KONGL CAROLINSKA MEDICO CHIRURGISKA INSTITUTET

TECHNICAL AUDIOLOGY

Report TA No. 103
December 1981

PROBLEMS AND METHODS IN JUDGMENTS OF PERCEIVED SOUND QUALITY

Alf Gabrielsson

Karolinska Institutet
Dept. of Techn. Audiology
KTH, S-100 44 Stockholm
SWEDEN

Report TA 103
December 1981

PROBLEMS AND METHODS IN JUDGMENTS OF PERCEIVED SOUND QUALITY

Alf Gabrielsson

Invited paper to special session on "Interdisciplinary Contributions to the Study of Noise Annoyance" at the 102nd Meeting of The Acoustical Society of America, 30 November - 4 December, Carillon Hotel, Miami Beach, Florida, USA.

Karolinska Institutet
Dept. of Techn. Audiology
KTH, S-100 44 Stockholm
SWEDEN

Report TA 103
December 1981

PROBLEMS AND METHODS IN JUDGMENTS OF PERCEIVED SOUND QUALITY

Alf Gabrielsson

ABSTRACT

The perceived sound quality of any auditory stimulus (music, speech, noise) may be expressed in terms of overall evaluations as well as in terms of specific perceptual qualities. Examples of the former are variables as pleasantness, annoyance, and fidelity, and of the latter perceptual variables as loudness, clarity, sharpness etc. Studying the relations between overall evaluations and separate variables and their respective relations to the stimulus characteristics may increase the understanding of factors influencing perceived sound quality. The associated methodological problems will be illustrated by examples from research on perceived sound quality of sound-reproducing systems (A. Gabrielsson and H. Sjögren, J. Acoust. Soc. Am. 65, 1019-1033 (1979)). Some possible parallels, as well as dissimilarities, between this research and research on noise-induced annoyance will be discussed.

The research have been supported by the Swedish National Board for Technical Development, the Swedish Council for Planning and Cooperation of Research, the Swedish Telecommunications Administration, the National Swedish Board for Consumer Policies and the Swedish Philips Group.

INTRODUCTION

Judgments concerning auditory stimuli (music, speech, noises etc.) may be expressed in terms of overall evaluations as well as in terms of specific perceptual dimensions. Examples of overall evaluations are expressions for acceptance/rejection, pleasantness/unpleasantness, preference, annoyance, and others. Examples of perceptual dimensions may be such as loudness, clarity, sharpness, dullness, and so on. There are in fact hundreds of adjectives which are used to characterize the sound quality or timbre of various sounds. This indicates that perceived sound quality is a multidimensional phenomenon, that is, it is constituted by a number of separate perceptual dimensions. It is important, both for basic research and for applications, to understand which these perceptual dimensions are, and how they are related to overall evaluations. It is equally important to understand the underlying psychophysical relations, that is, the relations between the physical/acoustical properties of the auditory stimuli and the various perceptual dimensions. If we know these psychophysical relations and the relations between separate perceptual dimensions and overall evaluations, we may have possibilities for manipulating the acoustical properties of the stimuli to bring about certain intended or desired perceptual qualities with consequences for the overall evaluation, look at Figure 1.

For methodological and other reasons the present knowledge in these matters is rather limited. In the following we will discuss some possible approaches, using examples from research on the perceived sound quality of sound-reproducing systems, such as loudspeakers, headphones, and hearing aids (Gabrielsson 1979; Gabrielsson & Sjögren 1979a, 1979b). However, the principles are easily generalized to other types of auditory stimuli.

METHODS AND RESULTS

To find the relevant perceptual dimensions in sound quality you can use multivariate techniques like multidimensional scaling and factor analysis. These are supplemented by unidimensional scaling and certain other techniques to analyze the relations to overall evaluations and to physical/acoustical properties.

Multidimensional scaling

Multidimensional scaling (MDS) started in the late 1930s and nowadays comprises many different models and associated

