Karolinska Institutet
Technical Audiology
KTH, 100 44 Stockholm 70

Report TA No. 82
March 1976

PREFERRED LISTENING LEVELS AND PERCEIVED SOUND QUALITY AT DIFFERENT SOUND LEVELS IN "HIGH FIDELITY" SOUND REPRODUCTION

Alf Gabrielsson and Håkan Sjögren

ABSTRACT

Ten "hi-fi" subjects and ten "non hi-fi" subjects made adjustments of preferred listening level for six music sections and a speech section reproduced by four loudspeakers. The results showed a large inter- individual variation in preferred level, and the adjustments were affected by characteristics of the loudspeakers, the programs, and of various psychological factors. The preferred level for most subjects corresponded approximately to the original level of the recordings. In a following experiment the same subjects rated the "true-to-nature" quality of the reproductions when they were reproduced at the preferred level and at another level 10 dB lower or higher. The effects of the different sound levels on the ratings varied with subjects and often interacted with characteristics of the loudspeakers and the programs. Suggestions are given for the selection of sound levels in "listening tests".

INTRODUCTION

Investigations on "preferred listening level" or "most comfortable listening level" are of great concern to many branches in sound engineering, for instance, broadcasting (Chinn & Eisenberg, 1945, 1947; Ilmonen, 1971), telephony (Gleiss, 1974), audiology (Kopra & Blosser, 1968; Ventry et al., 1971; Gabrielsson et al., 1974), and high-fidelity sound reproduction (for some short comments see Olson, 1967, p. 399). Other aspects of "comfortable listening levels" are treated by Pollack (1952) and Backman (1971). Several other reports may be found by consulting the references given in the above-mentioned papers.

There are two inter-related purposes with the present work. First, to get some information about preferred listening level for music and speech when reproduced by some different systems for high-fidelity sound reproduction in a listening room similar to an "average living room". Second, to study how the perceived sound quality of the respective sound-reproducing systems is affected by differences in the sound level of the presentation. This latter point is of direct relevance for the design of so-called "listening test" as explained below.

In most earlier research the preferred listening level or "most comfortable level" (abbreviated to MCL in the following) was investigated for sinusoidals, bands of noise, or speech sections as stimuli, while there are only meagre data as regards music. A general impression from the earlier work is that there is a considerable variation between individuals (inter-individual variation) in their MCL, while the variation within the single individual (intra-individual variation) may be much smaller. In the present context it may also be expected that the MCL:s will be a function of many different parameters related to the stimuli and listening conditions (type of music and its original sound level, characteristics of the sound-reproducing systems, acoustic characteristics of the listening room etc.) as well as to the listening subjects (their acquaintance and liking/disliking of the music, their earlier experi-

ences of sound-reproducing systems, temporary physical condition, temporary mood, motivational factors etc.). To investigate systematically the influence of all such factors on MCL would require a large research project. The present work is explorative as an attempt to find out some critical factors influencing the MCL and the way they seem to operate. This information might form the basis for continued work.

Investigations on MCL may be of importance for the design of "listening tests", in which a number of subjects makes judgments about the perceived sound quality of, say, some various loudspeakers. It is well-known that such judgments may be highly dependent on which sound level is used for the presentation of the selected program sections (music, speech) in the test. That is, the rank order and/or the perceptual differences between the loudspeakers may be different for different sound levels. One alternative could be to use a "true-to-nature" level, that is, a sound level corresponding to the original sound level at the recording of the music or speech in question. However, in many cases (especially for music in fortissimo) such a level is probably not used when listening at home since it would be too disturbing for other people around. Another alternative would then be to use a preferred listening level (that is, MCL) which has to be established for each subject before the listening test is started. There is sometimes an (implicit) expectation that it would be possible to find an average MCL which could be representative for a certain kind of listeners or for an "average listener". However, in view of the large inter-individual variation in MCL found for other stimuli and in view of the many factors which may be supposed to affect the MCL, it is doubtful whether this is a reasonable expectation.

To try to answer the questions raised above the investigation was performed in two parts. In the first part 20 subjects, representing a "hi-fi" group and a "non hi-fi" group, made adjustments of MCL for a number of music sections and a speech section as presented over four different loudspeakers. The data from this part were used in the design of the second part, in

which the same subjects made judgments concerning how "true--to-nature" the respective loudspeakers reproduced the music/ speech sections, when they were presented at two different sound levels: at the MCL as obtained in the first part of the experiment and at another sound level 10 dB lower or 10 dB higher. The detailed planning of the second part was not made beforehand but was made dependent on the results concerning MCL in the first part.

METHOD

Stimuli and listening conditions

There were seven program sections, six music sections and one speech section. The music sections were chosen to represent different kinds of music and music ensembles as well as different sound levels, ranging musically from piano to fortissimo. The sections were:

Program 1 (P1): Excerpt from the beginning of "Prélude à l'après-midi d'un faune" (Debussy), performed by the Stockholm Philharmonic Orchestra. Recorded in the Concert Hall of Stockholm using two omnidirectional microphones for the main sound image, two for the wood winds, and two at the rear end of the stalls for the reverberation. Sound level: 70-80 dB(A). Gramophone record: LYSSNA 6 (available from the Concert Hall, Stockholm).

Program 2 (P2): The very end of the finale chorale in "St. John Passion" (J.S. Bach), performed by the Bach Choir at Adolf Fredrik, Stockholm. Recorded in the empty church of Adolf Fredrik, Stockholm, using two omnidirectional microphones. Sound level: 85-97 dB(A). Gramophone record: PROPRIUS 7741.

Program 3 (P3): Excerpt from the end of "The Firebird Suite" (Stravinsky), performed by the Stockholm Philharmonic Orchestra. Recorded in the Concert Hall of Stockholm, using two omnidirectional microphones. Sound level: 90-97 dB(A). Special recording for the Cullberg ballet.

Program 4 (P4): Big band jazz music, performed by "Symfonikernas jazzband". Recorded in Nacka Concert Hall, using two omnidi-

rectional microphones. Sound level: 90-95 dB(A). Available from Tandberg Radio, Norway.

Program 5 (P5): "You have got a hot line to my heart", performed by female singer accompanied by piano, bass and drums. Originally four-channel recording made at the Swedish Broadcasting Corporation, transformed to two-channel stereo. Sound level: 80-88 dB(A). Gramophone record: LJUD (The Swedish Hi-Fi Institute).

Program 6 (P6):Speech section read by the experimenter (C.W., see Acknowledgements). Free field recording. Sound level: 65-70 dB(A).

Program 7 (P7):"Gammal fäbodpsalm", Swedish folktune performed by soprano voice and mixed choir. Recorded as program 2. Sound level: 70-85 dB(A). Gramophone record: PROPRIUS 7731.

The program sections lasted for about 30 seconds each and represented "musically homogeneous" sections. The stimuli tapes were copied directly from the respective master tapes; (tape copies of the program sections are available at self-cost). The signal levels on the stimuli tapes were chosen to give highest possible signal-to-noise ratio without introducing audible distortion.

In this experiment the stereo recordings were reproduced monophonic (tape speed 38 cm/sec, overall dynamic range 70 dB, A-weighted) over the following four loudspeakers:

Loudspeaker A : Electrodynamic, omnidirectional loudspeaker

Loudspeaker B : Electrostatic loudspeaker

Loudspeaker C : Electrodynamic, loudspeaker elements mounted in the front of the cabinet

Loudspeaker D : Radio receiver, loudspeaker turned upwards.

Loudspeakers A - C belong to the "hi-fi" category but are rather different as seen in Figures 1 - 3. The frequency curves for A and C are comparatively flat, while the frequency response of B has peaks in the bass region and falls towards higher

frequencies. B also has much distortion in the bass range.

Loudspeaker D of "non hi-fi" quality was included for comparison. As seen in Figure 4 its bandwidth is rather small (about 200-6000 Hz) and there are several marked peaks and dips within this range.

The acoustical level of the program sections was set to correspond approximately (within ± 3 dB) to the original level at the respective recording (given under Programs above) as measured over loudspeaker A. The level of the different loudspeakers was approximately equalized using a white noise signal measured at the octaves 500 and 1000 Hz and broad-band dB(A), see Table I. The equalization was also perceptually checked to give approximately the same loudness impression from all loudspeakers when they were listened to in rapid succession within each program section.

In the first part of the experiment the subjects themselves adjusted the listening level (see Procedure) by using a gain control of a preamplifier (CM Laboratory). The gain control was equipped with so-called loudness compensation, see Figure 5. The loudness compensation was used only for levels below the approximate original level.

The listening room is described in detail in an earlier report (Gabrielsson et al., 1971). The room is arranged to imitate a "normal" living-room, both visually and acoustically. There are armchairs for the listeners along one of the long sides, and the loudspeakers are placed along the other longside (loudspeakers A and B on the floor, C and D on a shelf about 1.5 meter above the floor), hidden by an acoustically transparent curtain. The room has about "normal" diffusity and the reverberation time sinks from 0.90 sec at 63 Hz to 0.34 sec at 6300 Hz.

Subjects

There were two groups of subjects, one "hi-fi" group and one "non hi-fi" group, ten members in each group.

The members in the "hi-fi" group were recruited from a club for people interested in sound engineering and high fidelity sound reproduction. They all had modern listening equipment of high quality at home. Only one of them performed music (subject 1; the numbering of the subjects appearing in the following refers to the numbering of the subjects given in the appendices). Their listening to "live" music (visits to concerts etc) varied from "very seldom" (subjects 2, 5, 7 and 9) to one-two times per month (most of them) or even once per week (subject 10). All of them were men, 20 - 37 years old.

The members of the "non hi-fi" group were recruited by means of advertisments put up at the Royal Academy of Music and in the Concert Hall of Stockholm. As conditions for participation it was requested that these subjects should often listen to "live" music but not be accustomed to listen to high-fidelity sound reproduction. (Parenthetically said,much more work was required to get hold of such subjects than of the "hi-fi" subjects.) They typically had smaller or old radio receivers, sometimes tape recorders and loudspeakers of lower quality. Seven of them (subjects 11, 12, 14, 15, 17, 18 and 19) performed music themselves since several years ago and all of them were frequent concert visitors (as a rule once a week or more). There were eight males and two females. Nine of them were 16 - 25 years old, the remaining one 68 years old (subject 20, frequent concert visitor since very long time ago).

All subjects were checked for normal hearing, less than 15 dB hearing loss 250-4000 Hz (ISO), and were paid for their participation.

Procedure

The experiment was divided into two parts. In the first part the subjects made adjustments of the MCL for the different program sections presented over the four loudspeakers. In the second part they made ratings of how "true-to-nature" the respective reproductions sounded at two different sound levels. There was an interval of 1 1/2-2 months between the two parts.

a) MCL adjustments

This part of the experiment was made individually for each subject. The subject was sitting in an armchair and beside him there was a small table on which the pre-amplifier had been placed. The pre-amplifier was, however, covered by a cardboard box and thus totally invisible. The gain control of the amplifier was equipped with a lengthened axis which protruded through a small hole in the box and ended with a steering wheel, which was operated in principally the same way as an ordinary gain control. However, the extended axis was arranged with a special device to make it slide over the end positions of the real gain control if the end positions were approached when the subject operated the steering wheel. This meant that, although the steering wheel actually affected the gain control only during some few rotations of the wheel, it could be turned infinitely in any direction and so it was impossible for the subject to get any indication of the end positions of the real gain control.

The following (here somewhat abbreviated) instruction was given, tape-recorded and also in written form:

"In this experiment you will listen to various sections with music or speech which are reproduced by different constructions for sound reproduction. Each single presentation lasts for about 30 seconds, and during that time you shall quite simply adjust the loudness so that the loudness is comfortable. (Note: the word "comfortable" is used here as a very approximate translation of the Swedish word "lagom", which actually means "something that is neither too much, nor too little"). The adjustment is made with the steering wheel beside you. Turning to the right increases the loudness and turning to the left decreases the loudness. While the music/speech is sounding you keep adjusting until you get a loudness which is comfortable for the respective case - it should be neither too loud nor too soft but a comfortable loudness as you would like to have it when you listen to the music/speech in this room. You need not bother about neighbours or other people (as you perhaps have to

do at home) but feel quite free to adjust the loudness as you want it to be comfortable just now in this room, with these recordings and in this listening situation! ... When you are ready with the adjustment, take your hand away from the wheel and do not touch the wheel during the intervals between the presentations. If you anyhow do not get satisfied with the adjustment, call for the experimenter and explain to him." Some practical instructions were added and thereafter fourteen preliminary trials were made (two presentations for each of the seven program sections).

There were 7 program sections x 4 loudspeakers = 28 cases. Each of these 28 cases appeared three times so in all there were 28 x 3 = 84 adjustments to be made. The order of these was randomized independently for each subject. The 84 adjustments were distributed over two sessions of about 80 minutes each. The subjects were also asked to write some free "verbal descriptions" as comments to their adjustments. This task was voluntary and it was stressed that it should not interfere with the adjustments.

To avoid any indication from the experimenter's side concerning comfortable loudness the gain control was operated solely by the subject himself (by his steering wheel) from the very beginning of the preliminary trials and all through the experiment. This means that the starting point for any new adjustment was the position that had been adjusted for the nearest preceding case.

The subject's setting of the gain control was measured using a sinusoidal signal passing the gain control in a parallell channel and a digital voltmeter read by the experimenter.

When the gain control was in its upper end position, the produced sound level was approximately the same as at the original recording. If this was not enough for the subject, he had to call for the experimenter (see last sentence of the instruction), who then raised the level still more according to the subject's wishes (linear rise, no loudness compensation). This

possibility was rather frequently used by subjects 1-3 in the hi-fi group and subject 11 in the non hi-fi group and in some few cases by other subjects.

