ISSN 0280-6819


# TECHNICAL AUDIOLOGY

Report TA No. 123
Dec 1991

## PERCEIVED SOUND QUALITY OF REPRODUCTIONS WITH DIFFERENT SOUND LEVELS

Alf Gabrielsson, Björn Hagerman, and
Tommy Bech-Kristensen

KAROLINSKA INSTITUTET
Teknisk Audiologi
Dept. of Technical Audiology

ISSN 0280-6819
TA123, Dec 1991

# PERCEIVED SOUND QUALITY OF REPRODUCTIONS WITH DIFFERENT SOUND LEVELS

Alf Gabrielsson[1], Björn Hagerman, and Tommy Bech-Kristensen

[1]Dept. of Psychology, Uppsala University, Uppsala, Sweden

## ABSTRACT

Four programs (female voice, acoustic guitar, pop music, and symphonic music) were reproduced using four different sound levels. Eighteen normal hearing subjects, familiar with high fidelity sound reproduction, listened to the reproductions in earphones and judged the sound quality on eight perceptual scales (loudness, clarity, fullness, spaciousness, brightness, nearness, softness/gentleness, absence of extraneous sounds), and an overall scale (fidelity). They also made ratings concerning an (imagined) ideal reproduction and adjustments of the preferred sound level. Increased sound level resulted in more loudness, fullness, nearness, and less softness for all programs. In the remaining scales there was an increase up to a certain sound level or an interaction with the physical properties of the programs. The difference between the ratings of the real reproductions and the imagined ideal reproductions was smallest at the preferred sound level, and this level was for most programs set higher by the subjects than by the authors.

This work was supported by The Bank of Sweden Tercentenary Foundation and the Swedish Ministry of Health and Social Affairs, Commission for Social Research.

Mailing address:
Teknisk Audiologi/KI
Box 70025
S-100 44 STOCKHOLM
Sweden

Visiting address:
Teknisk Audiologi
Lindstedtsvägen 1 & 7
Kungliga Tekniska Högskolan
Stockholm

Telephone:
Nat.: 08-11 66 60
Int.: +46 8 11 66 60

Fax:
Nat.: 08-791 71 30
Int.: +46 8 791 71 30

ISSN 0280-6819
TA123, Dec 1991

## INTRODUCTION

In a number of previous reports we studied how the perceived sound quality of various sound-reproducing systems is affected by different frequency responses and different sound levels. The methods used were direct manipulation of one or both of these factors (Gabrielsson, Schenkman, and Hagerman, 1988; Gabrielsson, Hagerman, Bech-Kristensen, and Lundberg, 1990) or post hoc analysis of frequency responses of loudspeakers (Gabrielsson, Lindström, and Till, 1991), that were used in a comprehensive listening test (Gabrielsson and Lindström, 1985). The sound quality has throughout been assessed by means of perceptual scales resulting from earlier multivariate analyses (Gabrielsson and Sjögren, 1979). The results generally show that the frequency response and the sound level affect practically all relevant perceptual dimensions. Sometimes their effects may be difficult to separate (Gabrielsson et al., 1990), and furthermore they often interact in complex ways with the physical properties of the programs used as stimuli.

In this report we focus on the effects of different sound levels. The available evidence from the above reports and from other researchers (Stevens and Davis, 1938; Kötter, 1968; Bismarck, 1974; Tannaka and Koshikawa, 1989; Ueda and Akagi, 1990) indicates that an increase of the sound level will usually increase the perceived fullness, spaciousness, nearness, and sharpness of the reproduction;

a decrease of the sound level gives the opposite results. Increased sound level may also contribute to increased clarity, although only up to a certain level. The results concerning brightness are not conclusive; effects have been reported (Stevens and Davis, 1938; Gabrielsson, Rosenberg and and Sjögren, 1974; Gabrielsson and Sjögren, 1979) as well as no effect (Gabrielsson et al, 1990; Ueda and Akagi, 1990). The effect on an overall evaluation, such as fidelity, usually reflects the effect(s) on one or more of the above perceptual dimensions (e.g., clarity) depending on how much the respective dimension contributes to the overall impression. Furthermore, the effect on any dimension may interact with the properties of the programs used as stimuli.

In order to further investigate the effect of the sound level an experiment was conducted in which the sound level was varied over a range of 30 dB and using programs widely differing in spectral composition and temporal density. Besides rating the reproductions, the listeners also made judgments concerning the properties of an ideal reproduction and made adjustments of the preferred sound level for each program.

## METHODS

Eighteen normal hearing subjects listened in earphones to four programs reproduced at four different sound levels and judged the sound quality on eight perceptual scales and a fidelity scale. Judgments were also made concerning an imagined ideal reproduction and of the preferred sound level.

**A. Programs**

Four programs, each lasting about one minute, were used:

1. Female voice, reading a list of sentences originally used for measurement of speech recognition in noise. The sentences were used without noise, and the silent intervals between them were reduced to about 1 sec. Monophonic recording in free field (Hagerman, 1982).

2. Chamber music. J.S. Bach: Sarabande from Suite in A minor, BWV 997 (original C minor; excerpt 0'00''- 1'08''), performed by John Williams on the guitar. Phonograph record: John Williams, Bach: The Four Lute Suites. CBS Records, CBS Masterworks, CB 801, MK 42204.

3. Pop music. Sting: "If you love somebody, set them free" (excerpt 0'13'' - 1'16''). Phonograph record: Sting: "The dream of the blue turtles", A & M Records, A&M 393 750 - 2.