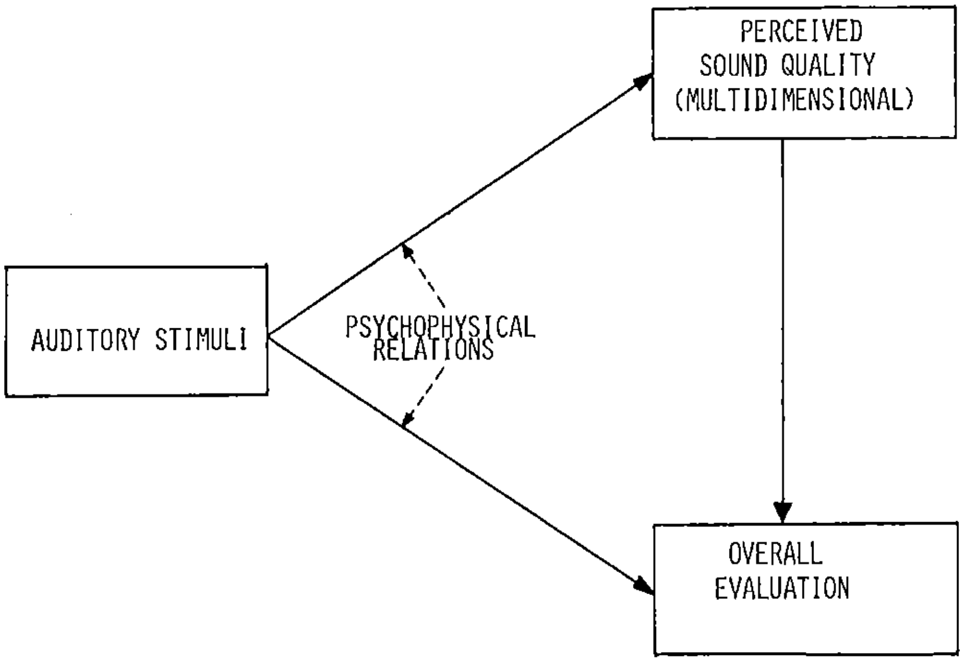
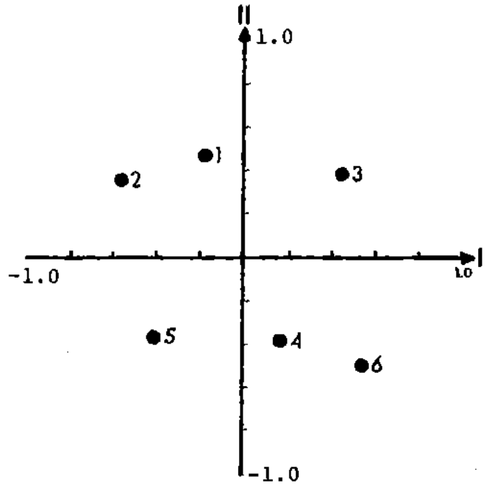


Figure 1. General framework for problems discussed in this paper.

methods. Only some basic features of MDS can be treated here. Suitable references for deeper studies are Torgerson (1958), Shepard, Romney & Nerlove (1972), Kruskal & Wish (1978), and Carroll & Arabie (1980).

The purpose of MDS is to find out the constituent dimensions of a multidimensional phenomenon (e.g. perceived sound quality of loudspeakers) and to perform a scaling of the actual stimuli in these dimensions. The most common model (often called the "distance model") operates with geometrical concepts. The stimuli (e.g. loudspeakers) are represented by points lying in a space of n dimensions. The distances between the points represent the perceived similarity (or dissimilarity) between the stimuli. The closer two points are to each other, the more similar the corresponding two stimuli are to each other, and conversely - the more distant they lie from each other, the less similar (more dissimilar) they are. The axes of the space correspond to the perceptual dimensions we are looking for, and the projections of the points on the different axes represent the scale values of the respective stimuli in the different dimensions.

For a simple example look at Figure 2a, which represents the configuration of six loudspeaker reproductions in a two-dimensional space. For instance, stimuli nos. 4 and 6



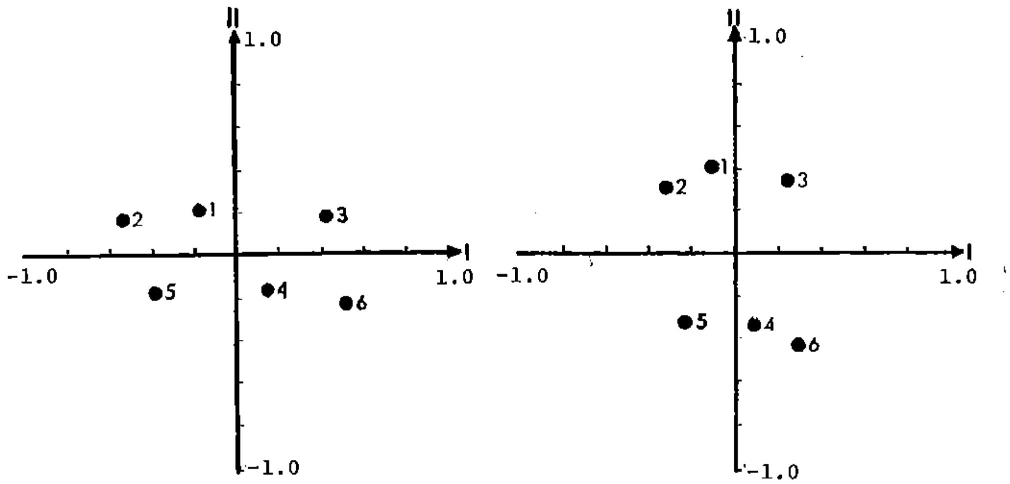


Figure 2a.(Upper part). Configuration of six reproductions in a two-dimensional space.

Figure 2b.(Lower part). Configuration of six reproductions for a high-fidelity experienced listener (left) and for a non high-fidelity experienced listener (right). (From Scand. J. Psychol., 1974, 15, p. 75 and 78.)

are closer/more similar to each other than stimuli nos. 3 and 6. Their scale values in the two dimensions are obtained by projecting the respective points on the two axes. For instance, nos. 3, 4 and 6 all have positive scale values in dimension I(horizontal axis), no. 3 also has a positive scale value in dimension II, in which nos. 4 and 6 have negative values, and so on (this example is treated further below).