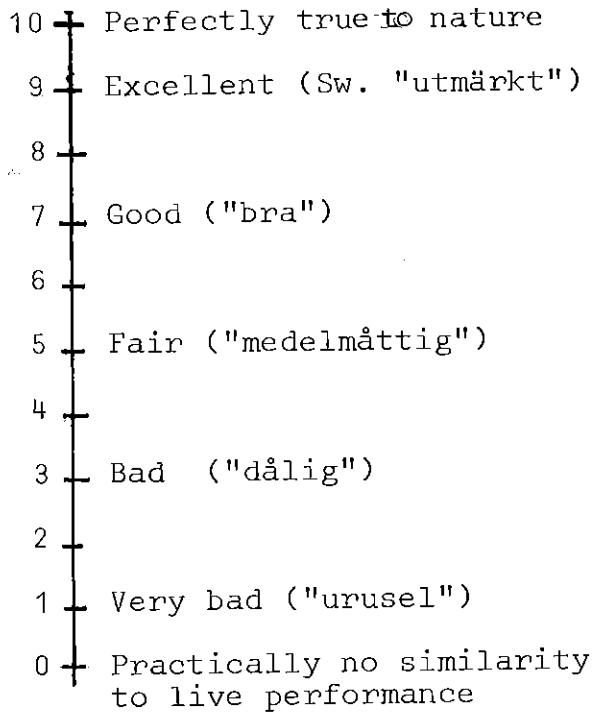
b) "True-to-nature" ratings

In this part of the experiment the subjects made "true-to-nature" ratings of the perceived sound quality. The sound level for each case was varied in two levels: one level corresponding to the MCL adjustment made by the subject for the respective case and another level 10 dB lower (if the MCL adjustment corresponded to the original sound level within ± 10 dB) or 10 dB higher (if the MCL adjustment corresponded to a sound level more than 10 dB lower than the original sound level, see further explanation under Results).

Eight subjects were run individually since their MCL adjustments were not similar to those of any other subject. The remaining twelve subjects could be run in six pairs, the MCL adjustments of one member of the pair being reasonably similar to those of the other member of the same pair (see further under Results).

The pre-amplifier was now removed from the listening room and operated by the experimenter. The subject was supplied with a short description of the program sections (name of music, composer, performers, and the room where the recording took place), a numbered paper for writing his ratings, and the following (here somewhat abbreviated) instruction:

"You will listen to the same music/speech sections as earlier and the sound reproduction of them will be varied in different ways. This time you have no wheel to operate but you shall concentrate on listening carefully with regard to how true-to-nature the sound reproduction is. The task is thus to judge for each presentation how true-to-nature the originally sounding music (speech) is reproduced. The judgment shall be made on a 0-10 scale as follows:



10 denotes a reproduction which is perfectly true to nature, it sounds exactly as if you listen to the originally sounding music (speech). You cannot hear any difference between such a reproduction and the original sound. The meaning of the numbers 9, 7, 5, 3, and 1 are seen in the figure. 0, finally, denotes practically no similarity at all to the original sound, a still worse reproduction cannot be imagined.

The fact that certain numbers are given definitions does not mean that they should be used more than the others. You may use any number from 0 to 10 which you think is the best to describe how true-to-nature the reproduction sounds. If you want, it is permitted to use one decimal in addition to the integer.

To judge how true-to-nature the reproductions sound, actually presupposes that you have heard or now get the possibility of hearing the music sounding in original. Since this is hard to arrange you have to trust your experience and imagination to imagine how the music originally sounded ... The speech section is read by the experimenter, whose voice you have heard directly... Observe finally that your judgments shall not be influenced of what you think of the music as such but only refer to how true-to-nature it is reproduced."

After some further practical information 15 preliminary trials were made for practice. The main experiment comprised 120 judgments (distributed over two sessions of about 1 1/2 hour each), viz. 5 program sections x 4 loudspeakers x 2 levels = 40 cases, each of which was judged three times. The order of the 120 presentations was randomized independently for each subject (pair of subjects, respectively). The number of programs was reduced

to five programs for each subject as explained under Results. In the last third of the experiment the subjects were asked to write free "verbal descriptions" as comments/explanations to their ratings (this task was voluntary), and finally they answered some questions related to both parts of the experiment.

RESULTS

a) MCL adjustments

Each subject made three adjustments for each of 28 cases (7 programs x 4 loudspeakers). The first step was to compute the arithmetic mean of the three adjustments for each case. It was apparent, however, that the arithmetic mean might be non-representative in certain cases, for instance, if two of the three adjustments were near each other while the third value was more or less deviating, or if there was a considerable dispersion between the three adjusted values. Therefore the median was computed, too, and accepted as the more representative measure of central tendency in the cases where the absolute difference between the arithmetic mean and the median was bigger than 1.5 dB. For most of the subjects this occurred in rather few cases. It was, however, rather frequently occurring for subject 7 in the hi-fi group and for subjects 16, 18 and 19 of the non hi-fi group.

The resulting values are shown in Appendix A (values denoted with crosses (x) are median values). The values are given in relation to the original sound level set to 0 dB. Thus values preceded by a plus sign mean that the adjusted level was so many decibels above the original level, and values preceded by a minus sign that the adjusted level was so many decibels below the original level. The arithmetic mean for the respective group as a whole and the corresponding standard deviation appear at the bottom of each column.

Inter-individual and intra-individual variation

It is immediately seen that the inter-individual variation is very large (and so the group means are rather useless information). For the hi-fi group there is a variation from about 5 - 8 dB above the original level (subject 1) down to about 20-25 dB below the original level (subject 10). For the non hi-fi group the corresponding variation goes from about 4-6 dB above (subject 11) to about 24-34 dB below the original level (subjects 19 and 20;the numbering of the subjects within each group was actually done to reflect the differences in adjusted levels, from high to low). The standard deviations over the subjects (see at the bottom of each column) are for the hi-fi group 7 - 10 dB, lowest (about 7 dB) for the speech program (program 6). For the non hi-fi subjects these standard deviations are somewhat higher, 8 - 12 dB in most cases, lowest for program 3. Although the inter-individual variation thus is large for both groups, the standard deviations given here are rather smaller than those reported in earlier works for inter-individual variation as regards pure tones, noise and speech (compare Ventry et al., 1971; Backman, 1971).

The intra-individual variation was considerably smaller. It is computed here as the average standard deviation of the three adjustments made per case (that is, the average of the standard deviation within each of the 28 cases). For the members of the hi-fi group this standard deviation was in general 2 - 4 dB (ranging from 2.2 dB for subjects 1 and 2 to 3.8 dB for subject 8). However, subject 7 showed a higher variability, his average standard deviation being 5.3 dB. In the non hi-fi group subject 14 was extremely stable with an average standard deviation of only 0.8 dB. Subject 17 had an average standard deviation of 2.5 dB, subjects 11, 12, 13, 15 and 20 had 3.4 - 4.0 dB and subjects 16, 18 and 19 were highly variable with an average standard deviation of 6.2 - 7.5 dB. For the three last-mentioned subjects it was fairly obvious that their higher variability was partly due to the circumstance that they successively increased the adjusted levels during the course of the experiment

(it is thus possible that their variability would have diminished if more adjustments were included). On the other hand, it is possible that "ceiling effects" are a contributing cause to the rather low intra-individual standard deviations for some of the subjects who adjusted the highest MCL values (that is, they approached an upper limit of the possible MCL and so their variability may be smaller).

MCL adjustments for loudspeakers and program

The group mean MCL adjustments for the four loudspeakers and the seven programs are given in Table II. As regards the loudspeakers, it is noted that loudspeaker A gets the highest adjusted level and loudspeaker B the lowest. This holds for both groups, but the levels for the hi-fi group are higher than those for the non hi-fi group (the lowest adjusted level in the hi-fi group, -8.0 for B, is about the same as the highest adjusted level in the non hi-fi group, -7.7 for A). The differences between the adjusted levels for the loudspeakers are bigger in the hi-fi group than in the non hi-fi group, and it is noted that in the non hi-fi group the adjusted level for the radio receiver (loudspeaker D) is about the same as for loudspeaker A.

As regards the programs, the three programs with the lowest original levels (program 1, 6 and 7) are adjusted at higher levels (relative to the original level) than the other programs in both groups. Here, too, the adjusted levels lie higher for the hi-fi group than for the non hi-fi group. The range of variation in adjusted levels for the programs is about the same in both groups; they differ, however, in many details (note, for instance, the low adjusted levels for the jazz programs 4 och 5 in the non hi-fi group).

Inference statistics in the form of variance analysis and F tests was performed for group data (randomized block factorial design, non-additive model) as well as on the data for each individual subject (treated as randomized factorial design with

loudspeakers and programs as independent variables; Kirk, 1968; Winer, 1971; Hays, 1973). The last-mentioned analyses showed that there were significant differences between the adjusted levels for the loudspeakers for almost all members of the hi-fi group but only for three members of the non hi-fi group. There were significant differences between the adjusted levels for the programs for almost all subjects. For four subjects (subjects 3, 4, 6 and 17) there was a significant interaction between programs and loudspeakers, the details of which may be studied in the individual data given in Appendix A. The common meaning of the interaction for these four subjects was essentially that the adjusted level for loudspeaker B and/or D was set much lower at certain of the programs (especially at program 3, 4 and 5) than at the other programs.

From the "verbal descriptions" and from answers to questions about which principles were used for the MCL adjustments, it is quite clear that the lower levels set for loudspeaker B is due to its emphasis on the bass region and distortion in the bass (Figure 2). To avoid the resulting negative effects from this ("too dark", "rumbling", sometimes "cracked") most subjects preferred to use a lower level for this loudspeaker. In an analogous way the hi-fi subjects and some non hi-fi subjects decreased the level for loudspeaker D (the radio receiver) since it sounded "distorted", "cracked", "hollow", "mushy", "sharp" etc at higher levels (due to overload). The reason why the two music programs in piano (program 1 and 7) was set at a relatively higher level by most subjects was of the type "I wanted to hear the flute solo clearly" (program 1) or to get the very soft beginning of the choir in program 7 to be heard clearer. Evidently, however, this raising led to a conflict in many subjects, because the noise from the tape also got louder, which was not wanted. It seems that complaints about the noise was more frequent in non hi-fi subjects than in hi-fi subjects (the latter knowing that a certain amount of noise is unavoidable).

Some subjects in the non hi-fi group explicitly pointed out that disliking of certain music sections may lead to that they set the level lower. The most obvious example occurred for subject 20, who disliked programs 4 and 5 and therefore set them 30 dB or more below the original level, see Appendix A (this fact is a contributing cause to the low levels for these programs in the means for the non hi-fi group in Table II). Statements concerning liking/disliking and about the influence on the settings due to the temporary mood from time to time was made by subjects 13, 16, 18, 19 and 20, some of which had high intra-individual variation. An important factor for MCL is, of course, if you intend to listen to the music (as the subjects had to do here) or if the music only serves as a background to other work (as pointed out by subject 12).

MCL in relation to original level

An interesting question for the present purposes is whether the adjusted MCL levels conform to a "true-to-nature" level, that is, to the approximate original sound level, or not. The instruction was to set a "comfortable" ("lagom") level, and nothing was said about a "true-to-nature" level or something like that. It might be expected that hi-fi subjects, according to their interest for high fidelity in sound reproduction, would equalize a "comfortable" level with a "true-to-nature" level more than non hi-fi subjects would do. And, as seen in Table II, the hi-fi group in general lies nearer to the original sound level than the non hi-fi group. However, there is in both groups a considerable inter-individual variation behind these group data, see Appendix A.

Since the original sound level as set by the experimenters has a tolerance of, say ± 3 dB, and since there is a tolerance (intra-individual variation) of some decibels in the subjects' MCL settings, it is reasonable to consider MCL settings in Appendix A ranging from about +5 dB down to about -6 or -7 dB as corresponding to an approximate "true-to-nature" level.

In the hi-fi group the MCL settings of subjects 2 - 6 for the most part lie within this range. The settings of subject 1 are sometimes higher than so, while those of subjects 7 - 10 lie still more lower down to about 20-25 dB lower for subject 10. In the non hi-fi group subjects 11 - 15 for the most part lie within the "true-to-nature" range and, to a certain part, subjects 16 and 17. Subjects 18 - 20 are marked "low-setters" down to 25 - 35 dB lower than the approximate original level (as noted earlier, however, subjects 18 and 19 tended to increase the MCL levels during the course of the experiment).

After the very end of the experiment (that is, after the "true--to-nature" ratings in the part to follow) the subjects were asked if they had deliberately tried to set an MCL so that it would correspond to a "true-to-nature" level, or if they had attempted to set an MCL so that it would sound louder or softer than in reality, or if they had not at all thought in terms of a "true-to-nature" level. In the hi-fi group subjects 3 - 6 answered that they had set the MCL to an approximate "true-to--nature" level, except for the cases where the reproduction sounded too "rumbling" or "distorted" at that level (referring to loudspeakers B and D). The two "high-setters" in this group, subjects 1 - 2, answered somewhat differently: subject 1 wanted a "loudness which 'filled' the whole room", and subject 2 aimed at a "revealing level" (probably meaning a level that would reveal defects in the reproduction). Subject 7 used a combined criterion of "pleasant loudness which also would be at least approximately true-to-nature". The "low-setters", subjects 8 - 10, mainly referred to a "pleasant/comfortable" loudness or to mixed principles (subject 9).