4. Symphonic music. P. Tchaikovsky: Symphony no. 4 in F minor, op. 36, 4th movement, Finale: Allegro con fuoco (excerpt 7'23'' - 8'30''). Oslo Philharmonic Orchestra, conductor Mariss Jansons. Chandos Records, CHAN 8361.

The female voice and the guitar music have most of their energy below 2.5 kHz; their power spectra have dropped 50 dB at about 5 kHz. The pop music and the symphonic music have much broader frequency range and decrease slowly down to -50 dB at about 18 kHz. A further difference between these two groups of programs is that programs 3-4 are continuously sounding at a rather constant level, whereas there are short silent intervals (1 s)

between the sentences in program 1 and less density of sound events in program 2.

**B. Reproduction system**

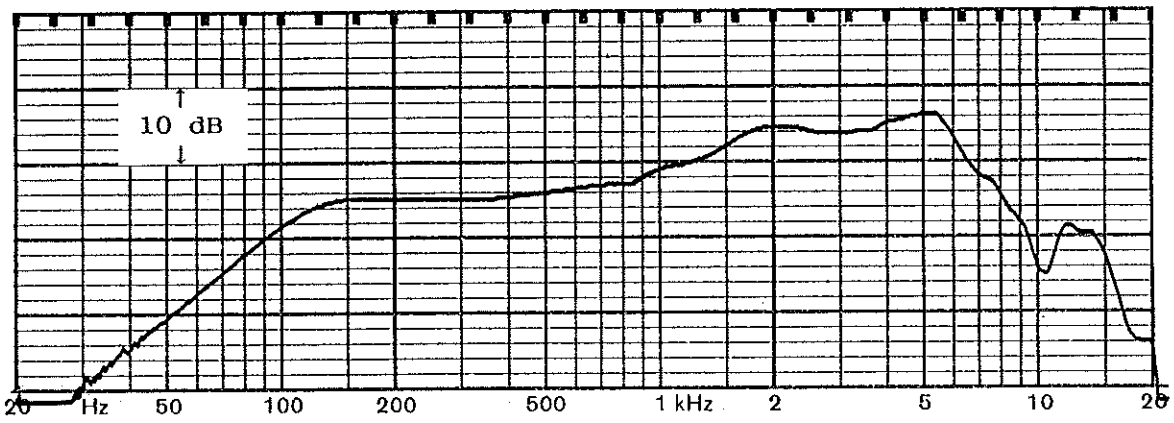
A stereo tape recorder, Revox B77 Mk II, was used to record and reproduce the programs. They were presented to the listeners in stereo through Sony Walkman MDR-E262 earphones. The frequency responses of the earphones are shown in Figure 1. Four attenuators (two Hewlett Packard 350 D and two built in our department), two for each channel, were used to get the different sound levels. The impedance matching was checked in order to assure the appropriate attenuations. The balance between the stereo channels was checked by the authors in listening to the reproductions.

The authors also set an approximately natural sound level for each program by listening to the respective program in the earphones with no filtering. The sound level for this setting was measured on a coupler according to IEC 711 fitted into the KEMAR manikin and registered by a level recorder (B & K 2305) with a writing speed of 50 mm/sec. The sound pressure level for the female voice varied between about 60 and 70 dB, for the pop music between about 80 and 90 dB, for the orchestral music between about 83 and 93 dB, and for the guitar music between about 58 and 72 dB.

Four sound levels were used for each program. For the two softer programs, that is, the female voice and the guitar music, these were the natural level, 10 dB and 20 dB above the natural level, and 10 below the natural level. For the louder programs, the pop music and the symphonic music, the four

levels were the natural level, 10 dB above the natural level, and 10 dB and 20 dB below the natural level. The highest level was thus roughly about 100 dB for the pop and symphonic music and about 90 dB for the female voice and the guitar.

Left earphone




Right earphone

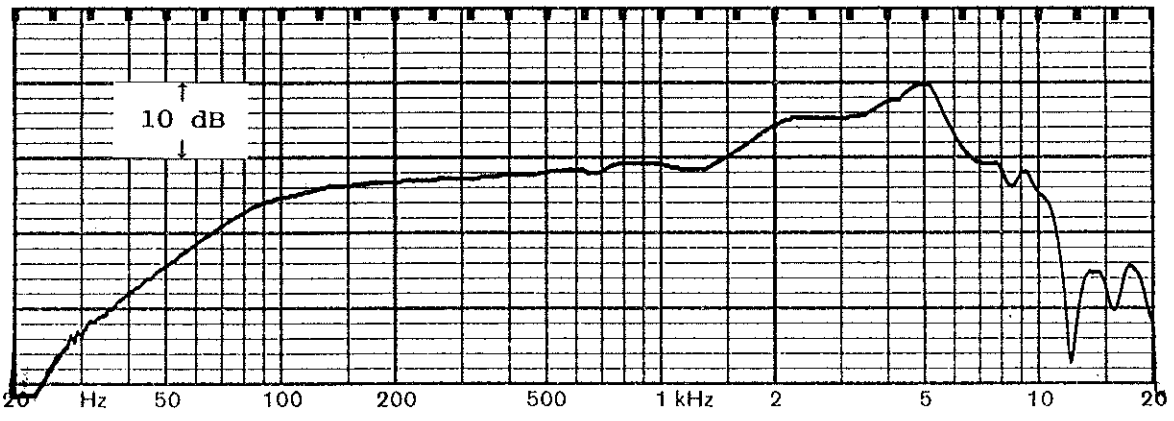



Figure 1. Frequency responses of the earphones measured on a coupler (IEC 711) fitted into a KEMAR manikin.

The listener was seated in a sound insulated chamber used for psychoacoustic experiments. All the equipment as well as the experimenter were in an adjoining room.