In practice we gather information about the perceived similarity (or dissimilarity) between our stimuli, e.g. by having listeners rate the similarity between any two stimuli (in all possible pairwise combinations) on a scale from "perfect similarity" to "minimum similarity". The similarity data are processed according to one of the available algorithms to result in a representation of the stimuli as points in an *n*-dimensional space. Since there are many possible solutions, it is necessary to have some criterion for the "best" solution. A usual criterion is that there should be as good fit as possible between the obtained (dis)similarities and the corresponding distances derived from the respective solution. Furthermore the number of dimensions should be "small" in relation to the number of stimuli, and it should be possible, of course, to give the dimensions a meaningful perceptual interpretation.

The example in Figure 2 refers to MDS of six different reproductions of a certain piece of music (Gabrielsson, 1974; Gabrielsson, Rosenberg & Sjögren, 1974). The reproductions were systematically varied in sound level and treble response. Thus reproductions 1 - 3 had a "normal" sound level and reproductions 4 - 6 a 6dB decreased level. With regard to the treble response, reproductions 1 and 4 had a flat response, 2 and 5 an increased response (up to +6dB at 10 kHz), and 3 and 6 a decreased response (down to -6dB at 10 kHz). The six reproductions were judged for similarity among themselves. As expected, a two-dimensional solution proved to be adequate. The first dimension (horizontal axis) is related to the types of treble response (reproductions 2 and 5 with increased response to the left, 3 and 6 with decreased response on the other side, and 1 and 4 with flat response in the middle) and was perceptually interpreted as "brightness vs dullness". The second dimension reflects the different sound levels (reproductions 1 - 3 with normal level versus reproductions 4 - 6 with decreased level) and was simply interpreted as "loudness".

This result was obtained both with so-called metric and non metric models for MDS (in the non metric case there is "only" a monotonic relationship between (dis)similarities and distances) and after rotating the axes to fit our hypothesis about the above-mentioned two dimensions. An interesting re-analysis of the same data was made according to the recent INDSCAL model (Carroll & Chang, 1970). This model allows for possible individual differences in the weightings of different dimensions. For instance, in the example here some individuals may find differences in loudness more important than differences in brightness - or conversely. The introduction of individual weightings for different dimensions results in a unique solution (no rotation required), and it is also possible to derive how the configuration of stimuli looks in the individual case. An example is given in Figure 2b. The left configuration was obtained for a "high fidelity experienced" listener, while the configuration to the right was found for a "non high fidelity experienced" listener. It is apparent that the former gives relatively more weight to the differences in brightness - dullness than to differences in loudness (the configuration is expanded along the horizontal axis but contracted along the vertical axis), while the situation is the opposite for the latter (non hi-fi) subject - which in fact seems plausible.

The interpretation of the dimensions obtained by MDS may be more difficult than in the example given here. Various kinds of information can be used for the interpretation: physical/acoustical data, verbal responses (formal or informal) from the listeners, results from earlier experiments etc. Further examples are given in Gabrielsson (1979) and Gabrielsson & Sjögren (1979a, 1979b). It is generally recommended to check the validity of your interpretations in follow-up experiments.

Factor analysis

Factor analysis (FA) was introduced into psychology in the beginning of this century as a means to study the structure of intelligence. Since then it has expanded enormously, both with regard to different models and as regards applications. In the present context only some basic ideas of FA can be given together with some examples on the study of perceived sound quality. For more thorough studies see Harman (1967), Gorsuch (1974) or Cureton & D'Agostino (1981).

The basic principle of FA is simple. Assume that we have a situation with data in a big number of variables. If we look at the statistical correlation between these variables, we may find that certain variables are substantially inter-correlated, positively or negatively. This is an indication that they have something in common, and that they may be seen as different manifestations of some common "fundamental" factor(s). There may be more such groups of inter-correlated variables, and generally the purpose of FA is to account for the variation in a big number of observed variables in terms of a small number of inferred fundamental factors - look at Figure 3.

The model for so-called component analysis, which is the alternative mostly used for investigations on perceived sound quality, can be written as

$$X_{iv} = w_{v1} F_{1i} + w_{v2} F_{2i} + w_{v3} F_{3i} + \dots + w_{vf} F_{fi} \qquad (1)$$

X_{iv} represents the observed score for individual i in variable v (expressed in a form called standard score). This score is written as a weighted linear function of individual i's (standard) score in factor 1 (denoted F_{1i}), factor 2 (F_{2i}), factor 3 (F_{3i}), and so on to f factors. Each of these scores is combined with a weighting coefficient w, which represents the weight for variable v in the respective factor: w_{v1} is the weight for variable v in factor 1, w_{v2} the weight for variable v in factor 2, and so on. (The designations are adapted from Gorsuch, p. 18-19). The individual's score in an observed variable is thus a function of his scores in a