While these statements from the hi-fi subjects seem rather consistent in relation to the actual settings of the subjects, the picture is not so clear for the non hi-fi subjects. Among the subjects who made their settings at an approximate original sound level (subjects 11 - 15), only one (subject 14) directly referred to a "true-to-nature" level. Subject 11 referred to "a pleasant level and at which the single instruments were

clearly heard". Subject 12 indicated that a higher level actually was needed in many cases but could not be attained due to too loud noise. Subjects 13 and 15 referred to "pleasant loudness", subject 13 also to the type of music and the temporary mood. Subjects 16 - 17, who had MCL settings sometimes within and sometimes below the"true-to-nature" range, both referred to "clearly heard instruments" but "simultaneously as little noise as possible" (subject 16). Among the "low-setters" subject 18 referred both to "pleasantness" and original level: "the loudness should not be so high as to cause irritation but should also correspond to a 'live'-impression". The two most marked "low--setters", subjects 19 and 20, referred both to a "true-to-nature" criterion, for example, "as it would sound from the front seats in the concert hall" (subject 20, excepting programs 4 - 5 as mentioned earlier).

Dynamic range in original and in MCL adjustments

The dynamic range of the original levels for the different programs is about 25 dB (from program 6 at 65-70 dB(A) to 90-97 dB(A) for programs 2-4). Is this range maintained in the MCL settings? As seen in Table II, and also in the group means in Appendix A, there is a certain restriction of the dynamic range in both groups, since the three programs with the lowest original level (programs 1, 6 and 7) actually are set some decibels higher (in relation to the original level) than programs 2-4. For instance, in Table II the mean settings for the hi-fi group are -3.4 to -3.9 dB for programs 1, 6 and 7, while they are from -7.5 to -9.3 dB for programs 2-4. The same tendency also holds for most of the single subjects as seen in Appendix A (for loudspeaker A) or in Table III. The original dynamic range between programs 2-4 (treated as a unit since they have very nearly the same level) and program 6, 1 and 2, respectively, is given in the upper row of Table III (computed from the mean dB(A) values given for the programs under Stimuli, for instance, 75 dB(A) for program 1, 91 dB(A) for program 2 etc.). In the rows below the corresponding values

are given for subjects representing different levels of their MCL settings. In comparison with the original difference there is in general a reduction in the individual difference data, sometimes slight (for instance, subject 1), sometimes rather big (for instance, subject 2). However, there is hardly any systematic relation between the degree of decreased dynamic range and the different subjects categories ("high setters", "true-to-nature setters" and "low-setters"). On the whole it seems that the dynamic range between the programs certainly is reduced to a certain extent but is about the same at highly different MCL adjustments.

b) True-to-nature ratings

Selection of levels and design

In the second part of the experiment the various reproductions should be rated for their "true-to-nature" quality with special regard to the influence of different sound levels (see Introduction). It was decided to use two levels 10 dB apart. For the subjects, whose MCL adjustments corresponded approximately to the original sound level (about half the subjects) or were somewhat higher or lower than so, one level was their MCL setting for the respective program x loudspeaker combination. The other level was 10 dB lower than the respective MCL setting. For the six "low-setters" (subjects 8 - 10 and 18 - 20) one level was likewise their MCL setting for the respective program x loudspeaker combination but the other level was instead 10 dB higher than the respective MCL setting (a further reduction with 10 dB would often lead to irrealistic low sound levels). A 10 dB increase was also used in those cases where the "true-to-nature" setters incidentally gave lower MCL adjustments, especially occurring at loudspeaker B and D as noted earlier. In summary, it may be stated that in cases where the MCL adjustment was higher than -10 dB (as given in Appendix A) the second level was decreased 10 dB. On the con-

trary, if the MCL adjustment was lower than -10 dB, the second level was increased 10 dB. It would thus be possible to see what would happen to the perceived "true-to-nature" quality, if a 10 dB reduction is made from a preferent level approximately coinciding with the original level, and what would happen if a 10 dB rise is made from a preferent level definitely below the original level.

In order to reduce the work for the subjects and the experimenter the following restrictions were introduced. Program 7 was taken away since the results for this program were very similar to those for program 1 (see Appendix A). Program 4 was rated only by half the hi-fi subjects and half the non hi-fi subjects, and program 5 by the remaining half of the two groups. Twelve subjects were combined to pairs (subject 2 and 11; 5 and 12; 8 and 9; 10 and 18; 14 and 15; 19 and 20) since the MCL adjustments of the two members in each of these pairs were reasonably similar. The MCL values for these pairs were taken as the means of the respective two individual MCL values, and the original individual adjustments were scrutinized to see that these means would fall within the intra-individual tolerance (in general within $\pm$ 1.5 x the intra- individual standard deviation). In the few cases where the mean would be non-representative to both members of the pair, the value was set to that of one of the members (this occurred in only nine cases out of 120 possible, all nine at different program x loudspeaker combinations, and so their effect for the data treatment could be neglected; these cases were, however, removed from the computations underlying Table 6). The remaining eight subjects were run individually.

Each subject made three ratings (on the 0 - 10 scale) for each of 40 cases (5 programs x 4 loudspeakers x 2 levels). The first step was to compute the arithmetic mean of the three ratings for each case. The resulting values are given in Appendix B as well as the group means. The two sound levels are simply designated as "high" and "low", respectively. In general the "high" level is equal to the preferred level for subjects 1 - 6 and 11 - 17, while the "low" level is equal to the preferred

level for subjects 7 - 10 and 18 - 20. Exceptions from these rules are marked by a cross (x) which is written in that column which denotes the preferred level in that case (for example, the cross written at the value 6.7 at program 1, loudspeaker A,"high" level, subject 7, means that in this case the "high" level was the preferred level).

Statistical treatment

Mean data for both groups at all combinations of loudspeakers, programs, and levels appear in Table IV. To facilitate the interpretation analyses of variance were performed on the underlying data. One analysis was made for each of the two groups, utilizing all data from all ten subjects in the respective group (called group ANOVA in the following). Further an analysis was made on all data for each single subject (individual ANOVA). The two group ANOVA:s were performed according to the model of a randomized block factorial design (mixed model, non-additive; Kirk, 1968) with loudspeakers, programs, levels, subjects and all possible interactions between these factors as variation sources. The F tests for the fixed factors (loudspeakers, programs, levels, and the corresponding interactions) were made using the mean square of the respective factor's interaction with the random factor (that is, subjects) as error term (for instance, the mean square for loudspeakers was tested against the mean square for loudspeaker x subjects, the mean square for the interaction loudspeaker x program against the mean square for the interaction loudspeaker x program x subjects, and so on). All terms including subjects were tested against the "within cell" error term (that is, replications). The individual ANOVA:s were performed according to a three factor (loudspeaker, program, level) randomized factorial design with the "within cell" mean square as error term for all F tests.

There was a host of significant F tests in the group ANOVA:s. In the hi-fi group all main factors and all two- and three-factor interactions were significant at the 1% level, except for the main effect of program and level which were significant at

the 5% level. In the non hi-fi group there were no significances for programs, levels, and the interaction loudspeaker x level. The remaining main effects and two- and three-factor interactions were significant at the 1% level.

Because of the many interactions (including interactions with the subjects) it is in a sense more interesting to look at the results from the individual ANOVA:s, Table V. This table presents which significances occurred for all subjects. Moreover the percentage variance accounted for by the different sources of variation are given as computed by the ω^2 index proposed by Hays (1973; Kirk, 1968, p. 198, generalized to a three factor design). The higher the percentage variance accounted for, the less the error variance. The percentage error variance is thus given by 100 minus the value given in column "Sum of variance accounted for". Finally the average intra-individual standard deviation (average of standard deviation for the ratings within each of 40 loudspeaker x program x level combinations) is given in the column to the right.

Combining the information given in Tables IV and V (and from various further statistical tests, the details of which are omitted here) the most important results are the following:

Loudspeakers:

There are obvious differences between the loudspeakers. Loudspeakers A and C get about the same ratings and are superior to B and D. The reasons for the inferior perceived quality of B and D were mentioned earlier (too much emphasis on the bass region and bass distortion in B, limitations in bandwidth and power handling capacity in D).

Programs:

There were significant differences between the mean ratings at the different programs (the column utmost to the right in Table IV) in the hi-fi group (but not in the non hi-fi group). The three fortissimo programs (programs 2 - 4) lie lower than the remaining programs as rated by the hi-fi group. This may be interpreted so that the fortissimo programs are more demanding for a "true-to-nature" reproduction than the other programs. However, there is an interaction between programs and loudspeakers. It is noted that the lower mean ratings for programs 2 - 4 are solely due to the worse reproductions of these programs by louspeakers B and D. Loudspeaker A and C, however, reproduce these programs about as well as the other programs. The interaction may also be described so that the range of the ratings is bigger for the fortissimo programs (especially program 4) than for the other programs (smallest range at program 1).

The last mentioned fact also fits for the meaning of the loudspeaker x program interaction in the non hi-fi group. The differences between the loudspeakers are in general more pronounced for the louder programs, while there are small differences for program 1.

Sound levels:

There are for the most part negligible differences between the ratings at the two different sound levels when the ratings are averaged over loudspeakers (see the margin to the right in Table IV) or programs (bottom margin). There is one sizeable difference for program 4 in the non hi-fi group, where the higher level gets higher ratings.

There are interactions, however, with loudspeakers and with programs in the hi-fi group and with programs in the non hi-fi group. In the hi-fi group the better loudspeakers A and C get higher ratings when reproducing the fortissimo programs 3 and 4 at the higher level than at the lower level, while the situation is quite opposite for loudspeaker B (and partly D). For program 2 there is no difference between levels

for loudspeaker A and C, while there is a big difference favoring the lower level for loudspeaker B (and a smaller difference in the same direction for D). For the speech program (program 6) there is a preference for the lower level in all loudspeakers, especially for A and C.

In the non hi-fi group the reproduction at the higher level is rated the more "true-to-nature" for loudspeakers A and C reproducing program 2 - 4, but - unlike the situation in the hi-fi group - this is also partly the case for loudspeakers B and D. For the speech program there is a tendency to prefer the lower level for all loudspeakers as in the hi-fi group.

It must be remembered that, although the difference between the two sound levels is always 10 dB, these levels vary from subject to subject, since they were taken from the individual MCL adjustments. It is thus possible that a more detailed analysis about level preferences should be made, dividing the subjects into further groups according to their MCL adjustments. Such an analysis is made in Table VI. The subjects are grouped into "high setters" (HS, subjects 1, 2 and 11, not at all programs, however),"true-to-nature setters" (TTNS, the majority of both groups), and "low setters" (LS, in general subjects 8 - 10 and 18 - 20). The number of subjects in each category is written in parenthesis after the respective symbol (this number sometimes varies between two values because some subjects were TTNS when listening to certain loudspeakers but LS when listening to other loudspeakers). For program 4 - 5 the data for the hi-fi and the non hi-fi group are combined to get enough data for means (only half the subjects in each group listened to either of these two programs).

Since the number of subjects in each category is sometimes rather small the data should be regarded with caution. As regards the few HS (at programs 1, 2 and 4), it seems that they (or only he at programs 2 and 4) actually rated the reproduction at the 10 dB lower level as more "true-to-nature" in most cases for programs 1 and 2; however, for program 4 the higher level is rated more "true-to-nature" for loudspeakers

A and C.

As regard the LS, it seems that the non hi-fi LS (who in fact were more extreme "low setters" than those in the hi-fi group) actually rated the reproduction at a 10 dB higher level as more "true-to-nature" in most cases. (This partly contradicts, of course, the statements made by these subjects that they aimed at a "true-to-nature level" when they made their MCL adjustments, see above.) The same phenomenon occurs some times for the hi-fi LS (program 3) and for LS of the both groups combined (program 4 and 5). However, the dominant impression for the hi-fi LS is that they rate the lower level (that is, their MCL) as giving a more "true-to-nature" reproduction in the case of the worse loudspeakers B and D, while there are small differences between the levels in the case of the better loudspeakers A and C. For the speech program the LS of both groups tend to regard the reproduction at their MCL as more "true-to-nature" than at a 10 dB higher level.

As regards the TTNS, it might be expected that they would rate the reproductions at the higher level (= their MCL) as the more "true-to-nature", or that there will be no sizable differences in the ratings between the two levels. This is also in general the case. The differences in the ratings between the levels are for the most part rather small. Sizable differences to the preference of the higher level occur in some cases (program 2 at loudspeaker D, non hi-fi; program 3 at loudspeaker C, hi-f; program 4 at loudspeakers A - C). Sizable differences to the preference of the lower level also occur in some cases for loudspeakers B or D (program 1 at D; program 2 at B, hi-fi; program 3 at B,non hi-fi), and for loudspeakers A and C in the speech program.

Although there were in general negligible differences between the ratings at the two levels when data are averaged over subjects, loudspeakers and programs (Table IV), there may in fact appear sizable differences if a more detailed analysis is done as shown here. Whether a sizable difference between the reproductions at the different levels shows up or not seems to depend on many

factors: the characteristics of the loudspeakers, of the programs, of the absolute level of the MCL adjustments, and on interactions between these factors. A quite general answer to the question which effect the difference in sound level has on the perceived "true-to-nature" quality is thus hardly possible to give.

Two further facts will be noted concerning the effects of different sound levels. First, the importance of differences in sound level relative to differences in other factors varies in different subjects, see Table V. The amount variance accounted for by the level factor is for most subjects small and lower than the variance accounted for by the loudspeakers and the programs. For some subjects in the non hi-fi group it is relatively important (subjects 12, 14 and 19). However, including also interactions with the level factor (see the last four sources of variation in Table V), it is seen that almost all subjects somehow take the level difference into account in its interaction with loudspeakers and/or programs. This is also reflected in the subjects' answers to a question concerning how the differences in sound level affected the "true-to-nature" judgments. Subjects 3 - 5, 9 and 14 tried to disregard differences in sound level (however, it did affect subject 14 as seen above). All other subjects stated that level differences influenced the ratings in various ways. For instance, some loudspeakers sound better at a lower level (referring to loudspeakers B and D) due to "clipping", "distortion", etc at higher levels; some of the programs must be reproduced at a high level to give a "true-to-nature" impression (referring especially to programs 2 - 4); too soft reproductions sound "indistinct" or "blurred", etc.