## C. Subjects

Eighteen male subjects, age 22-34 years, participated. They were selected on the basis of their answers to a questionnaire concerning music preferences and experience of high fidelity reproduction. About half of them preferred listening to pop and rock music or other popular music, the other half mainly to classical music; however, several listened to both popular and classical music. All were used to listen to high fidelity sound reproduction of relatively high quality.

The subjects were tested for normal hearing (threshold not exceeding 20 dB hearing loss 250 - 8000 Hz, ISO 389). None of them had any experience from this type of experiment. They were paid for their participation.

## D. Response variables

The reproductions were rated on nine scales. Eight of them refer to perceptual dimensions: loudness (Swedish: ljudstyrka), fullness (fyllighet), brightness (ljushet), softness/gentleness (mjukhet), nearness (närhet), spaciousness (rymdkänsla), clarity (tydlighet), and absence of extraneous sounds (störningsfrihet). The ninth scale required an overall evaluation in terms of fidelity (naturtrohet). All scales were graded from 10 (maximum) to 0 (minimum) and with definitions for 9, 7, 5, 3, and 1 as seen in Figure 2.

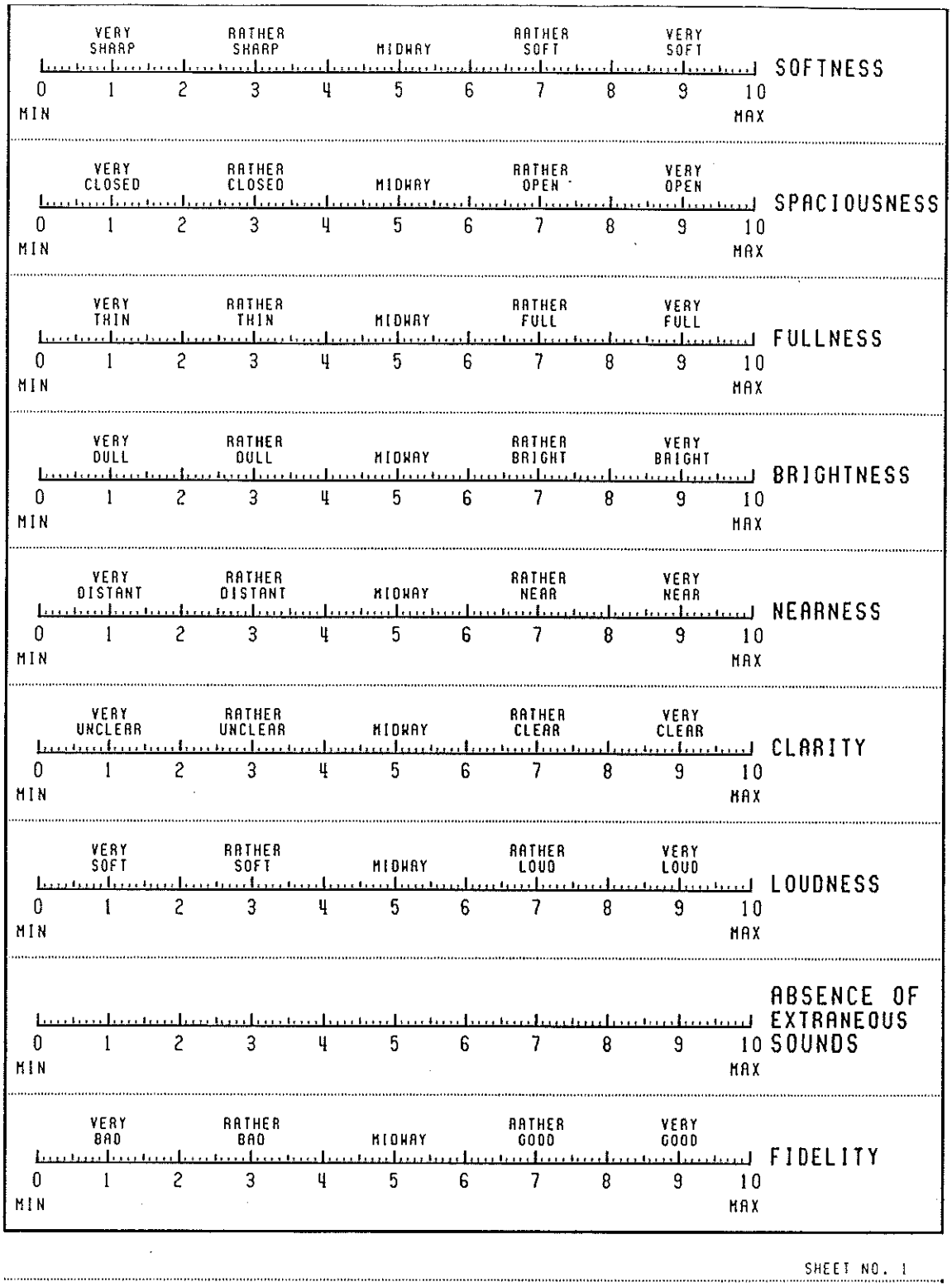



Figure 2. Example of the response form (translated from Swedish).

Decimals were included, since many subjects in earlier investigations used decimals in their ratings (Gabrielsson & Lindström, 1985; Gabrielsson et al., 1988; Gabrielsson et al., 1990). Further explanations are given in the instructions, see Appendix A.

### E. Design and procedure

The experiment comprised three parts: sound quality ratings, ratings of the "ideal" sound quality, and adjustment of the ideal sound level for each program.