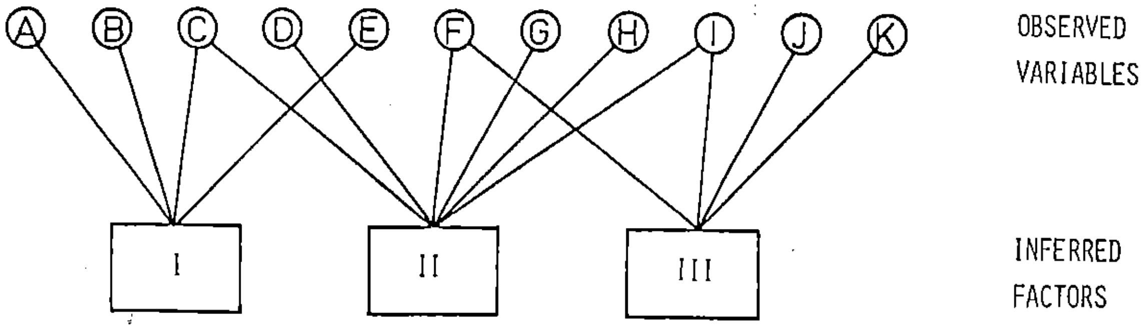


Figure 3. Correlations between observed variables indicate that they are different manifestations of common fundamental factors to be inferred by factor analysis.

number of fundamental factors multiplied by the weight coefficient for the variable in the respective factors. The weight coefficents are usually called the factor loadings of the variable in the respective factors. It can further be shown that the correlation between two observed variables is a function of the cross products of their factor loadings. The correlation r_{ab} between the variables a and b is thus

$$r_{ab} = w_{a1} w_{b1} + w_{a2} w_{b2} + \dots + w_{af} w_{bf} \quad (2)$$

(assuming uncorrelated factors). Since the correlations are given from our empirical data, and since the number of such correlations is bigger than the number of the unknown factor loadings, it is possible to solve the equations for the factor loadings. Mathematically there are many possible solutions, and thus a criterion must be established to determine which is the "best" solution. A general criterion is that of "simple structure", the detailed meaning of which may be dependent on the specific application (it could mean, for instance, that each variable should be loaded in as few factors as possible). The process by which the "best" solution is sought is called "rotation of factors", hinting at a geometrical representation of FA, which is partly similar to that in MDS. Further problems to be considered concern the number of factors that should be judged as sufficient to account for the variation in the observed variables and, of course, how to interpret the factors.

It is impossible here to go deeper into these and many other problems in FA. In practice there are many efficient computer programs, by which you can compare the results from different decisions regarding rotation, number of factors etc. It is recommended to make such comparisons in order to get a feeling for how FA works in your own application area. In the following we will briefly describe some applications of FA for studying perceived sound quality.

In such investigations the "individuals" are usually reproduction systems (loudspeakers, headphones etc.), and the observed variables are a number of adjective scales. In our investigations (Gabrielsson, 1979; Gabrielsson & Sjögren, 1979a, 1979b) the adjectives were selected by sound engineers, audiologists and others to characterize possible features of perceived sound quality. In the experiments the subjects listen to various reproductions, one at a time, and rate the respective reproduction on each adjective scale graded from a "minimum" to a "maximum". The number of adjectives varied between 30 and 55 in different experiments. FA was then applied to the correlations between the adjective scales (correlations over all reproductions, programs and listeners) to see if the variation in the many adjective scales could be accounted for in terms of a smaller number of "fundamental" factors (compare Figure 3). The result is a factor loading matrix, in which each of the adjectives has a loading (ranging from -1.00 to +1.00) in each of the inferred factors. These factor loadings are those obtained after rotation to the "best" solution. An example is given in Table I. This refers to an investigation of headphones, which were rated in 30 adjective scales by 20 subjects for five different music programs.

The perceptual interpretation of the factors is made by looking at which adjectives have high loadings, positive and/or negative, in each factor and trying to determine what those adjectives have in common. The interpretation of the present five factors presents no big problems. Factor I has one high positive loading for "soft" (+0.84) and some high negative loadings ("hard" -0.76, "jarring/grating" -0.83, "loud"-0.92, and "sharp/keen"-0.66). This was interpreted as "loudness/sharpness/hardness vs. softness". Using the analogous procedure, factor II was interpreted as "clearness/distinctness", factor III as "disturbing sounds", factor IV as "brightness vs. dullness", and factor V as "feeling of space/spaciousness".

A further result of the FA is a matrix of so-called factor scores. These correpond to the values denoted by F_{1i}, F_{2i} etc. in equation (1). In the present case they mean the scores of each reproduction in each of the factors. (The computation of these scores is rather intricate and is skipped here.) Thus we have a kind of scaling of the reproductions within each perceptual factor (dimension). This also gives a possibility to study the psychophysical relations by looking at the physical properties of the reproductions which differ