Second, the differences between the rating values for the different loudspeakers are in general more pronounced at a higher level (not unexpected, of course). This may be seen in Table IV and is also shown in the two columns to the right in Table VI, which present the range in ratings between the loudspeakers at "high" and "low" level, respectively. The range (the difference

between the highest and lowest rating) is almost invariably bigger at the higher level than at the lower. However, the speech program may be an exception from this.

Hi-fi subjects versus non hi-fi subjects

In addition to differences already mentioned the following characteristics may be noted. The mean of all ratings is higher for the non hi-fi group (5.7 than for the hi-fi group (5.0), (see the "total mean" down to the right in the respective parts of Table IV). This is mainly due to the fact that the worse loudspeakers B and D got higher ratings by the non hi-fi:s than by the hi-fi:s (Table IV). The reasons may be several. The non hi-fi subjects in general listened to a level somewhat lower than the hi-fi subjects (Table II) and so the negative effects of a high level for loudspeakers B and D were not so pronounced. The non hi-fi subjects are more accustomed to listen to radio receivers with limited bandwidth (as D here), often prefer a "darker" sound (as in B), and dislike noise which is more noticeable in loudspeakers like A and C (for all these points compare Gabrielsson et al., 1971, 1972, 1974). It follows that the differences in rating values for the loudspeakers are smaller in the non hi-fi group (Table IV), and that the variance accounted for by the loudspeakers is smaller for the non hi-fi:s than for the hi-fi:s (Table V).

The reliability of the ratings is for the most part higher in the hi-fi group: higher percentage variance accounted for (thus smaller error variance) and lower intra-individual standard deviations than in the non hi-fi group (Table V). On the whole a general (and rather natural) impression is that the hi-fi group is more homogenous in the ratings and in the judgment principles as stated in the verbal descriptions. Extreme ratings occur more often in the non hi-fi group (see, for instance, subject 12, Appendix B).

However, it cannot be emphasized too strongly that the above statements refer to characteristics of groups. When comparing single subjects from the respective groups the situation may

be quite different. There are hi-fi subjects who are less reliable in their ratings than non hi-fi subjects (Table V). It happens that hi-fi subjects rate the radio receiver as high as or higher than loudspeakers A or C, etc. Details may be studied by means of Tables III, V and VI and in the appendices.

Judgment principles

The verbal descriptions and answers to various questions reveal that there is a multitude of factors which are considered by the subjects when they do their ratings concering the "true--to-nature" quality of the reproductions. Some factors often mentioned are the following: presence or absence of bass and/or treble in the reproduction and balance between bass and treble; presence or absence of distortion, clipping and noise; clearness/distinctness in the reproduction (for instance, the possibility of distinguishing single instruments in the orchestra); the "impression of room" or "feeling of space" in the reproduction; the loudness impression as described earlier. It is also apparent that different subjects weight these factors in sometimes very different ways, which is probably an important reason to the fairly biginter-individual variation in the ratings of the same reproduction (Appendix B). The "true--to-nature" measure is composed by many different components, and there is an urgent need of research to discover which perceptual dimensions enter into perceived sound quality (Gabrielsson et al. 1971, 1972, 1973, 1974, 1975). It is possible that an investigation on the influence of sound level upon perceived sound quality would reveal less complex relations if some selected perceptual dimensions were used instead of an "overall" measure as here (for instance, how does the sound level affect the perceived distinctness of the reproduction? the balance between bass and treble? etc).

DISCUSSION

There was a twofold purpose with this investigation. First, to get some data about MCL for music and speech as presented by various sound reproduction systems of "high fidelity" type and which factors may affect the MCL in such situations. Second, to study how the perceived "true-to-nature" quality of the reproducing systems is influenced by different sound levels of the reproduction - with special regard to the question whether to use the original, "true-to-nature" level or some kind of MCL for the presentation of stimuli in a "listening test".

The results may be briefly summarized in the following way:

a) MCL adjustments

There was a wide inter-individual variation in the MCL adjustments in both groups (hi-fi group and non hi-fi group). The majority of subjects made MCL settings corresponding approximately to the original sound level. There were one or two "high setters" in each group with MCL settings over or at the upper limit of the original level. There were three or four "low setters" in each group with MCL settings more or less below the original level (down to about 30 dB below). The original dynamic range between the different programs was for the most part reduced by some decibels. However, the reduction was about the same for all three categories ("high setters", "true-to-nature setters", and "low setters"). The intra-individual variation was 2 - 4 dB (average standard deviation) for most subjects.

The MCL adjustments was influenced by characteristics of the loudspeakers (lower MCL:s for the "bass emphasizing" loudspeaker B and for the radio receiver, loudspeaker D) and of the programs (somewhat higher MCL:s, relative to the original level, for the "softer" programs 1, 6 and 7). Sometimes the MCL was lowered to avoid audible noise in the recording. The mean MCL was somewhat lower for the non hi-fi group than for the hi-fi group, mainly due to three marked "low-setters" in the non hi-fi group. Other factors said to influence the MCL adjust-

ments in various ways was liking or disliking of the music in question, "active" or "passive" listening to the music, temporary physical condition (tiredness etc.), and temporary swingings in the mood. About half the subjects stated that they set their MCL to correspond to the original level (even some of the "low setters" in the non hi-fi group). Other subjects referred to a "pleasant/comfortable" loudness or to various "mixed" principles.

b) "True-to-nature" ratings

The "true-to-nature" ratings for the reproductions were made at two sound levels: one the MCL, the other 10 dB lower (for "high setters" and "true-to-nature setters") or 10 dB higher (for "low setters"). There were obvious differences in the ratings of the loudspeakers (A and C rated higher than B and D), and the mean ratings at the fortissimo programs (programs 2 - 4) were lower than for the other programs due to bad reproductions of these programs by loudspeakers B and D (as rated by the hi-fi group). The non hi-fi subjects rated loudspeakers B and D higher than the hi-fi subjects and differentiated less between the loudspeakers than what the hi-fi subjects did. Considered as a group the hi-fi subjects were somewhat more reliable in their ratings (less error variance, smaller intra-individual standard deviations) than the non hi-fi subjects. However, characteristics of the groups do not always apply when comparing single individuals from the different groups.

The effect of the different sound levels on the ratings was in general small when the ratings were averaged over subjects and/or over loudspeakers and programs (Table IV). A few sizable effects were noted, for instance, that the hi-fi subjects rated loudspeaker B higher at the lower level for the fortissimo programs. A more detailed analysis on the data from "high setters", "true-to-nature setters", and "low setters" separately, revealed more instances of sizable rating differences between the two levels (Table VI), for instance, that a "high setter" often rated the 10 dB lower reproduction as more "true-to-nature",

and conversely that the most extreme "low setters" rated the reproduction at 10 dB above their MCL as more "true-to-nature". The data in Tables IV and VI indicated complex interactions between many factors which was verified in the individual analyses of variance (Table V).The importance of differences in sound levels was found to vary between subjects, which was also seen in the subjects' own statements in this question. A few subjects wholly disregarded the level difference.Most subjects, however, somehow took the level difference into account often in interaction with the loudspeakers and/or the programs (for instance, some loudspeakers sound more "true-to-nature" at a lower level, while others sound better at a higher level etc).

The conclusion is thus that the effect of differences in sound level on "true-to-nature" ratings cannot be stated in general terms. It will depend on characteristics of loudspeakers and programs and varies with subjects in a complex (but rather understandable) way.

c) Implications for "listening tests"

Both phenomena studied here - MCL and perceived "true-to-nature" quality at different sound levels - are apparently dependent on a multitude of factors. These may be systematically studied in continued research to get a more detailed understanding of the relevant relations and interactions. However, on the basis of the present results it seems possible to draw certain conclusions as regards the sound levels for stimulus presentation in "listening tests".

It has been discussed (see Introduction) whether to use the original level ("true-to-nature" level) or a preferred level (MCL), or both,or levels chosen according to other criteria. The following results seem relevant for this discussion:

(1). There is a wide inter-individual variation in MCL, and MCL varies as a function of many technical/acoustical and psychological variables. It is thus hard to find an average MCL value that would be representative for many different

listeners and for varying stimuli and listening conditions. And even if MCL:s are established for individuals and stimulus situations before the test, this requires a lot of work and probably results in a multitude of levels to be handled in the test, which may make it harder to draw conclusions. With new subjects and new stimuli, new MCL:s must be established.

(2). In the present experiment the MCL for the majority of subjects actually coincided approximately with the original level. There were, however, certain "low setters" in each group and one or two "high setters" at certain of the programs.

(3). There was in general a bigger differentiation in the ratings of the loudspeakers when the higher level was used, for "true-to-nature setters" as well as for "high setters" and "low setters" (Table VI). The speech section may be an exception from this. A higher sound level will be more demanding for the reproduction properties of the loudspeakers. (Consequently at least one program section in a "listening test" should represent music in fortissimo.)

(4). The fact that most subjects somehow took the difference in sound level into account when doing the "true-to-nature" ratings suggests the desirability of including more than one level in a "listening test".

In the authors' opinion these results point to the conclusion that "listening tests" should be performed using the original level of the selected program sections and a lower level, say, 10 - 15 dB lower. There are rather self-evident reasons for including the original level - it is definitely in line with the principle of "high fidelity" in sound reproduction. There is no need to include levels higher than the original level. The original level will also correspond approximately to the MCL of many subjects, at least when they need not bother about neighbours or other people around. Including a 10-15 dB lower level will also probably correspond to the MCL of many "low-setters", and will present a reduction in sound level of a size that may be necessary at home listening for subjects who actually want to listen at the original level.

If only one level could be included in the test, it seems more reasonable to use the original level than a lower level. Reproduction at the original level will in general give more information about the reproducing properties of the loudspeakers than at the lower level. However, for a user who knows that he will never listen at the original level (to avoid disturbing people or due to defects in other equipment, etc) this may be partly irrelevant or "redundant" information.

It is noted that, if these suggestions are accepted, the actual original level of every program must be known. This is important to consider when selecting the programs for use in "listening tests".

It is finally pointed out that the above conclusions are based on results from a single experiment with a limited number of stimuli and subjects. Replications with other stimuli, other subjects etc are highly wanted. And, as suggested earlier, it would be more satisfying if"listening tests" included judgments in a number of dimensions relevant for perceived sound quality rather than only an "overall" judgment of the "true-to-nature" type.

ACKNOWLEDGEMENTS

The authors wish to express their gratitude to Cenneth Wallén, who made an heroic work as experimenter and took part in many discussions, to Bodil Johansson for much help with computations, and to "Statens Provningsanstalt" for providing the listening room and technical assistance. The investigation was supported by a grant from the Swedish Council for Social Science Research and from the Swedish Board for Technical Development.

REFERENCES

BACKMAN, J. (1971). Preferensbedömningar av ljudintensiteter Beteendevetenskapliga institutionen, Umeå Universitet, Sweden.

CHINN, H.A. & EISENBERG, P. (1945). Tonal range and sound-intensity preferences of broadcast listeners. Proc.IRE, 33, 571-581.

CHINN, H.A. & EISENBERG, P. (1947). New C.B.S. program transmission standards. Proc. IRE, 35, 1547-1555.

GABRIELSSON, A., JOHANSSON, Bertil, JOHANSSON, Bodil, LINDBLAD, A-C. & PERSSON, L. (1974). Assessment of comfort and discomfort levels for pure tone, a methodological study. Reports from Technical Audiology, Karolinska Institutet, Stockholm, Sweden, No.74.

GABRIELSSON, A., ROSENBERG, U. & SJÖGREN, H. (1971, 1972). Judgments and dimension analyses of perceived sound quality of sound-reproducing systems. Reports Psychol. Lab., University of Uppsala, No. 115 and 116. (Also as Reports from Technical Audiology, Karolinska Institutet, Stockholm, No. 66 and 68.)

GABRIELSSON, A., ROSENBERG, U. & SJÖGREN, H. (1974). Judgments and dimension analyses of perceived sound quality of soundreproducing systems. Journal of the Acoustical Society of America, 55, 854-861.

GABRIELSSON, A., ROSENBERG, U. & SJÖGREN, H.(1973). Adjective ratings and dimension analyses of perceived sound quality of soundreproducing systems. Reports Psychol.Lab., University of Uppsala, No. 141. (Also as Reports from Technical Audiology, Karolinska Institutet, Stockholm, No. 72.)

GABRIELSSON, A. & SJÖGREN, H. (1975). Adjective ratings and dimension analyses of perceived sound quality of hearing aids II. Reports from Technical Audiology, Karolinska Institutet, Stockholm, No. 76.

GLEISS, N. (1974). Preferred listening levels in telephony. TELE, Swedish Telecommunications Administration, No. 2, 24-27.

HAYS, W. L. (1973). Statistics for the social sciences (2nd ed). New York: Holt, Rinehart & Winston.

ILMONEN, K. (1971) Listener preferences for loudness balance of broadcasts, with special consideration to listening in noise. Research reports from the section for Long-Range Planning, the Finnish Broadcasting Company, No. 12.

KIRK, R.E. (1968). Experimental design: Procedures for the behavioral sciences. Belmont, California: Brooks/Cole.

KOPRA, L.L. & BLOSSER, D. (1968). Effects of method of measurement on most comfortable loudness level for speech. J.Speech Hearing Res., 11, 497-508.

OLSON, H.F. (1967). Music, physics and engineering.(2nd. ed.). New York: Dover Publications.