In the first part the subject listened to the programs at the different sound levels and rated the sound quality. There were 16 stimuli, 4 programs x 4 sound levels. They were rated twice by each subject on all nine scales. The presentation order of the stimuli was randomized, differently for each subject. However, to avoid possible TTS (Temporary Threshold Shift) none of the loudest reproductions of any program was allowed to appear after one another. The order of the perceptual dimensions on the response form was also randomized differently for each subject; however, the fidelity scale was always last. After introducing the subject to the situation and trying out the earphones, the instructions were given (see Appendix A), followed by 12 practice trials. The main experiment included 32 trials (16 stimuli x 2 trials). A short break was made in the middle.

In the second part the subject was asked to rate how each program should sound in the different perceptual dimensions in order to represent an ideal reproduction (Appendix B). This was done

without listening to the programs. The subject had to remember the programs and imagine how an ideal reproduction of each program should sound. These ratings were made on the same scales as before except fidelity and absence of extraneous sounds, for which the ideal value is 10 by definition. Each program was rated twice.

In the third part the subject listened once more to the programs and was asked to adjust the sound level of each program to become ideal. The experimenter set an initial sound level, the subject listened and indicated by hand movements if he wanted the level to increase or decrease. This was repeated until he was satisfied. This was made twice for each program, one time with an initially too soft level, the other with an initially too loud level. The subject was told that no program would be allowed to get louder than in the first part of the experiment.

The total time required for each subject was about two hours.

### F. Data treatment

The subjects' ratings were subjected to analysis of variance, separately for each scale. This was done both for each subject (sources of variance: sound levels and programs; fixed model) and over all subjects (sources: sound levels, programs, and subjects; mixed model). These analyses provide estimates of intra- and inter-individual reliability and tests of the possible effects of the experimental variables. For general principles concerning analysis of variance and related

questions, see Winer (1971) or Kirk (1982), and for application in listening tests Gabrielsson (1979).

## RESULTS

### A. Reliability of ratings

The intra-individual reliability was studied by the "within cell mean square" (MSw) in the individual analyses of variance, that is, the estimated average variance of the two ratings made for each stimulus in each scale (MSw is the error term for the F tests in the fixed model). The smaller this variance, the better the reliability.

The mean MSw values across subjects appear in Table 1. With regard to ratings of the real reproductions, the reliability is especially high for loudness (MSw = 0.38), which is, of course, due to the apparent differences in this dimension. For the other scales the reliability is also very good with most MSw values below 1.00. These values are lower than for unselected normal hearing subjects (Gabrielsson et al., 1988, 1990) but somewhat higher than for very experienced high fidelity listeners (Gabrielsson & Lindström, 1985). For the ratings of the imagined ideal reproductions the reliabilities are still better, which also is in accordance with the results in our earlier studies.

Table 1. Mean value across subjects for MSw in ratings of real reproductions and of ideal reproductions.

| | Real | Ideal |
|---|---|---|
| Loudness | 0.38 | 0.14 |
| Clarity | 1.06 | 0.30 |
| Fullness | 1.00 | 0.27 |
| Spaciousness | 0.97 | 0.31 |
| Brightness | 0.83 | 0.17 |
| Softness | 0.94 | 0.25 |
| Nearness | 0.68 | 0.32 |
| Abs.extr.sounds | 0.91 | |
| Fidelity | 0.94 | |

The inter-individual reliability (the agreement between the subjects) was estimated by the $r_b$ index (Winer, 1971, p. 283; Gabrielsson, 1979). Its maximum value is 1.00. This index was generally high, varying for the different scales between 0.77 and 0.99.

## B. Effects of sound levels

The results from the analyses of variance are summarized in Table 2.

Table 2. Significant differences due to sound level (L), program (P), and interaction sound level x program (LxP). *** = p <.001, ** = p <.01, * = p <.05.

| Scale | L | P | LxP |
|---|---|---|---|
| Loudness | *** | *** | *** |
| Clarity | | ** | ** |
| Fullness | *** | * | |
| Spaciousness | *** | *** | *** |
| Brightness | | *** | *** |
| Softness | *** | *** | *** |
| Nearness | *** | *** | *** |
| Abs.extr.sounds | *** | *** | *** |
| Fidelity | * | *** | *** |

In all scales, except clarity and brightness, there are significant effects due to the sound levels. However, there is also a significant interaction between sound levels and programs, except in fullness. This means that the effects of the different sound levels vary with the different programs. The meaning of the effects and the interactions can be studied by means of Table 3.

### C. Loudness

Of course, the rated loudness increases with increased sound level. The differences among the levels are highly significant (Table 2). There is also a highly significant sound level x program interaction meaning that the **increase** in loudness is different for different programs. As seen in Table 3, all four programs get the same rated loudness at the lowest level used (3.1-3.3). At the following levels loudness increases faster for programs 3 and 4, especially between the two highest levels, than for programs 1 and 2. It should be remembered that programs 3 and 4 are throughout about 10 dB louder than programs 1 and 2.

### D. Clarity

In clarity there is a significant sound level x program interaction. For program 1 clarity increases up to the second level and then declines. For program 2 there is little difference among the first three levels but worse clarity at the highest level. For programs 3 and 4 clarity increases with increased sound level up to the third level and drops at the highest level, at least for program 3.

Table 3. Mean ratings across subjects. (Sound levels are numbered from 1 = lowest level to 4 = highest level.)