	FACTORS				
ADJECTIVE	1	2	3	4	5
"BALANCED"	.28	.59	-.06	-.32	.18
"PLEASANT"	.42	.39	-.20	-.21	.34
"WHITE NOISE/HISSING"	.48	-.10	.88	-.01	.05
"DIFFUSE"	.12	-.90	-.05	-.17	.18
"DULL"	.18	-.31	-.25	-.71	-.12
"EMPHASIZED BASS"	.00	.14	.11	-.93	-.21
"EMPHASIZED TREBLE"	-.43	.22	.39	.37	-.16
"HISSING"	-.04	.04	.88	.07	-.10
"FULL/-TONED"	.22	.18	-.13	-.45	.47
"HARD"	-.76	.31	-.12	.16	-.31
"HOLLOW"	.01	-.43	-.10	.20	-.50
"SHUT UP/CLOSED"	-.11	-.35	-.09	-.10	-.71
"CLEAR"	-.21	.85	-.02	-.02	.01
"CRACKLING/CRUNCHING"	.07	.04	.95	-.08	-.07
"BRIGHT/LIGHT"	-.20	.07	.18	.77	.00
"FAINT/FEEBLE"	.34	-.55	-.15	-.07	-.45
"SOFT"	.84	-.06	-.04	-.17	.12
"RUMBLING"	-.26	-.20	-.06	-.93	.10
"NASAL"	-.25	-.45	.07	.52	-.18
"TRUE TO NATURE"	.09	.75	-.06	-.23	.20
"FEELING OF PRESENCE"	.23	.64	.01	-.17	.31
"PURE/CLEAN"	-.01	.83	-.13	.13	.11
"FEELING OF SPACE"	.02	-.01	-.13	.07	.92
"JARRING/GRATING"	-.83	-.03	.09	-.11	-.15
"LOUD"	-.92	.01	-.24	-.08	.42
"HARSH"	-.49	-.04	.30	.29	-.17
"DRY"	-.20	-.15	.33	.30	-.44
"THIN"	-.07	-.25	.06	.65	-.31
"SHARP/KEEN"	-.66	.09	.20	.37	-.06
"WHISTLING/WHIZZING"	-.25	-.09	.82	.05	.18

<u>Table I</u>. Example of factor loading matrix (Gabrielsson, Frykholm & Sjögren, 1977).

in their factor scores. For instance, the reproduction with the highest factor score on the "sharp/hard" side in factor I had a pronounced peak in its frequency response around 3 kHz, while the reproductions on the "soft" side of this factor did not have such a peak. However, for more detailed knowledge about scale values, psychophysical relations etc. it is definitely recommended to make special follow-up experiments.

Comparison MDS - FA

There are various problems associated with the use of MDS and FA for the applications described here. With adjective ratings followed by FA it is apparent that which factors will result from the FA depend on the selection of adjectives - look at Figure 3 again. Thus this selection has to be made with great care. If you omit some adjectives of importance for a certain aspect of perceived sound quality, this aspect may not turn up as a factor in the FA. On the other hand, you have the risk of including more adjectives than necessary to reflect a certain factor. By your selection you can direct the listener's attention to various aspects. This may be an advantage in the sense that he may hear things, that he would not have done otherwise. But some of these are perhaps rather irrelevant in the listener's opinion, while he might ask for some other adjectives which you have forgotten. There are also problems with possible individual differences in how words are used and interpreted. These problems may be partly handled by checking the inter-rater reliability for different adjectives.

With (dis)similarity ratings and MDS some language problems are circumvented, and the listener can make his own selection of which aspects to attend. However, since there are many aspects to be combined into a single (dis)similarity rating, this may not be an easy task for the listener. He may deliberately omit certain aspects because he cannot manage to include them in his judgments, or he may vacillate between different principles from time to time. Furthermore, different listeners do this in different ways.

It is obvious that none of these techniques is "perfect", nor is any of them definitely superior to the other. They have to be used in combination with each other and also in combination with other relevant information. You will find a direct comparison between the two methods in two experiments described in Gabrielsson (1979). Eight hearing aids were used in both experiments, one with adjective ratings + FA, one with similarity ratings + MDS (INDSCAL). In both experiments the best solution contained three factors/dimensions. Two of these were similar for both experiments, while the interpretation of the remaining dimension was different (but meaningful) in the different experiments.

Generally both methods are sensitive to context effects. The context of stimuli, listeners, specific rating scales, adjectives etc. may affect the results in different ways. Unless the results should apply to a specific context, it is therefore necessary to perform many experiments with varying contexts to see which results remain invariant. The combined results from eight experiments on sound-reproducing systems are summarized in Gabrielsson & Sjögren (1979a). Eight perceptual dimensions are identified and labelled as follows (translated from Swedish): Clearness/Distinctness, Sharpness/Hardness vs. Softness, Brightness vs. Dullness, Fullness vs. Thinness, Feeling of space/Spaciousness, Nearness/Closeness, Disturbing sounds, and Loudness. Their relations to the physical properties of the sound-reproducing systems (properties as the frequency response, non-linear distortion etc.) are suggested on the basis of the positions of the different systems in the perceptual dimensions as found by MDS or FA.

Supplementary unidimensional scaling

The MDS and FA models are fairly complex and rely on certain assumptions, which may not be satisfied in various applications. The results are also sensitive to various factors as discussed above. It is therefore generally recommended to check the validity of the obtained dimensions in separate experiments. This can be done by taking one dimension at a time, that is, perform unidimensional scaling of stimuli in each of the suggested dimensions. If the results are consistent, this may be taken as evidence for the validity of the respective dimension. Of course, accurate unidimensional scaling is generally desirable and necessary if you want to study the underlying psychophysical relations with good precision.