POLLACK, I. (1952). Comfortable listening levels for pure tones in quiet and noise. J. Acoust. Soc. Amer., 24, 158-162

VENTRY, I. M., WOODS, R. W., RUBIN, M. & HILL, W. (1971). Most comfortable loudness for pure tones, noise, and speech. J. Acoust. Soc. Amer., 49, 1805-1813

WINER, B.J. (1971). Statistical principles in experimental design (2nd ed). New York: McGraw-Hill.

TABLE I. Relative level (dB) at the octaves 500 and 1000 Hz, and broadband (dB, A-weighted) for loudspeakers A - D using a white noise signal and with loudspeaker A as reference.

Loudspeaker	500 Hz	1000 Hz	Broadband
A	0	0	0
B	+1	-1	+1
C	+2	0	+1
D	+4	-3	0

TABLE II. Group mean MCL adjustments for loudspeakers A - D (averaged over programs) and for programs 1 - 7 (averaged over loudspeakers). The values refer to number of decibels relative the original sound level set to 0 dB.

	Loudspeaker				Program						
	A	B	C	D	1	2	3	4	5	6	7
Hi-fi	-4.2	-8.0	-5.4	-6.3	-3.4	-7.5	-8.8	-9.3	-6.1	-3.8	-3.9
Non hi-fi	-7.7	-9.4	-8.6	-8.0	-7.2	-8.2	-9.3	-11.5	-10.9	-4.8	-7.3

TABLE III. Dynamic range between programs in original and in the MCL adjustments of thirteen subjects, representing "high setters", "true-to-nature setters", and "low-setters". The values refer to the difference between dB(A) level for programs 2-4 treated as a unit (designated P_{2-4}) versus program 6 (P_6), program 1 (P_1) and program 7 (P_7), respectively. Data refer to reproductions by loudspeaker A.

	P_{2-4} - P_6	P_{2-4} - P_1	P_{2-4} - P_7
Original difference	23-26	16-19	13-16
"High setters":			
Subject 1	26-27	14-15	12-13
" 2	17-20	8-11	6-9
" 11	19-22	10-13	7-10
"True-to-nature":			
Subject 3	21-24	13-16	12-15
" 4	18-21	16-19	13-16
" 12	22-23	13-14	13-14
" 14	17-20	14-17	11-14
"Low setters":			
Subject 8	18-25	11-18	8-15
" 9	17-18	15-16	12-13
" 10	17-19	10-12	11-13
" 18	23-25	15-17	17-19
" 19	16-19	18-21	15-18
" 20	19-21	21-23	13-15

TABLE IV. Mean "true-to-nature" ratings (averaged over subjects) in the hi-fi group and the non hi-fi group for all combinations of loudspeakers, levels and programs.

			Loudspeaker A		B		C		D		Means for programs High Lev.	Means for programs Low Lev.	Means for programs Total
	Program	Level	High	Low	High	Low	High	Low	High	Low			
Hi-fi	1		6.0	6.2	4.8	5.4	5.5	5.6	3.9	4.8	5.1	5.5	5.3
	2		6.5	6.6	2.9	4.7	5.7	5.8	3.7	4.1	4.7	5.3	5.0
	3		6.4	5.7	2.7	3.2	6.0	5.0	3.1	3.1	4.6	4.3	4.5
	4		7.4	6.0	3.1	4.2	6.5	5.9	2.3	2.7	4.8	4.7	4.8
	5		6.5	6.6	4.4	4.4	6.8	6.1	4.3	4.6	5.5	5.4	5.5
	6		6.0	6.9	3.7	4.0	5.4	6.3	4.1	4.5	4.8	5.4	5.1
	Mean		6.5	6.3	3.6	4.3	6.0	5.8	3.6	4.0	4.9	5.1	5.0
	Mean loudspeaker		6.4		4.0		5.9		3.8				
Non Hi-fi	1		5.4	5.4	5.3	5.3	5.1	5.7	4.8	5.3	5.2	5.4	5.4
	2		6.8	6.5	5.7	5.3	7.0	6.0	6.2	4.8	6.4	5.7	6.1
	3		7.3	6.3	5.1	5.3	7.1	6.1	4.4	4.6	6.0	5.6	5.8
	4		6.7	4.7	5.5	4.1	7.1	4.8	4.5	4.1	6.0	4.4	5.2
	5		6.6	7.1	5.6	5.0	6.2	7.5	4.5	4.2	5.7	6.0	5.9
	6		6.1	6.5	4.9	5.1	5.7	7.3	5.4	5.6	5.5	6.1	5.8
	Mean		6.5	6.1	5.4	5.0	6.4	6.2	5.0	4.8	5.8	5.5	5.7
	Mean loudspeaker		6.3		5.2		6.3		4.9				

TABLE V. Significances, variance accounted for, and intra-individual standard deviations for all subjects. x means "significant at 5% level", xx means "significant at 1% level". The number following after the significance symbol denotes per cent variance accounted for by the respective source of variation.

		Sources of variation								
	sub-jects	Loud-speaker(L)	Level (LEV)	Program (P)	LxLEV	LxP	LEVxP	LxLEVxP	Sum of variance accounted for,%	Intra-individual standard deviation
Hi-fi	1	xx 61%	x 1%	xx 6%	xx 1%	xx 10%	xx 5%	xx 3%	87%	0.80
	2	xx 19	xx 3	x 3	xx 5	xx 13		xx 9	52	0.98
	3	xx 61		xx 10		1		1	73	0.80
	4	xx 60	xx 1	xx 14	xx 2	xx 5	xx 2	x 2	86	1.19
	5	xx 31		xx 20					51	1.27
	6	xx 45	xx 2	xx 12		xx 8	xx 7	xx 8	82	0.52
	7	xx 67	xx 2	xx 2		xx 10	xx 2	xx 2	85	0.71
	8	xx 67		x 2		xx 9	x 2		80	0.77
	9	x 6		xx 19					25	1.33
	10	xx 27		xx 16	x 3	x 4	x 3	2	55	1.06
Non hi-fi	11	xx 17		xx 13	xx 6	xx 16	x 2	xx 12	66	1.14
	12	xx 11	xx 28	xx 13	xx 3	xx 7	xx 6		68	1.22
	13	xx 26	x 2	xx 9	x 3	4	xx 5	1	50	1.27
	14	xx 14	xx 19	xx 31			xx 3		67	0.85
	15	xx 8		xx 19	x 3	x 6	xx 10		46	1.78
	16	xx 21	xx 3	xx 17		xx 7	xx 9		57	1.04
	17	xx 52	xx 2	xx 9		xx 17			80	0.79
	18	xx 9	xx 13	x 3	3		xx 11		39	1.66
	19		xx 21	xx 16		1	xx 6		43	1.60
	20	x 3		xx 33			xx 9		45	1.24

TABLE VI. Mean "true-to-nature" ratings for "high setters"(HS) "true-to-nature setters"(TTNS), and "low setter" (LS) at all combinations of loudspeakers, programs and levels. The range of ratings within each category appears in the right margin for "high" and "low" level, respectively.

PROGRAM 1

		Loudspeaker A		Loudspeaker B		Loudspeaker C		Loudspeaker D		Range of ratings	
	Level	High	Low	High	Low	High	Low	High	Low	High	Low
Hi-fi	HS(2)	4.2	5.8	4.4	4.6	4.0	4.2	3.5	4.4	0.9	1.6
	TTNS(5-6)	6.9	6.6	5.1	5.5	6.4	6.4	4.0	4.9	2.9	1.7
	LS(2-3)	5.9	5.7	4.8	5.9	5.0	5.1	4.0	4.8	1.9	1.1
Non Hi-fi	HS(1)	5.0	6.7	6.7	7.7	5.0	6.0	4.0	6.0	2.7	1.7
	TTNS(6)	4.9	5.3	4.9	5.0	4.8	5.6	4.2	4.9	0.7	0.7
	LS(3)	6.7	5.3	5.6	5.1	5.7	5.7	6.1	6.6	1.1	1.5
	Program 2										
Hi-fi	HS(1)	6.0	7.8	1.3	4.8	7.5	8.8	3.5	3.8	6.2	5.0
	TTNS(4-5)	6.4	6.5	2.6	4.8	5.2	5.6	4.1	3.9	3.8	2.6
	LS(4-5)	6.7	6.4	3.6	4.6	5.6	5.4	3.4	4.2	3.3	2.2
Non Hi-fi	TTNS(7)	6.6	6.7	4.9	5.3	6.5	5.9	5.6	4.6	1.7	2.1
	LS(3)	7.5	5.9	7.4	5.2	9.1	6.4	7.5	5.3	1.7	1.2
	Program 3										
Hi-fi	TTNS(4-6)	6.2	5.8	2.9	3.3	6.1	4.8	4.4	3.6	3.3	2.5
	LS(4-6)	6.6	5.6	2.3	2.9	6.0	5.3	2.3	2.7	4.3	2.9
Non Hi-fi	TTNS(5-7)	6.8	6.4	3.1	4.7	6.6	6.2	3.6	4.2	3.7	2.2
	LS(3-5)	8.6	6.1	7.0	5.9	8.0	5.8	4.9	4.9	3.7	1.2
	Program 4										
Both groups	HS(1)	8.3	5.8	2.7	3.0	8.2	6.5	2.8	3.3	5.6	3.5
	TTNS(3-7)	7.3	5.5	5.0	3.9	6.7	4.9	3.0	2.6	4.3	2.9
	LS(2-6)	5.8	4.6	3.7	4.6	6.6	5.6	4.1	4.4	2.9	1.2

		Program 5									
Both groups	TTNS(6)	6.3	7.2	5.1	4.4	6.4	7.0	3.9	4.4	2.5	2.8
	LS(4)	7.0	6.4	5.0	5.1	6.6	6.5	5.2	4.5	1.8	2.0
		Program 6									
Hi-fi	TTNS(5-7)	6.0	7.1	3.8	3.8	5.6	6.3	3.8	4.4	2.2	2.7
	LS(3-5)	5.9	6.5	3.6	4.1	5.2	6.4	4.9	4.8	2.3	2.4
Non Hi-fi	TTNS(7)	6.4	6.6	4.9	4.7	6.1	7.2	5.2	5.4	1.5	2.5
	LS(3)	6.2	6.4	4.9	5.9	5.4	7.6	5.6	5.9	1.3	1.7

SAMMANFATTNINGSMATRISER

Music section 1

Loudspeaker

Subjects	Category	A		B		C		D		E	
	L	6.5	5.9	6.6	6.5	5.6	4.6	5.4	4.8	4.3	3.3
		7.6	6.6	8.3	6.5	5.8	5.0	6.1	5.0	4.1	3.1
		7.4	8.3	9.5	9.7	4.9	3.4	5.6	3.2	1.1	0.7
		7.5	7.5	8.0	7.3	6.3	4.8	7.0	4.6	3.0	2.4
	M	7.4	7.6	8.6	8.7	6.6	4.6	6.8	5.9	4.1	3.6
		7.0	7.1	6.3	6.7	4.8	5.1	5.5	5.1	3.0	3.0
		8.0	6.6	7.0	6.5	3.8	2.9	4.3	1.9	1.0	1.0
		4.5	4.6	7.8	7.3	4.0	3.6	2.5	3.6	2.0	2.3
	HF	6.0	7.1	7.8	7.9	3.0	3.3	3.3	4.2	2.0	1.5
		7.2	7.4	8.2	8.2	4.1	3.8	4.6	3.7	1.4	1.0
		7.0	8.1	8.6	8.7	5.1	5.5	4.8	5.6	3.6	4.0
		6.2	4.7	8.2	6.0	4.2	2.7	4.1	3.1	3.2	2.1

Music section 2

Subjects	Category										
	L	7.6	7.2	6.4	7.4	6.4	6.9	6.0	6.6	5.1	5.3
		7.2	7.6	6.7	7.3	6.4	6.5	5.7	6.4	5.7	4.8
		9.5	8.8	9.4	7.6	6.0	4.8	5.5	4.6	2.9	2.3
		5.8	6.6	7.0	6.8	5.0	4.8	4.0	4.3	3.3	3.4
	M	7.8	7.6	8.0	7.3	5.5	5.4	5.0	5.3	3.8	4.1
		7.3	7.6	6.8	7.1	5.0	5.1	5.5	5.2	3.5	3.3
		8.8	8.0	6.8	7.0	5.8	5.4	5.5	5.8	3.5	3.0
		5.3	6.9	6.3	6.1	5.0	4.9	4.5	4.4	3.5	3.5
	HF	7.8	7.9	8.0	7.4	5.5	5.6	4.5	5.6	2.8	2.6
		7.3	7.6	7.9	7.9	4.6	4.2	5.0	4.0	2.4	2.4
		7.9	8.9	8.0	8.5	5.6	6.8	5.2	6.0	5.0	5.0
		8.2	7.3	7.5	7.3	5.7	5.7	5.2	5.1	3.5	4.0

Music section 3

Category											
	L	7.3	7.3	8.0	6.9	7.6	6.9	7.4	7.5	6.0	5.9
		7.9	7.1	7.0	6.7	7.2	6.8	7.0	6.9	5.4	5.9
		8.4	6.9	8.6	7.5	6.3	5.3	7.9	7.6	4.4	2.5
		6.3	7.5	8.5	8.0	5.3	6.5	6.3	6.5	2.8	2.8
	M	7.6	7.3	7.4	7.6	7.1	7.2	7.5	7.3	4.9	4.4
		7.5	7.8	7.0	7.3	4.9	5.5	5.8	6.4	3.5	3.6
		8.0	8.0	6.3	6.5	5.3	5.5	6.5	7.0	3.0	3.5
		5.3	5.6	4.5	6.3	5.0	6.0	5.5	7.5	3.3	4.1
	HF	7.3	7.4	7.3	7.6	6.0	6.3	6.8	6.9	2.8	3.2
		7.0	7.7	7.6	7.9	5.1	4.9	6.4	6.4	3.0	3.0
		7.6	8.3	7.6	8.0	6.2	7.1	6.2	6.6	4.7	5.0
		6.8	6.8	6.2	6.6	5.3	5.7	6.4	6.1	4.0	4.1