Loudness

| | Sound levels 1 | 2 | 3 | 4 | Mean |
|---|---|---|---|---|---|
| Program 1 | 3.3 | 4.8 | 6.0 | 8.0 | 5.5 |
| Program 2 | 3.2 | 4.3 | 5.7 | 7.6 | 5.2 |
| Program 3 | 3.1 | 4.9 | 6.2 | 8.7 | 5.7 |
| Program 4 | 3.1 | 4.6 | 6.3 | 8.8 | 5.7 |
| Mean | 3.2 | 4.6 | 6.1 | 8.3 | |

Clarity

| | Sound levels 1 | 2 | 3 | 4 | Mean |
|---|---|---|---|---|---|
| Program 1 | 5.8 | 6.2 | 5.7 | 5.4 | 5.8 |
| Program 2 | 6.8 | 6.4 | 6.9 | 5.9 | 6.5 |
| Program 3 | 5.8 | 6.2 | 6.8 | 6.1 | 6.2 |
| Program 4 | 5.3 | 5.6 | 6.2 | 6.0 | 5.8 |
| Mean | 5.9 | 6.1 | 6.4 | 5.9 | |

Fullness

| | Sound levels 1 | 2 | 3 | 4 | Mean |
|---|---|---|---|---|---|
| Program 1 | 4.7 | 5.5 | 6.0 | 6.2 | 5.6 |
| Program 2 | 5.5 | 5.7 | 6.7 | 6.9 | 6.2 |
| Program 3 | 4.3 | 5.1 | 5.9 | 6.5 | 5.4 |
| Program 4 | 4.6 | 5.2 | 5.9 | 6.4 | 5.5 |
| Mean | 4.8 | 5.4 | 6.1 | 6.5 | |

Spaciousness

| | Sound levels 1 | 2 | 3 | 4 | Mean |
|---|---|---|---|---|---|
| Program 1 | 3.6 | 4.0 | 4.0 | 3.7 | 3.8 |
| Program 2 | 5.4 | 5.4 | 6.3 | 6.1 | 5.8 |
| Program 3 | 5.1 | 6.2 | 7.0 | 6.9 | 6.3 |
| Program 4 | 5.6 | 5.9 | 6.4 | 7.0 | 6.2 |
| Mean | 4.9 | 5.4 | 5.9 | 5.9 | |

Table 3. (Continued)

Brightness

| | Sound levels 1 | 2 | 3 | 4 | Mean |
|---|---|---|---|---|---|
| Program 1 | 4.5 | 5.0 | 4.6 | 4.5 | 4.7 |
| 2 | 5.1 | 4.8 | 5.0 | 4 7 | 4.9 |
| 3 | 5.6 | 6.0 | 6.1 | 6.6 | 6.1 |
| 4 | 5.3 | 5.4 | 5.7 | 6.5 | 5.8 |
| Mean | 5.1 | 5.3 | 5.3 | 5.6 | |

Softness

| | Sound levels 1 | 2 | 3 | 4 | Mean |
|---|---|---|---|---|---|
| Program 1 | 6.7 | 6.4 | 6.1 | 5.2 | 6.1 |
| 2 | 6.7 | 6.3 | 6.5 | 5.8 | 6.3 |
| 3 | 5.5 | 5.3 | 4.8 | 3.5 | 4.8 |
| 4 | 6.1 | 5.7 | 5.1 | 3.7 | 5.1 |
| Mean | 6.2 | 5.9 | 5.6 | 4.5 | |

Nearness

| | Sound levels 1 | 2 | 3 | 4 | Mean |
|---|---|---|---|---|---|
| Program 1 | 4.9 | 6.5 | 6.9 | 7.6 | 6.5 |
| 2 | 4.7 | 5.5 | 6.3 | 7.2 | 5.9 |
| 3 | 3.4 | 4.7 | 5.8 | 7.1 | 5.2 |
| 4 | 3.1 | 4.2 | 5.2 | 7.2 | 4.9 |
| Mean | 4.0 | 5.2 | 6.1 | 7.2 | |

Abs.extr.sounds

| | Sound levels 1 | 2 | 3 | 4 | Mean |
|---|---|---|---|---|---|
| Program 1 | 7.6 | 5.9 | 3.9 | 3.2 | 5.1 |
| 2 | 8.1 | 5.4 | 3.4 | 2.5 | 4.9 |
| 3 | 7.6 | 7.5 | 7.6 | 7.6 | 7.6 |
| 4 | 7.4 | 7.6 | 7.5 | 7.5 | 7.5 |
| Mean | 7.7 | 6.6 | 5.6 | 5.2 | |

Fidelity

| | Sound levels 1 | 2 | 3 | 4 | Mean |
|---|---|---|---|---|---|
| Program 1 | 5.3 | 5.3 | 4.9 | 4.0 | 4.9 |
| 2 | 6.7 | 6.3 | 6.0 | 4.9 | 6.0 |
| 3 | 4.8 | 5.6 | 6.2 | 5.9 | 5.6 |
| 4 | 4.9 | 5.4 | 5.9 | 5.3 | 5.4 |
| Mean | 5.4 | 5.7 | 5.7 | 5.0 | |

### E. Fullness

Fullness increases significantly with higher sound levels. Although the increase is somewhat different for the different programs, there is no significant interaction.

### F. Spaciousness

Of course, spaciousness is considerably less for program 1 (speech in anechoic chamber) than for the three music programs. The significant effect of sound levels and the significant sound level x program interaction mean that spaciousness increases with increased sound level up to the next highest or highest level for the music programs (programs 2-4), whereas there is no effect for the speech program (program 1).

### G. Brightness

Programs 3 and 4 are rated as brighter than programs 1 and 2. This is related to the fact that the spectra for programs 3 and 4 extend far higher in frequency than for programs 1-2 (cf. Programs). This may also explain why brightness increases with increased sound level for programs 3 and 4, especially at the highest sound level, but not for programs 1 and 2 (significant sound level x program interaction).