Attempts at unidimensional scaling of the above-mentioned dimensions are described in two recent reports (Gabrielsson, Frykholm & Lindström, 1979; Gabrielsson & Lindström, 1981). Besides ratings in perceptual scales, ratings are also made on one or two evaluative scales, namely Fidelity and Pleasantness. All scales are graded from 0 (minimum) to 10 (maximum) with definitions also attached to certain scale numbers as seen in Figure 4. The results indicate that it is possible to differentiate among different loudspeakers etc. in these scales, and that all perceptual scales are correlated with the evaluative scales. In the continued work special consideration will be given to problems associated with the use of "internal references" and "perceptual references" for doing the scaling, and with differences in judgments when the stimuli are presented separately and in pairwise combinations, respectively.

The stimuli we use are generally varying over time (music, speech etc.). A single rating value for such a varying stimulus probably represents some kind of averaging over time for the characteristic in question. It can be assumed, however, that certain parts of the stimulus are more important than others for the rating (e.g., dynamic peaks). Quite

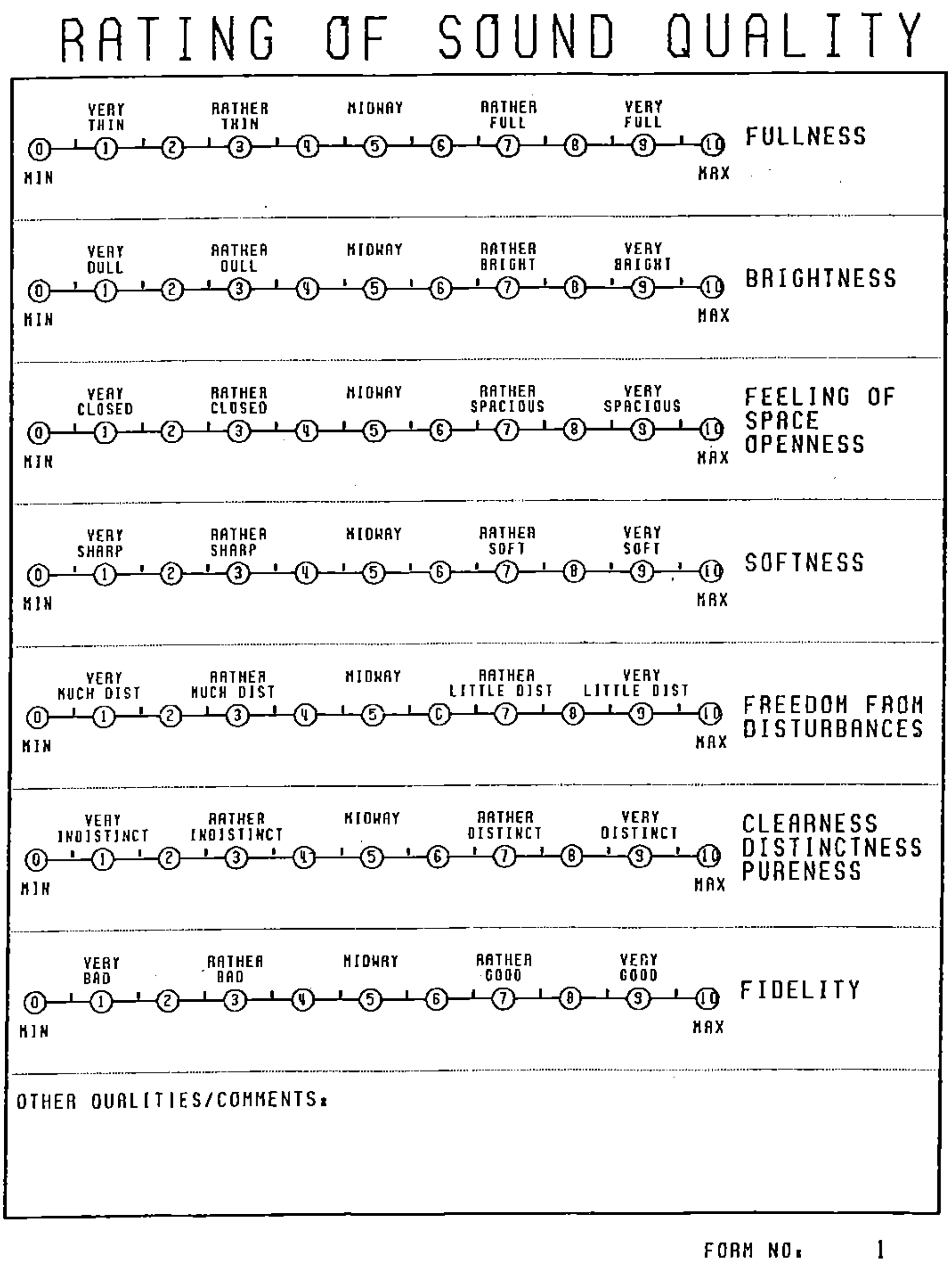


Figure 4. Examples of rating scales used for unidimensional scaling and for "Fidelity".

generally it would be highly informative to trace the percept continuously during the whole duration of the stimulus. Preparations are now being made to realize a continuous response registration in various scales. A variant of this, "continuous judgment by category", has been tried for "noisiness" of aircraft noises by Namba & Kuwano (1980).