Music section 4

Subjects		A		B		C		D		E	
Category L		7.8	6.9	7.6	7.1	6.0	6.0	6.8	6.6	4.9	5.3
		7.8	7.5	8.1	7.4	6.4	6.2	6.9	6.9	4.6	5.8
		8.9	8.0	9.8	9.5	4.9	4.7	4.3	4.2	2.1	1.9
		7.0	7.8	7.8	7.6	5.5	6.3	6.5	5.5	2.8	3.0
Category M		7.8	8.1	7.8	8.3	6.8	6.6	6.6	6.3	4.4	4.3
		7.5	8.0	6.5	7.2	5.0	5.3	5.3	5.3	3.0	3.0
		8.5	7.9	7.0	7.0	5.8	5.3	6.8	6.5	2.5	3.0
		6.0	6.3	7.0	7.5	6.5	5.0	4.3	4.4	3.3	3.0
Category HF		7.5	7.7	7.5	8.0	5.0	5.2	4.5	5.6	2.1	2.6
		7.5	7.8	8.4	8.0	5.0	4.3	5.4	4.6	2.3	2.5
		8.8	9.0	8.6	8.8	5.6	6.6	6.0	6.9	4.6	4.9
		8.0	7.4	7.9	7.3	5.9	5.4	5.7	5.5	3.8	4.0

Music section 5

Subjects											
Category L		7.4	6.4	7.3	6.6	6.9	6.9	7.8	6.9	6.1	5.8
		7.4	7.1	7.3	7.6	7.0	7.1	6.3	7.0	5.3	5.9
		8.9	6.4	8.9	8.1	8.3	6.7	8.3	9.4	4.8	3.6
		7.5	8.4	8.3	7.9	6.3	6.6	8.5	7.1	3.8	3.1
Category M		7.9	7.8	7.9	7.6	7.6	7.3	7.9	7.2	5.5	5.3
		7.0	8.0	7.0	7.5	5.5	5.5	6.0	6.1	4.0	4.1
		8.3	7.8	7.5	7.0	6.5	5.8	7.0	6.8	4.0	3.9
		6.0	5.6	5.8	6.3	5.5	5.5	7.0	7.1	4.0	3.9
Category HF		6.8	6.7	6.3	7.3	6.0	5.9	6.8	6.5	3.5	3.2
		7.6	7.8	8.1	7.9	6.2	4.9	6.9	5.7	3.2	2.5
		8.1	8.6	8.4	8.3	7.0	7.4	8.0	7.8	5.3	5.3
		8.2	6.9	7.8	7.1	5.6	5.8	6.7	6.1	3.5	4.1

Legend to Figures 1-4:

Frequency curve and distortion (sum of 2nd and 3rd harmonic) for loudspeakers A-D measured in a reverberation chamber.
Test signal: 30 Hz narrow-band noise.
Frequency curve, zero level = 50 dB rel. 1pW
Distortion, zero level = 30 dB rel. 1pW