### H. Softness

There is a significant decrease in softness/gentleness with increased sound level for all programs. The accompanying significant interaction means that this decrease is more pronounced for programs 3 and

4 (especially between the two highest sound levels) than for programs 1 and 2. In other words, although an increased sound level in general gives more sharpness (= less softness), this is especially obvious for the two programs that extend highest in frequency due to their high frequency components. These two programs are in fact generally perceived as sharper than programs 1-2.

## I. Nearness

Nearness increases strongly with increased sound level. At the lowest sound level programs 1 and 2 (recorded in an anechoic chamber and chamber music, respectively) sound nearer than programs 3 and 4, which were recorded in large halls. However, there is also a strong program x sound level interaction: the perceived nearness increases faster for programs 3 and 4 so that they are rated about as near as programs 1 and 2 at the highest level. This is similar to the situation for loudness.

## J. Absence of extraneous sounds

Extraneous sounds (hissing noise) are more apparent for programs 1-2, which have little energy above 2.5 kHz and less density in the time domain than programs 3-4. The latter extend into higher frequency regions and sound continuously why the noise is almost totally masked. This shows up in a strongly significant interaction. For programs 3-4 there is no effect at all of the different sound levels (ratings 7.4 - 7.6 at all levels), whereas for programs 1-2 the ratings get successively much lower as the level increases (that is, the hissing noise becomes more and more disturbing).

## K. Fidelity

The most important result is the strong sound level x program interaction. For programs 1-2 fidelity is rated best at the lowest or next lowest sound level and then drops, especially at the very highest level. However, for programs 3-4 fidelity improves with increased sound level up to the third level but not further.

## L. Ratings of ideal reproductions

Table 4 shows the results of the subjects' ratings for an imagined ideal reproduction of the respective programs.

Table 4. Mean ratings of ideal reproductions.

| | Program | | | |
|---|---|---|---|---|
| Scale | 1 | 2 | 3 | 4 |
| Loudness | 5.5 | 6.0 | 6.6 | 7.1 |
| Clarity | 8.3 | 8.0 | 7.8 | 8.0 |
| Fullness | 6.5 | 7.4 | 7.8 | 8.0 |
| Spaciousness | 5.8 | 7.0 | 7.5 | 7.9 |
| Brightness | 6.0 | 5.6 | 5.8 | 6.0 |
| Softness | 6.7 | 6.6 | 5.3 | 5.6 |
| Nearness | 6.8 | 6.8 | 6.5 | 6.2 |

The differences among the programs are statistically significant ($p < .001$) in loudness, fullness, spaciousness, and softness. According to the results in Table 4, programs 3 and 4 should be louder than programs 1 and 2, as expected. They should also have more fullness and more spaciousness, but less softness, than programs 1 and 2. The differences in clarity, brightness, and nearness were not significant. All programs should have high

clarity, somewhat above midway value in brightness, and sound rather near.

Comparing the ratings of the reproductions in Table 3 with the values for an ideal reproduction in Table 4 shows that no reproduction of any program reaches the corresponding ideal value in clarity, spaciousness, or fullness. For the remaining scales the situation varies for different programs; sometimes the reproduction at one of the levels comes rather close to the ideal value.

Absence of extraneous sounds is not included in Table 4, since it was not rated for its ideal value, which simply must be 10. No reproduction of any program comes close to this value (cf. Table 3), and the discrepancy is especially large for programs 1-2 at the higher sound levels due to more hissing noise.

Forming the absolute difference between the ratings of the real reproductions and the ratings of the ideal reproduction (including absence of extraneous sounds) and averaging over all scales gives the results in Table 5.

Table 5. Average absolute difference over all scales between real and ideal reproduction for the different programs at different levels.

| Program | Sound level 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.8 | 1.4 | 1.7 | 2.3 |
| 2 | 1.5 | 1.7 | 1.3 | 1.8 |
| 3 | 2.2 | 1.5 | 1.0 | 1.4 |
| 4 | 2.4 | 1.9 | 1.3 | 1.5 |

For program 1 the smallest difference occurs at the second level, for programs 2-4 at the third level. In other words, the best reproduction occurs at the third level for programs 2-4, and at the second level for program 1. As seen in the following, these levels also agree fairly well with the levels that the subjects considered to be the ideal levels for the respective programs.

## M. Comparison of loudness ratings

If we compare the ratings of ideal loudness for the programs in Table 4 with the corresponding ratings of perceived loudness in Table 3, we find that the rated ideal loudness for program 1 (5.5) would correspond to a sound level between the second and third sound levels (rated 4.8 and 6.0, respectively), that is, somewhat higher than the supposed natural level (= the second level, cf. Reproduction system). For program 2 the ideal loudness (6.0) would about correspond to the third sound level (rated 5.7), that is, one level higher than the supposed natural level. For program 3 the ideal loudness (6.6) would roughly correspond to the third level (rated 6.2), the supposed natural level. For program 4 the rating of the ideal

loudness was 7.1, corresponding to a level between the third (6.3) and fourth (8.8) level, that is, somewhat higher than the supposed natural level. There is thus a tendency that the subjects want a higher sound level than that set by the authors, especially regarding program 2. Similar results appear in the adjustment data below.

## N. Adjustments of ideal sound level

The results of the subjects' adjustments of the sound level to represent an ideal level for each program appear in Table 6.

Table 6. Mean attenuation (dB) across subjects re the maximum level. The values in parentheses are the attenuations set by the authors to represent an approximate natural level.

| Program | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 14.9 (20.0) | 11.7 (20.0) | 10.0 (10.0) | 7.6 (10.0) |

The differences among the programs were strongly significant [$F(3,51) = 18.66, p < .001$], despite large inter-individual differences in the adjustments of sound level. For program 1 the attenuations varied from 8.5 to 26.5 dB, for program 2 from 6 to 26.5 dB, for program 3 from 2 to 21.5 dB, and for program 4 from 1 to 23 dB. The large dispersion necessitates cautious conclusions. However, the values concerning preferred levels in Table 6 indicate that the subjects usually wanted a higher sound level, especially for program 2, than that

set by the authors. Only for program 3 there is a good correspondence.