Overall evaluations

Overall evaluations of the reproduction systems are usually made on scales referring to the Fidelity (in the "high fidelity" sense) and the Pleasantness of the reproduction. In some experiments also expressions for negative evaluations were included, e.g. "painful", "irritating", or "tiring".

The relations between the perceptual scales and the overall evaluation can be studied in different ways. In all methods described here you have some kind of scale values for the stimuli in the different perceptual scales as well as in the evaluative scale(s). You can then compute the correlation between each perceptual scale and the evaluative scale(s). You can also compute the multiple correlation between all perceptual scales together and the evaluative scale(s), and use multiple linear regression procedures to describe the overall evaluation as a weighted linear function of the ratings in the perceptual scales, that is,

$$R_e = w_1 R_{p1} + w_2 R_{p2} + w_3 R_{p3} + \dots + w_n R_{pn} \qquad (3)$$

where R_e is the overall evaluation, R_{p1} rating in first perceptual scale, R_{p2} rating in second perceptual scale etc. and w_1, w_2 etc. the appropriate weights to obtain maximum multiple correlation.

In the FA:s on adjectives we also have the simple procedure to look at the factor loadings for the evaluative adjectives in the respective factors and/or to look at their correlations with single adjectives.

The combined results from these procedures applied to many different experiments suggest the following general relations between the perceptual dimensions and the overall evaluations. In order to sound "pleasant" and or "natural" (fidelity) the sound reproduction should be "clear/distinct", "full-toned", "bright" (rather than "dull"), "soft" (rather than "sharp"), provide "feeling of space" (not sound "closed") and be free from any disturbing sounds. To write this in mathematical form as in Equation (3) above is somewhat risky, because the relative weights of the different perceptual dimensions can vary very much depending on the stimulus context. It is interesting to note that the relations suggested here are confirmed in ratings of so-called "ideal" values for sound reproduction systems (Gabrielsson,Frykholm & Lindström, 1979; Gabrielsson & Lindström, 1981). The evaluative scales, Pleasantness and Fidelity, are in general highly positively inter-correlated. However, they may differ somewhat in their relations to the perceptual scales. For instance, a "pleasant" reproduction is often "softer" and not as "clear/distinct" as what is required for fidelity.

Turning now to expressions for negative evaluations - such as "painful", "irritating" and "tiring" - they prove to be highly correlated with the "Sharpness/Hardness" dimension. In certain experiments there is an almost complete identification between the negative evaluation and "Sharpness/Hardness", in other experiments the negative evaluation is also correlated with "Indistinctness" and various "Disturbing sounds". (Remember that these data refer to evaluations of sound-reproducing systems, and that they may depend much on the stimulus context.) The physical factors behind "Sharpness/Hardness" in these experiments are mainly peaks in the frequency response at 1 - 4 kHz (that is, in the most sensitive region of the ear) and to some extent nonlinear distortion. Details about the frequency responses are given in Gabrielsson (1979) and Gabrielsson & Sjögren (1979b).

Quite generally the relations between various perceptual dimensions and negative evaluations can be given as the opposite of what was stated above concerning their relations to Pleasantness and Fidelity. Thus an "unpleasant/painful" sound reproduction is one that is "unclear/indistinct", "thin", "dull", "sharp", "closed" and with much disturbing sounds. Which of these components are most detrimental may in fact vary depending on the type of stimuli. A complicating factor as regards sound reproduction is also that the stimulus is in fact a combination of an original auditory stimulus with certain physical characteristics and a reproduction system with certain other physical characteristics. There may thus be complex interactions on the physical side, which sometimes make the study of psychophysical relations very difficult.

It would indeed be desirable to be able to write the overall evaluation as a function of various physical parameters - like equation (3) but with physical variables instead of perceptual. This can be done, to a certain extent, in single experiments using the obtained data. However, considering the many complexities discussed above it should be clear that the validity of such equations can be very limited, and that much remains to be done before they can be stated at a more general level.

DISCUSSION

The emphasis here has been on various methodological approaches to the study of the relations depicted in Figure 1 with examples taken from sound reproduction. To what extent these approaches may be of value for the study of annoyance is an open question. However, we may venture some speculations.

For sound reproduction the key words are such as preference, pleasantness, fidelity and the like - not annoyance. However, if annoyance is regarded as a kind of opposite to preference, you may get some information of interest from data on the least preferred systems and from ratings in characteristics as "painful", "irritating" and "tiring", which are close to annoyance in meaning. The outstanding feature in this respect here is the relation between such negative evaluations and the "Sharpness/Hardness" dimension - which in its turn is mainly due to somehow increased sound levels at high frequencies (about 1 - 4 kHz).