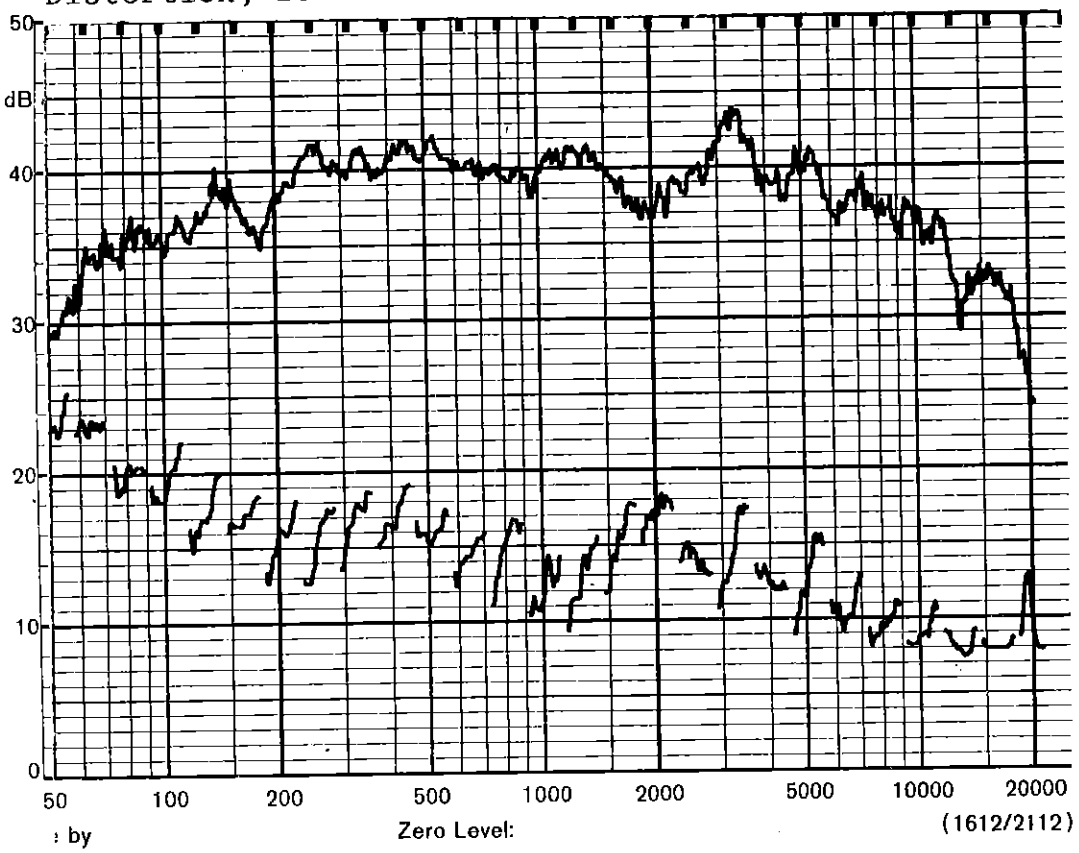


Figure 1. Loudspeaker A

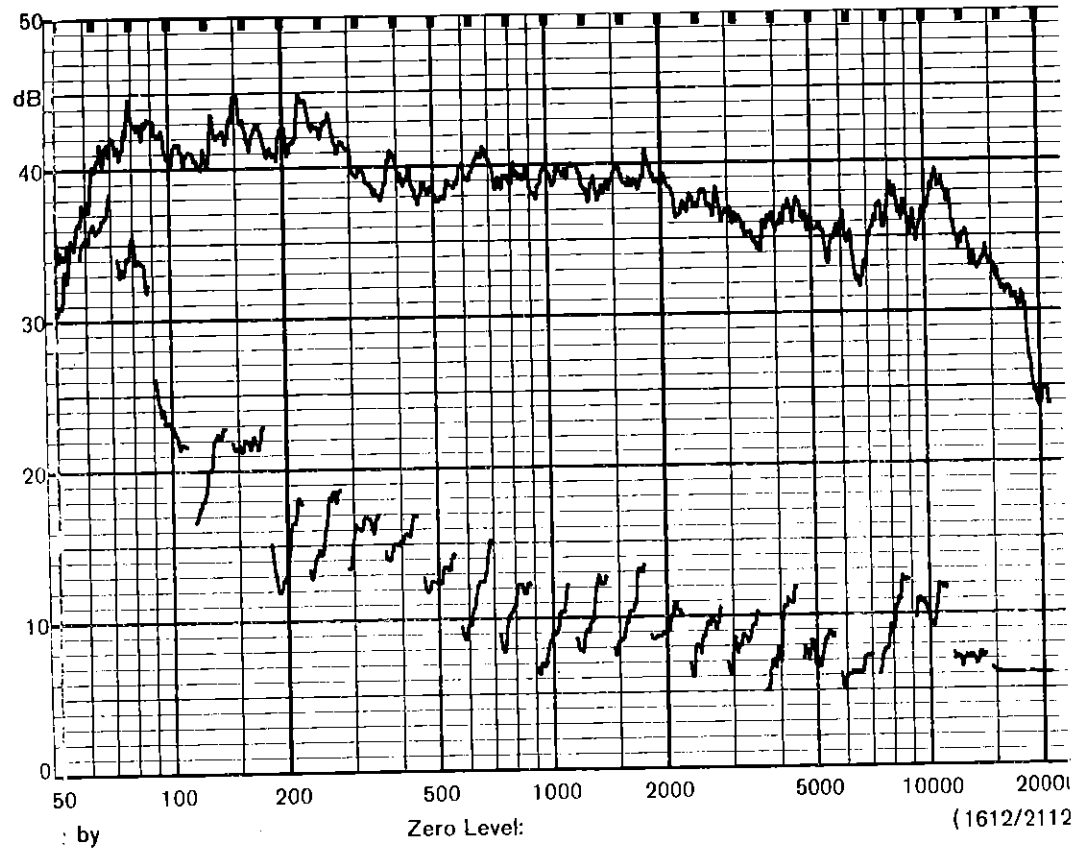


Figure 2. Loudspeaker B

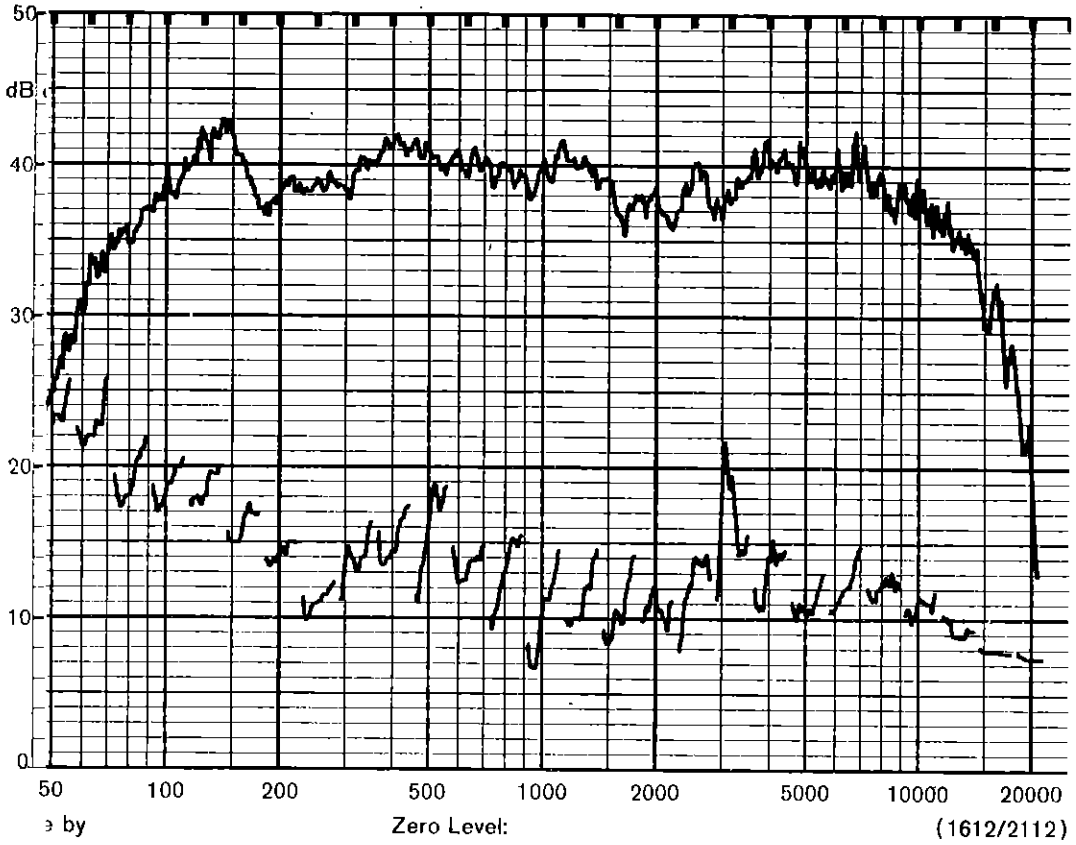


Figure 3. Loudspeaker C

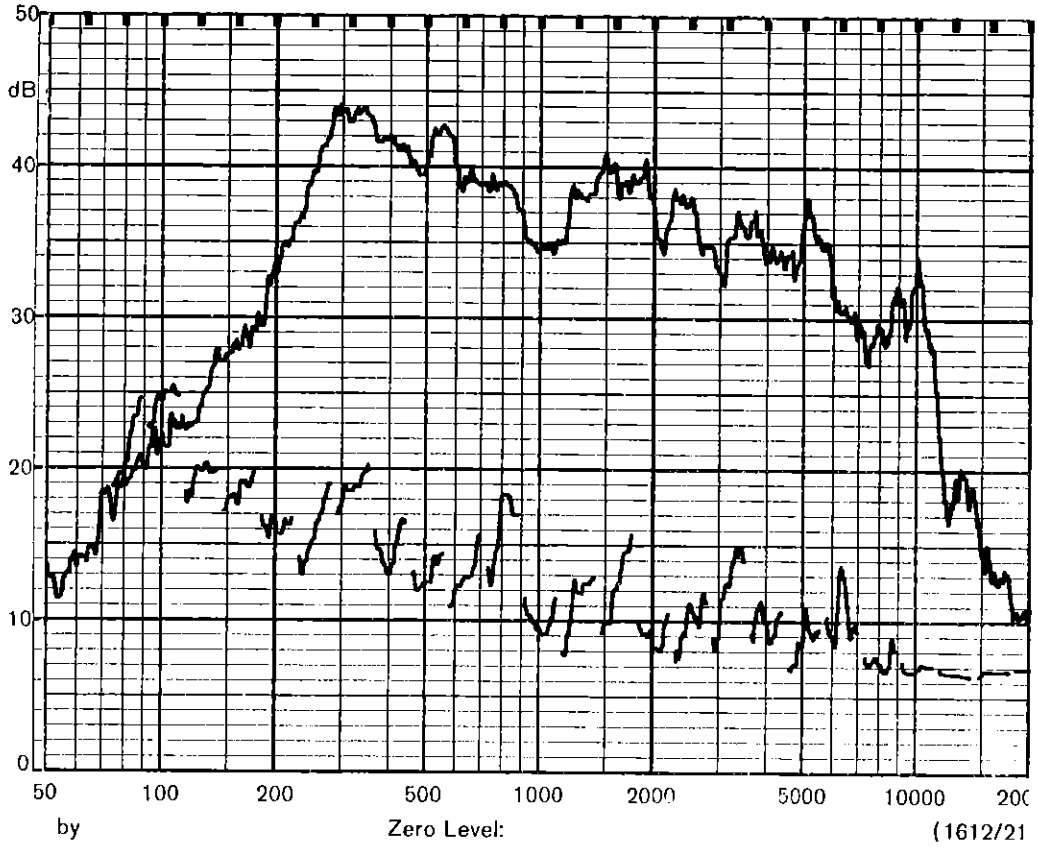


Figure 4. Loudspeaker D

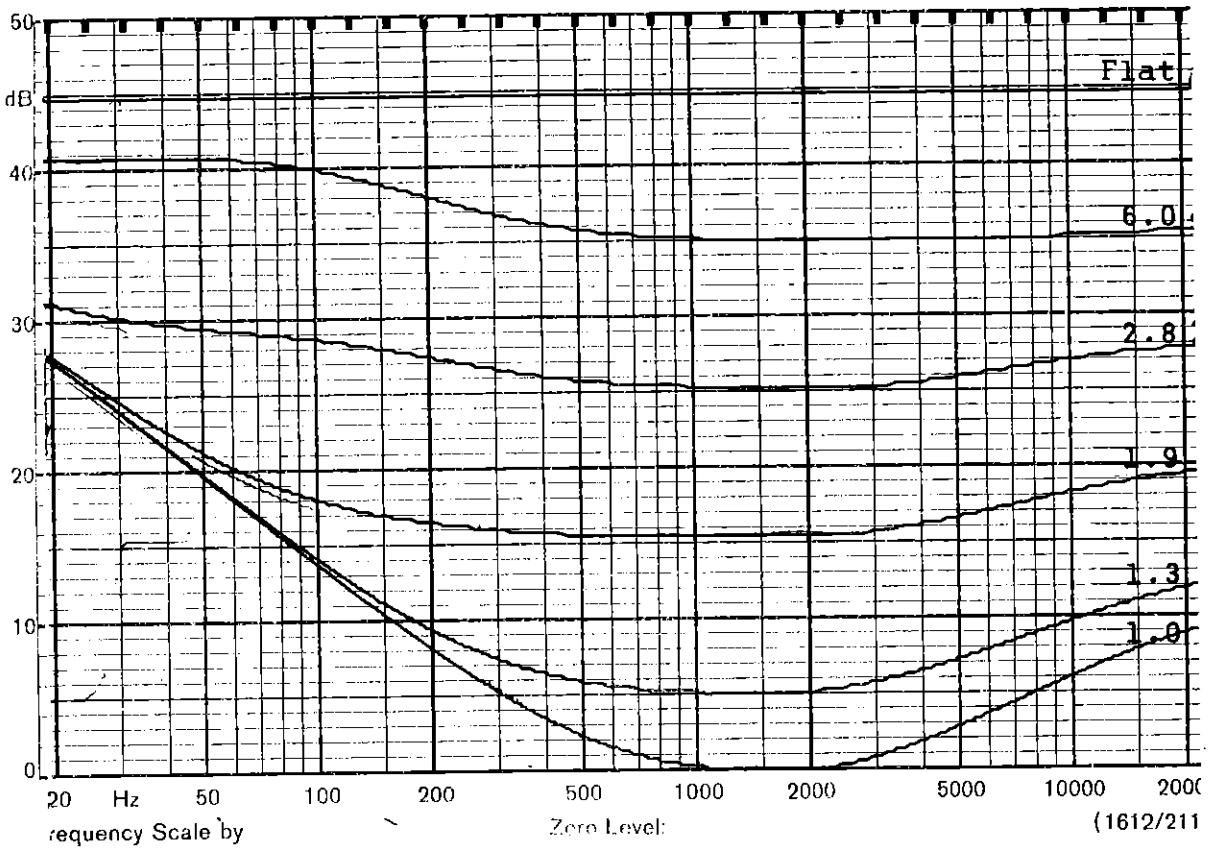


Figure 5. Frequency response of pre-amplifier with "loudness-control" at six different gain control settings.

APPENDIX A: INDIVIDUAL PREFERENCE LEVELS

Individual adjustments of preferent listening level for programs 1-7 reproduced by loudspeakers A-D. The values refer to number of decibels above (denoted by +sign) or below (denoted by -sign) the approximate original sound level at the respective recording session. See further comments in text.

HI-FI SUBJECTS

LOUDSPEAKER A

subject	Program 1	2	3	4	5	6	7
1	+ 7.3	+ 5.9	+ 1.8	+ 3.9	+ 6.0	+ 1.8	+ 5.9
2	+ 6.6	- 0.5	- 0.8	- 0.3	- 0.1	+ 4.9	+ 5.9×
3	+ 2.5	+ 0.2	- 0.6	+ 0.8	+ 1.5	+ 2.0	+ 1.0
4	0.0	+ 0.2	- 0.1	+ 1.3	- 0.5	+ 5.0	- 0.1
5	- 0.9	- 0.9	- 2.1	- 4.6×	- 3.5	- 1.6×	+ 3.0
6	- 0.4	- 3.6	- 0.8	- 2.2	- 0.1	+ 1.7	- 0.1
7	- 3.9	-13.6×	- 7.5×	- 8.3	- 6.3	- 2.1	- 6.3
8	- 6.7	- 9.8	- 8.4	-13.5	-11.9	- 7.1	- 6.8
9	-11.6	-13.0	-15.0	-13.7	-11.6	- 7.3	-11.9×
10	-15.7×	-20.2	-25.2×	-23.6	-21.3	-16.1×	-20.2
mean	- 2.3	- 5.5	- 5.9	- 6.0	- 4.8	- 1.9	- 3.0
stand.dev.	7.4	8.2	8.5	8.7	8.1	6.6	8.4

LOUDSPEAKER B

subject	Program 1	2	3	4	5	6	7
1	+ 5.2	+ 3.9×	- 1.9	+ 1.4	+ 2.2	- 8.8	+ 3.3
2	- 0.1×	- 1.2	- 1.2	- 1.4	- 0.8	+ 4.3	+ 3.2
3	+ 2.6	- 1.4	- 2.8	-13.2×	- 0.1	- 3.9	+ 1.5
4	- 1.1	- 3.2	-15.6	-20.0	-16.8	- 2.8	- 1.1
5	- 2.3×	- 2.4	- 7.5×	-10.5	- 1.8	+ 0.6	- 4.7
6	- 2.4	- 4.1	- 7.0×	-13.3	- 0.1	- 2.0	- 2.0×
7	- 2.2	-13.1	-10.1×	-18.8×	-20.6×	- 6.1	-11.3×
8	- 9.0	-16.3	-18.1	-19.6×	-16.1	-11.1	- 7.4
9	-13.6	-15.5	-16.9	-15.1	-12.5	- 7.9	-12.7
10	-20.2	-27.1×	-21.9	-26.2	-24.8	-20.4	-19.2×
mean	- 4.3	- 8.0	-10.3	-13.7	- 9.1	- 5.8	- 5.0
stand.dev.	7.7	9.5	7.4	8.5	10.1	6.9	7.5

LOUDSPEAKER C

	Program 1	2	3	4	5	6	7
1	+ 8.6	+ 5.9	+ 0.2	+ 3.9	+ 5.9	- 3.8×	+ 5.9
2	+ 5.2	- 0.6	- 1.4	- 0.9	- 0.1	+ 4.9	+ 5.9×
3	+ 0.8	- 0.5	- 0.4	+ 0.1	+ 1.4	+ 0.6	+ 0.8
4	- 1.0	- 2.8	- 0.6	- 2.7	0.0	+ 2.7×	- 0.1
5	- 2.4	- 2.4	- 3.0	- 3.7	- 2.2	+ 2.9×	- 4.8
6	- 1.1	- 8.6	- 1.8×	- 8.4×	- 1.4	- 0.3	- 0.2×
7	- 2.9	-17.8	-13.9	- 9.1×	- 3.2×	- 7.7×	- 3.9×
8	-10.1×	-12.8	-14.8	-12.4	-10.1	-15.3	- 5.6
9	-13.0	-12.2	-14.5	-13.4	-11.7	- 6.8	-12.9
10	-22.1	-21.2	-26.5	-23.9	-22.4	-16.2	-21.7
mean	- 3.8	- 7.3	- 7.7	- 7.1	- 4.4	- 3.9	- 3.7
stand.dev.	9.0	8.6	9.1	8.2	8.2	7.5	8.5

LOUDSPEAKER D

	Program 1	2	3	4	5	6	7
1	+ 6.8	+ 3.2	- 1.9	+ 1.4	+ 5.9×	- 1.9	+ 5.2
2	+ 5.2	- 1.2	- 0.8	- 0.4	- 0.1	+10.9×	+ 3.2
3	+ 1.1	- 6.3×	-13.8×	- 0.8×	- 1.0	- 2.3	- 0.6
4	- 3.1	- 6.3	-12.2	- 5.9×	- 9.1	- 2.2	- 3.2
5	- 2.9	- 2.2	- 3.3	- 2.3	- 1.6×	+ 2.7	- 0.5
6	- 0.1	- 9.8	- 6.7	-13.2	- 6.0×	+ 0.4	- 0.2
7	- 5.0	-16.3×	-19.3	-19.6×	-10.3	- 2.5	-10.9
8	- 2.5×	- 9.3×	-12.7	- 9.8×	- 7.8×	- 8.1	- 3.2
9	-12.2	-13.8	-13.3	-14.3	-11.3	- 7.2	-10.1
10	-17.6×	-24.7	-24.6	-24.3	-19.6	-16.2	-19.7
mean	- 3.0	- 8.7	-10.9	- 8.9	- 6.1	- 2.6	- 4.0
stand.dev.	7.4	8.1	7.7	8.8	7.2	7.1	7.5

NON HI-FI SUBJECTS

LOUDSPEAKER A

subject	Program 1	2	3	4	5	6	7
11	+ 5.9	- 0.1	0.0	-10.7^{X}	- 0.9	+ 3.8	+ 5.9^{X}
12	+ 1.9	- 0.8	- 2.8^{X}	- 2.5	- 0.2^{X}	- 0.3	- 0.5
13	- 6.3	0.0	- 0.2	- 0.2	0.0	+ 3.9	- 0.8
14	- 0.2	- 1.5	- 2.1	- 3.4	- 2.1	+ 3.5	- 0.1
15	- 0.1^{X}	- 3.1^{X}	- 4.1	- 2.1	- 5.6	- 2.8	- 0.6
16	- 5.1	- 1.2	- 0.5	- 0.1^{X}	-16.9^{X}	- 2.9	- 2.1^{X}
17	- 4.1	- 5.4	- 5.9	- 8.1	- 4.5	- 3.9	- 5.3
18	-16.6	-18.3	-18.8	- 8.4^{X}	-22.3^{X}	-17.9^{X}	-22.0
19	-24.4^{X}	-21.8^{X}	-22.4	-32.4	-20.6^{X}	-14.5	-23.8
20	-23.4	-18.1	-19.2^{X}	-34.0	-30.9	-14.2	-18.3^{X}
mean	- 7.2	- 7.0	- 7.6	-10.2	-10.4	- 4.5	- 6.8
stand.dev.	10.6	8.7	8.9	12.7	11.2	8.2	10.5

LOUDSPEAKER B

subject	Program 1	2	3	4	5	6	7
11	+ 5.9	- 0.4	- 0.8	-13.6^{X}	- 5.2	+ 4.4^{X}	+ 5.9
12	- 0.1	- 1.7	- 6.6	- 0.6^{X}	- 3.4	+ 3.1	- 0.8
13	- 0.1	- 5.4	- 2.7	- 1.1^{X}	0.0	+ 3.5	0.0
14	- 0.1	- 2.8	- 2.4	- 5.9	- 4.7	+ 2.1	- 0.3
15	- 0.8	- 4.9	- 4.2^{X}	- 9.5	-15.7	- 0.5^{X}	- 6.7
16	- 8.3^{X}	- 4.6	-14.6^{X}	- 8.4^{X}	-12.1^{X}	- 4.6	- 9.6
17	- 2.8	- 6.0	-12.4	-15.7	-11.3	- 2.4^{X}	- 4.9
18	-16.0	-30.0	-26.2	-17.3	-13.8^{X}	-22.4	-15.1^{X}
19	-19.2	-19.2	-19.7	-26.1^{X}	-24.4^{X}	-21.8	-25.2^{X}
20	-23.6^{X}	-17.4	-17.8	-32.0	-32.8	-18.3	-27.1^{X}
mean	- 6.5	- 9.2	-10.7	-13.0	-12.3	- 5.7	- 8.4
stand.dev.	9.8	9.7	8.7	10.2	10.1	10.9	11.0

