4S_2	P_1S_2	P_1S_2	P_1S_2
P_5S_4	P_2S_4	P_3S_5	P_3S_2
P_2S_4	P_5S_4 (.14)	P_6S_4	P_4S_4
P_1S_4 (.17)	P_1S_4 (-.03)	P_2S_5	P_5S_4
P_1S_1 (-.15)	P_1S_1	P_1S_4	P_3S_4
P_5S_1	P_3S_4	P_5S_3 (.01)	P_1S_4 (.18)
P_3S_4	P_4S_4	P_6S_5 (-.15)	P_6S_4 (-.02)
P_6S_4	P_5S_1	P_5S_4	P_5S_5
P_2S_1	P_6S_4	P_1S_5	P_4S_2
P_6S_1	P_6S_1	P_2S_2	P_2S_4
P_3S_1	P_3S_1	P_5S_5	P_1S_5
P_4S_4	P_4S_1	P_2S_3	P_3S_5
P_1S_3	P_2S_1	P_3S_1	P_6S_5
P_5S_3	P_4S_3	P_2S_4	P_3S_1
P_4S_1	P_1S_3	P_6S_1	P_5S_1
P_3S_3	P_3S_3	P_4S_1	P_6S_1
P_6S_3	P_6S_3	P_1S_1	P_1S_1
P_2S_3	P_2S_3	P_5S_1	P_2S_1
P_4S_3 (-1.60)	P_5S_3 (-1.60)	P_2S_1 (-2.59)	P_4S_1 (-2.07)

TABLE III. Rank order of P x S combinations in factor scores a at the direct oblimin solution on corrected data.

F II	F III
P_6S_3 (1.27)	P_4S_5 (1.70)
P_5S_2	P_6S_3
P_1S_2	P_2S_5
P_1S_3	P_1S_3
P_4S_2	P_3S_3
P_4S_3	P_4S_3
P_3S_3	P_5S_3
P_3S_2	P_2S_3
P_6S_2	P_5S_2
P_2S_2	P_6S_2
P_3S_4	P_5S_5
P_4S_4	P_6S_4
P_5S_4	P_2S_2
P_2S_3	P_1S_5
P_5S_3	P_4S_4
P_1S_4	P_3S_5
P_2S_4 (.25)	P_1S_4 (.07)
P_6S_4 (-.38)	P_5S_4 (-.03)
P_2S_5	P_3S_4
P_6S_1	P_1S_2
P_6S_5	P_3S_2
P_5S_1	P_2S_4
P_1S_5	P_6S_5
P_2S_1	P_4S_2
P_4S_5	P_3S_1
P_5S_5	P_1S_1
P_3S_5	P_5S_1
P_3S_1	P_6S_1
P_1S_1	P_4S_1
P_4S_1 (-1.63)	P_2S_1 (-1.93)