In this connection it should be noted that loudness, which is certainly related to annoyance, is in principle neutralized in studies on sound reproduction - it is in fact very important to listen to various systems at the same loudness, since differences in loudness also affect ratings in other perceptual scales. There is evidence, however, that for many listeners *loudness may be more or less synonymous to "sharpness"*. This in its turn may be informative when discussing the relations between loudness and annoyance.

When dealing with sound reproduction most stimuli belong to speech and music and constitute wanted signals for the listener. The purpose is therefore to optimize the systems with regard to perceptual and evaluative aspects. It is necessary to study the relations between evaluative and perceptual aspects in some detail, since different systems for sound reproduction may all of them be evaluated as "good", "pleasant" etc. - but still sound very different among themselves. They are "good" in different ways - in other words, the same overall evaluation may result from very different constellations of perceptual factors. It seems probable that this applies to annoyance as well, i.e. the "same" annoyance may arise from many different combinations of perceptual constituents.

One could then imagine to do the same thing as has been described above but now with noise stimuli of various kinds. Given representative samples of such stimuli you could *use multivariate and other techniques to study the perceptual dimensions for noise stimuli and their relations to overall evaluations and physical/acoustical properties*. With this knowledge at hand you could try to *minimize the annoyance reaction as a function of various perceptual and/or physical factors*. You could think of an analogue to equation (3) with the annoyance reaction in the left member and solving for the appropriate constellations of members to the right to minimize the annoyance reaction. There are, of course, many problems in such an approach, e.g. how the results depend on the specific context of stimuli. Often the stimuli are varying over time, and it may be very informative to use some kind of *continuous response registration* to see which parts of the stimulus are most responsible for the annoyance reaction.

The discussion in this paper is limited to descriptions and illustrations of various methods for studying the perception and evaluation of various auditory stimuli. Noise stimuli may be considered as a special class of auditory stimuli, but the described methods are general enough to include noise stimuli as well. There are lots of factors influencing our reactions to noise stimuli: what type of activity you are involved in when the noise occurs, the time of day, the physical surroundings, the social surroundings, your earlier experiences of the noise stimuli and various expectations on the basis of such experiences, possible habituation to lasting or repeated stimuli, various motivational factors, your physical and psychological condition, individual differences in hearing and in sensitivity to noise, the society's views on noise problems etc. These questions are outside the scope of this paper, but generally seen it should be possible to study the effects of many of these factors by means of the above-mentioned methods or various modifications of them to suit the actual applications.

ACKNOWLEDGEMENTS

The author wants to express his gratitude to Bertil Johansson for valuable discussions and to Björn Lindström for continuous cooperation and much practical help.
The research have been supported by the Swedish National Board for Technical Development, the Swedish Council for Planning and Cooperation of Research, the Swedish Telecommunications Administration, the National Swedish Board for Consumer Policies and the Swedish Philips Group.

REFERENCES

Carroll, J.D. & Arabie, P. (1980). Multidimensional scaling. Ann. Rev. Psychol. 31, 607-649.

Carroll, J.D, & Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. Psychometrika 35, 283-319.

Cureton, E. E. & D'Agostino, R.B. (1981). Factor analysis: An applied approach. Lawrence Erlbaum Associates Ltd, London.

Gabrielsson, A. (1974). An empirical comparison between some models for multidimensional scaling. Scand.J.Psychol. 15, 73-80.

Gabrielsson, A. (1979). Dimension analyses of perceived sound quality of sound-reproducing systems. Scand.J.Psychol. 20, 159-169.

Gabrielsson, A., Frykholm, S.Å. & Lindström, B. (1979). Assessment of perceived sound quality in high fidelity sound-reproducing systems. Technical Audiology Reports, Karolinska Institutet, Stockholm, no. 93.

Gabrielsson, A., Frykholm, S.Å. & Sjögren, H. (1977). Adjective ratings and dimension analyses of perceived sound quality of headphones. Technical Audiology Reports, Karolinska Institutet, Stockholm, no. 86.

Gabrielsson, A. & Lindström, B. (1981). Scaling of perceptual dimensions in sound reproduction. Technical Audiology Reports, Karolinska Institutet, Stockholm, no. 102.

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.

Gabrielsson, A. & Sjögren. H. (1979a). Perceived sound quality of sound- reproducing systems. J. Acoust. Soc. Amer. 65, 1019-1033.

Gabrielsson, A. Sjögren, H. (1979b). Perceived sound quality of hearing aids. Scand. Audiol. 8, 159-169.

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

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

Kruskal, J.B. & Wish, M. (1978). Multidimensional scaling. Sage, Beverly Hills.

Namba, S. & Kuwano, S. (1980). The relation between overall noisiness and instantaneous judgment of noise and the effect of background noise level on noisiness. J. Acoust. Soc. Japan (E) 1, 99-106.

Shepard, R.N., Romney, A.K. & Nerlove, S. (Eds.) (1972). Multidimensional scaling: Theory and applications in the behavioral sciences, Vol. 1. Seminar Press, New York.

Torgerson, W.S. (1958). Theory and methods of scaling. Wiley, New York.