These results are in good agreement with the results from the ratings of the ideal sound levels. For program 1 both methods (ratings and adjustments) indicate that the ideal level is about midway between the second and third level, for program 2 close to the third level, for program 3 also close to the third level, and for program 4 a level somewhat above the third level.

## 0. Sound level and fidelity

One could expect that the reproductions should be considered best at the levels that were considered as ideal. In fact the data on the smallest difference between real and ideal reproductions in Table 5 mainly agree with this expectation: For program 1 the difference was smallest at the second level, for programs 2-4 at the third level. The agreement with the fidelity ratings in Table 3 is not complete. Programs 3-4 were rated best in fidelity at the third level and program 1 at the second level (together with the first), which fits the picture. However, program 2 was rated highest in fidelity at the lowest sound level, not the third. We have no ready explanation for this discrepancy. It may be that the subjects consider the background noise from the recording so annoying at higher levels for this program (cf. the ratings in Absence of extraneous sounds, Table 3) that they consider the fidelity best at the lowest sound level (even if this perhaps sounds too soft). However, when imagining the ideal sound level the subjects probably do not think of this negative relation

between sound level and noise but rate the sound level as such independent of noise.

## DISCUSSION

The manipulation of the sound level affected the ratings in all scales, although often differently for different programs. There were thus interactions between programs and sound levels.

Clarity increased with increased sound level up to a certain level, which usually was close to the preferred level for the respective program (the second level for program 1 and the third level for programs 2-4), and then decreased. One may hypothesize that the preferred level is chosen as the level that gives the best clarity.

Fullness increased up to the highest level used for each program.

Spaciousness increased with increased sound level up to the highest or next highest level for programs 2-4. However, for program 1, that was recorded in free field, there was no significant effect.

Brightness increased with increased sound level for the two programs which extend into high frequency regions (programs 3-4) but not for programs 1-2, that have most of their energy below 2.5 kHz.

Softness decreased (= sharpness increased) with increased sound level throughout, especially for programs 3-4 with their high frequency contents and especially for the two highest levels used.

Nearness increased with increased sound level, especially for the programs extending into high frequency regions (programs 3-4).

Extraneous sounds (hissing noise) increased with increased sound level for the programs with no high frequency content (programs 1-2) and less density in the time domain. There was no effect for programs 3-4 that include high frequencies and sound continuously, both factors contributing to masking of hissing noise.

Fidelity was rated to be best at about the preferred sound level for programs 3-4 and may be for program 1, but not for program 2 that was rated best in fidelity at a lower sound level than the preferred one.

The above results agree in most cases with those in our earlier reports or reported by other researchers (cf. Introduction). An interesting fact was that brightness, for which earlier results have been inconclusive, increased with increased sound level only for the two programs with high frequency content (programs 3-4) and especially at their highest level. In a similar manner softness decreased (= sharpness increased) much at the highest level for those two programs, but less for the other two programs with little high frequency content. Increased sound level increased spaciousness only for the programs including reverberation, but not for the voice program recorded in free field.

Ratings of the "ideal" reproduction proved to be informative in providing a reference for the ratings of the real reproductions. The difference between real and ideal reproduction was smallest at

about the preferred sound level for each program. The ratings of ideal sound level and the adjustments of ideal sound level agreed and indicated that the subjects wanted a higher sound level, for most programs, than what the authors considered to be an approximately natural level.

## ACKNOWLEDGMENTS

We wish to express our gratitude to Göran Lundberg for valuable assistance in many matters.

## REFERENCES

Bismarck, G. von (1974). "Sharpness as an Attribute of the Timbre of Steady Sounds," Acustica 30, 159-172.

Gabrielsson, A. (1979). "Statistical Treatment of Data from Listening Tests on Sound-Reproducing Systems," Technical Audiology Reports No. 92 (Karolinska Institute, Stockholm).

Gabrielsson, A., Hagerman, B., Bech-Kristensen, T. and Lundberg, G. (1990). "Perceived sound quality of reproductions with different frequency responses and sound levels," Journal of the Acoustical Society of America 88, 1359-1366.

Gabrielsson, A. and Lindström, B. (1985). "Perceived Sound Quality of High-Fidelity Loudspeakers," Journal of Audio Engineering Society 33, 33-53.

Gabrielsson, A., Lindström, B., and Till, O. (1991). "Loudspeaker Frequency Response and Perceived Sound Quality," Journal of the Acoustical Society of America 90 (2), 707-719.

Gabrielsson, A., Rosenberg, U., and Sjögren, H. (1974). "Judgments and Dimension Analyses of Perceived Sound Quality of Sound-Reproducing Systems," Journal of the Acoustical Society of America 55, 854-861.

Gabrielsson, A., Schenkman, B.N., and Hagerman, B. (1988). "The Effects of Different Frequency Responses on Sound Quality Judgments and Speech Intelligibility," Journal of Speech and Hearing Research 31, 166-177.

Gabrielsson, A., and Sjögren, H. (1979). "Perceived Sound Quality of Sound-Reproducing Systems," Journal of the Acoustical Society of America 65, 1019-1033.

Hagerman, B. (1982). "Sentences for testing speech intelligibility in noise," Scandinavian Audiology 11, 79-87.