LOUDSPEAKER C

subject	Program 1	2	3	4	5	6	7
11	+ 5.1	- 2.0[x]	0.0	-7.3	- 0.2	+ 5.0[x]	+ 5.9
12	- 0.2	- 0.9	-10.6	- 8.2	- 0.9	+ 4.9[x]	- 1.9
13	- 0.7	- 0.1	- 0.8	- 1.3[x]	- 1.5	+ 3.0	- 2.8
14	- 0.4	- 1.4	- 3.2	- 5.0	- 2.7	+ 2.3	- 0.1
15	- 2.3	- 1.4	- 5.8	- 0.9	- 4.2	- 5.0	- 1.8
16	-13.2[x]	- 4.1	- 1.7	- 7.8[x]	- 7.2	- 2.6	- 2.1[x]
17	- 3.9	- 7.4	- 6.6	- 7.2	- 3.4	- 6.1	- 6.4
18	-20.1[x]	-27.3	-23.6[x]	-14.0	-18.3[x]	-17.9[x]	-24.3
19	-24.9	-22.3[x]	-16.6[x]	-30.2	-31.2[x]	-14.2[x]	-17.6[x]
20	-29.6[x]	-15.8	-18.1	-34.0	-31.9	-22.6	-20.9
mean	- 9.0	- 8.3	- 8.7	-11.6	-10.2	- 5.3	- 7.2
stand.dev.	12.1	9.9	8.2	11.5	12.4	9.9	10.1

LOUDSPEAKER D

subject	Program 1	2	3	4	5	6	7
11	+ 5.8	- 0.1	- 0.2	-1.2[x]	- 1.3	+ 4.9	+ 5.9
12	- 0.1[x]	- 0.2[x]	- 0.2[x]	- 3.0	- 0.1	+ 3.7	- 0.1
13	+ 0.9	- 4.7	- 1.7	- 0.9	- 2.0	+ 4.0	0.0
14	- 0.1	- 1.6	- 1.6	- 3.2	- 1.8	+ 4.4	- 0.2
15	- 0.7	- 2.0	- 3.1[x]	- 3.6[x]	- 1.0[x]	+ 1.6	- 0.4
16	- 2.1	- 3.8[x]	-20.2	-15.6	-17.2[x]	- 4.0	- 3.8
17	- 4.9	- 9.8	- 9.4	- 9.7	- 6.7	- 3.9	- 7.3
18	-10.7	-26.1	-20.4	-21.1[x]	-21.7	-18.1	-23.2
19	-25.2	-14.6[x]	-17.6[x]	-19.6[x]	-26.9	-15.8	-19.7
20	-22.4	-19.0	-19.3	-33.4	-31.3	-14.8	-18.8[x]
mean	- 6.0	- 8.2	- 9.4	-11.1	-11.0	- 3.8	- 6.8
stand.dev.	10.3	9.0	9.0	10.9	12.1	9.2	10.1

APPENDIX B: INDIVIDUAL "TRUE-TO-NATURE" RATINGS

Individual "true-to-nature" ratings (0-10 scale) for loudspeakers A-D reproducing programs 1-6 at two different sound levels. Each value represents the mean of three judgments. Group means are given in the bottom margin for the respective group.

PROGRAM 1 (DEBUSSY)

		Loudspeaker A		B		C		D	
	Level	High	Low	High	Low	High	Low	High	Low
Hi-fi subjects	1	4.7	5.8	4.5	4.2	4.3	4.3	2.7	3.8
	2	3.7	5.7	4.3	5.0	3.7	4.0	4.3	5.0
	3	5.3	6.3	3.7	4.0	5.0	5.3	3.0	4.0
	4	9.0	7.7	5.3	7.0	9.0	7.7	5.0	5.3
	5	7.3	7.3	6.0	5.3	5.7	6.3	4.7	6.0
	6	6.2	5.5	5.5	6.0	6.2	5.8	3.3	5.0
	7	6.7×	6.3	5.0×	5.3	6.0×	6.7	3.3×	4.7
	8	6.7	6.3	4.3	5.0	6.3	6.3	5.0×	4.3
	9	3.7	4.3	3.7	4.0	2.3	3.0	2.7	3.3
	10	7.2	6.6	5.8	7.8	6.3	6.0	5.2	6.3
	Group mean	6.0	6.2	4.8	5.4	5.5	5.6	3.9	4.8
Non hi-fi subjects	11	5.0	6.7	6.7	7.7	5.0	6.0	4.0	6.0
	12	2.7	2.7	3.3	3.0	3.0	4.0	3.3	3.7
	13	2.3	5.3	3.7	6.0	3.3	5.3	3.0	4.0
	14	4.7	3.7	5.3	3.3	4.7	4.3	5.3	3.3
	15	5.7	4.7	5.7	5.0	6.3	5.0	3.7	5.7
	16	4.7	6.3	4.3	4.7	4.3	6.3	4.0	5.7
	17	9.0	9.0	6.8	7.7	7.3	8.5	5.7	6.7
	18	7.8	4.7	6.0	3.7	5.0	4.5	7.0	4.3
	19	5.2	4.0	5.7	4.5	4.5	4.3	5.8	6.5
	20	7.2	7.3	5.0	7.0	7.5	8.2	6.3	6.7
	Group mean	5.4	5.4	5.3	5.3	5.1	5.7	4.8	5.3

PROGRAM 2 (BACH)

Loudspeaker

		A		B		C		D	
	Level	High	Low	High	Low	High	Low	High	Low
Hi-fi subjects	1	6.0	7.8	1.3	4.8	7.5	8.8	3.5	3.8
	2	4.3	5.7	1.0	5.7	5.3	5.0	6.3	5.3
	3	7.5	5.3	2.7	4.0	5.0	4.3	3.0	3.0
	4	7.3	9.0	0.0	5.0	4.3	7.0	0.7	1.7
	5	8.0	7.0	4.7	4.3	6.3	6.0	6.3	5.7
	6	5.0	5.5	4.8	4.8	3.5	4.2×	2.5	2.8×
	7	6.3	7.7	3.3	4.0	5.7	7.0	3.3	5.0
	8	7.3	6.0	3.3	4.0	7.0	6.3	3.0	3.3
	9	6.0	5.0	3.3	4.3	5.3	4.3	4.3	4.7
	10	7.2	6.7	4.3	6.0	6.5	5.3	3.8	5.3
	Group mean	6.5	6.6	2.9	4.7	5.7	5.8	3.7	4.1
Non hi-fi subjects	11	8.7	8.0	2.3	8.0	8.7	7.3	8.0	6.0
	12	0.7	5.0	0.0	2.3	0.3	3.7	1.0	3.0
	13	5.3	7.7	5.0	5.0	6.0	7.3	5.3	3.7
	14	7.3	5.3	6.0	4.3	7.0	5.0	6.3	4.0
	15	8.5	6.0	8.3	6.5	9.0	5.7	7.3	4.0
	16	7.7	7.0	7.0	5.0	6.7	5.3	6.0	5.7
	17	7.7	8.2	6.0	6.3	7.5	7.3	5.0	5.7
	18	6.6	4.8	5.3	2.8	6.6	6.0	6.7	3.3
	19	6.8	4.7	9.2	5.5	9.7	4.5	7.3	4.2
	20	9.0	8.2	7.7	7.3	8.5	8.2	8.5	8.5
	Group mean	6.8	6.5	5.7	5.3	7.0	6.0	6.2	4.8

PROGRAM 3 (STRAVINSKY)

Loudspeaker

		A		B		C		D	
	Level	High	Low	High	Low	High	Low	High	Low
Hi-fi subjects	1	7.7	6.3	2.8	2.8	8.0	5.7	2.7	3.2
	2	6.7	6.0	1.7	4.3	6.0	6.0	6.3	5.0
	3	4.7	4.3	2.3	3.0	3.7	3.7	1.3	2.0×
	4	6.0	7.0	0.0	1.3×	7.0	5.3	0.3	0.7×
	5	6.3	5.3	2.0	2.0	6.0	3.3	4.5	2.7
	6	6.0	5.8	5.5	4.5	5.8	4.8	4.2	3.5
	7	7.7	8.0	2.0	3.7	7.0	6.3	3.0	3.0
	8	7.7	6.0	3.7	3.7	7.7	7.0	3.7	3.3
	9	3.8	4.0	2.3	2.7	3.0	3.0	2.7	4.0
	10	7.0	4.3	4.9	3.7	6.2	5.0	3.0	3.3
	Group mean	6.4	5.7	2.7	3.2	6.0	5.0	3.1	3.1
Non hi-fi subjects	11	10.0	8.7	4.0	8.0	9.3	9.7	7.0	6.3
	12	1.3	5.0	0.0	2.3	0.0	3.7	1.0	3.3
	13	5.7	4.7	3.7	3.7	7.0	5.0	3.0	3.3
	14	6.7	5.3	4.7	3.7	6.3	5.3	5.7	4.0
	15	7.3	5.3	3.3	5.7	7.3	4.3	1.3	4.0
	16	8.0	6.7	6.7	5.7×	8.3	6.7	3.7	5.0×
	17	8.5	8.8	5.2	6.5×	8.3	8.5	1.0	4.0×
	18	8.6	4.2	5.8	3.5	7.2	4.3	5.2	2.5
	19	9.5	6.7	9.5	6.8	8.7	5.0	7.0	5.8
	20	7.7	7.5	7.7	6.8	8.2	8.0	7.8	7.3
	Group mean	7.3	6.3	5.1	5.3	7.1	6.1	4.4	4.6

PROGRAM 4 (JAZZ)

Loudspeaker

		A		B		C		D	
	Level	High	Low	High	Low	High	Low	High	Low
Hi-fi sub-jects	1	8.3	5.8	2.7	3.0	8.2	6.5	2.8	3.3
	3	7.0	6.3	3.3	3.3×	4.7	4.0	2.0	2.7
	4	8.0	6.3	2.3	6.0×	6.7	6.0	0.7	0.7
	6	6.2	5.7	4.3	5.0×	5.3	5.7×	3.2	3.7×
	7	7.7×	5.7	3.0	3.7	7.7	7.3	3.0	3.0
	Group mean	7.4	6.0	3.1	4.2	6.5	5.9	2.3	2.7
Non hi-fi sub-jects	14	7.0	4.7	5.3	3.3	7.0	5.3	5.0	3.3
	15	6.7	3.0	1.7	1.3	7.0	3.3	2.0	3.0
	16	8.3	6.7	8.0	7.0	8.0	6.0	5.0	5.3×
	19	5.8	4.0	6.8	3.7	7.2	4.5	5.5×	3.3
	20	5.8	5.2	5.8	5.0	6.3	4.8	5.2	5.7
	Group mean	6.7	4.7	5.5	4.1	7.1	4.8	4.5	4.1

PROGRAM 5 (SINGER)

Loudspeaker

		A		B		C		D	
	Level	High	Low	High	Low	High	Low	High	Low
Hi-fi sub-jects	2	5.3	6.0	5.3	4.7	5.7	5.3	5.0	4.7
	5	7.3	8.0	5.3	4.7	7.3	6.0	3.3	4.3
	8	7.0	7.0	3.3	3.0	6.3	7.0	2.3	3.3
	9	5.0	4.7	3.3	4.0	6.0	5.0	5.0	5.7
	10	7.7	7.5	5.0	5.7	8.5	7.1	6.0	4.8
	Group mean	6.5	6.6	4.4	4.4	6.8	6.1	4.3	4.6
Non hi-fi sub-jects	11	10.0	7.3	6.7	7.0	9.0	8.3	5.3	5.7
	12	0.0	5.7	0.0	0.7	1.0	5.7	0.0	2.7
	13	5.3	7.0	5.7	4.3	6.0	7.7	5.0	3.3
	17	9.7	9.2	8.3	7.8×	9.3	9.0	4.7	5.5
	18	8.2	6.5	7.3×	5.0	5.7	6.8	7.5	4.0
	Group mean	6.6	7.1	5.6	5.0	6.2	7.5	4.5	4.2

PROGRAM 6 (SPEECH)

Loudspeaker

		A		B		C		D	
	Level	High	Low	High	Low	High	Low	High	Low
Hi-fi subjects	1	7.3	8.3	3.2	3.8×	6.8	8.0	3.8	5.8
	2	5.0	5.7	3.3	3.7	5.7	5.3	5.7	6.7
	3	5.8	7.0	3.7	3.7	4.5	4.7	3.0	3.0
	4	6.3	8.7	3.7	4.0	7.0	7.7	1.3	0.7
	5	6.7	5.3	4.3	3.7	5.0	5.0	3.7	4.7
	6	4.5	7.2	4.0	4.0	4.5	7.0	4.0	4.0
	7	6.3×	7.7	4.7	4.3	6.0	7.3	5.0×	5.7
	8	5.7	6.7	3.0	3.3	4.7	5.7	4.3	5.3
	9	4.7	5.0	3.0	3.7	4.0	5.3	4.7	4.7
	10	7.3	7.7	4.0	5.3	6.2	7.3	5.6	4.5
	Group mean	6.0	6.9	3.7	4.0	5.4	6.3	4.1	4.5
Non hi-fi subjects	11	6.3	5.3	7.3	6.0	6.7	8.7	5.3	4.7
	12	4.3	6.3	0.7	2.3	3.0	5.0	4.3	5.7
	13	5.7	8.0	4.0	4.0	5.7	6.7	3.3	3.3
	14	7.8	7.0	7.0	4.7	7.3	8.3	6.3	7.3
	15	3.7	6.0	4.0	6.0	4.7	7.7	5.7	7.2
	16	6.3	5.3	5.3	3.7	7.3	6.0	4.7	3.0
	17	7.7	7.7	6.0	6.5	6.8	8.5	7.0	6.8
	18	5.7	6.0	2.8	4.2	5.0	7.3	4.2	5.2
	19	5.7	4.8	6.2	5.5	5.2	6.3	6.2	4.2
	20	7.3	8.3	5.7	8.0	5.5	8.8	6.5	8.3
	Group mean	6.1	6.5	4.9	5.1	5.7	7.3	5.4	5.6