Kirk, R.E. (1982). Experimental Design: Procedures for the Behavioral Sciences (Brooks/Cole, Belmont, California), 2nd ed.

Kötter, E. (1968). Der Einfluss Übertragungstechnischer Faktoren auf das Musikhören (Arno Volk, Köln).

Stevens, S.S., and Davis, H. (1938). Hearing: Its Psychology and Physiology (Wiley, New York).

Tannaka, Y., and Koshikawa, T. (1989). "Correlations between Sound Field Characteristics and Subjective Ratings on Reproduced Music Sound Quality," Journal of the Acoustical Society of America 86, 603-620.

Ueda, K., and Akagi, M. (1990). "Sharpness and Amplitude Envelopes of Broadband Noise," Journal of the Acoustical Society of America 87, 814-819.

Winer, B.J. (1971) Statistical Principles in Experimental Design (McGraw-Hill, New York), 2nd ed.

## APPENDIX A

### A. Instructions for sound quality ratings

You are going to listen to some programs with speech and music through earphones. We will change the reproduction of the programs in various ways.

Your task is to judge the sound quality of the different reproductions by means of the scales that you find on the response form.

The scales refer to various properties of the sound reproduction. They are all graded from 10 (maximum) to 0 (minimum).

For instance, in the scale for fullness 10 means maximum (highest possible) fullness, 9 = very full, 7 = rather full, 5 = midway, 3 = rather thin, 1 = very thin, and 0 means minimum fullness. The other scales work in similar ways. As you can see on the response form, it is possible to use decimals if you like.

You shall judge each reproduction separately without thinking absout the previous reproductions. There is a new response form for each reproduction.

Mark your judgment on each scale by a straight and distinct vertical line. Do your ratings on the scales in the order they appear on the response form. Do not look at earlier response forms.

Note that your judgments shall refer to the sound quality, not to what you think about the music or speech as such.

There are no right or wrong answers. It is solely your own opinion about the sound quality that should be decisive.

The scales are defined as follows:

Clarity: The reproduction sounds clear, distinct, and pure. The opposite is that the sound is diffuse, blurred, thick, and the like.

Fullness: The reproduction sounds full, in opposition to thin.

Spaciousness: The reproduction sounds open and spacious, in opposition to closed and shut up.

Brightness: The reproduction sounds bright, in opposition to dull and dark.

Softness/Gentleness: The reproduction sounds soft and gentle, in opposition to sharp, hard, keen, and shrill.

Nearness: The sound seems to be close to you, in opposition to at a distance.

Loudness: The sound is loud, in opposition to soft (faint).

Absence of extraneous sounds: The reproduction has no extraneous sounds like noise, hum, or the like.

Fidelity: Judge how similar the reproduction is to the original sound. 10 = perfect fidelity, 9 = very good, 7 = rather good, and so on.

First you will have several trials for practice. The examples used in these trials are representative for the reproductions that appear in the experiment.

## APPENDIX B

### A. Instructions for ratings of ideal reproduction

You have now heard the different progams several times and rated them on different scales. Now you are going to rate how each of the programs should sound in order to represent an ideal reproduction. Without listening to them again, try to imagine how each program should sound in clarity, fullness, spaciousness, brightness, loudness, nearness, and softness/gentleness to make the reproduction ideal. You use the same response form as earlier, one for each program.

You do not have to make judgments concerning absence of extraneous sounds and fidelity. In these cases the ideal is 10 by definition.

# LIST OF REPORTS FROM THE DEPARTMENT OF TECHNICAL AUDIOLOGY, KAROLINSKA INSTITUTET

TA123 1991 Perceived sound quality of reproductions with different sound levels.
Alf Gabrielsson, Björn Hagerman and Tommy Bech-Kristensen.

TA122 1991 Measurements with various loads on three sizes of zinc-air batteries for hearing aids.
Björn Hagerman.

TA121 1991 A renovated anechoic room:
Some aspects of requirements and measurements.
Ove Till and Björn Hagerman

TA120 1990 Perceived sound quality in hearing aid with vented and closed earmold equalized in frequency response.
Göran Lundberg, Anders Ovegård, Björn Hagerman, Alf Gabrielsson, and Ulrika Brändström

TA119 1990 Repeated measurements on hearing protectors on acoustic test fixture.
Eva Svensson

TA118 1990 Hearing-aid to ear impedance matching: A litterature survey.
Per-Eric Sandborn

TA117 1988T Perceived sound quality for different frequencyresponses and sound levels.
A Gabrielsson, B. Hagerman, T. Bech-Kristensen and G. Lundberg

TA116 1987 Influence of nonlinear distortion on speech intelligibility: Hearing impaired listeners.
A-C Lindblad

TA115 1987 Loudspeaker frequency response and perceived sound quality: Comparison between measurements in listening room, anechoic room and reverberation room.
A Gabrielsson, Björn Lindström and Ove Till

TA114 1986 Loudspeaker frequency response and perceived sound quality: Measurements in listening room.
A Gabrielsson, B Lindström and O Till

TA113 1986 Relations between speech intelligibility and psychoacoustical modulation transfer function (PMTF).
B Hagerman, Å Olofsson and A-C Lindblad

*TA112 1985 The effect of different frequency responses on sound quality judgments and speech intelligibility.
A Gabrielsson, B N Schenkman and B Hagerman

TA111 1985 Hearing aid measurements on occluded-ear simulator compared to simulated in-situ and in-situ measurements.
U Olsson

TA110 1984 Improvement of induction loop field characteristics using multi-loop systems with uncorrelated currents.
Å Olofsson

TA109 1984 Questionnaires on desirable properties of hearing aids.
B Hagerman and A Gabrielsson

TA108 1983 Audo versus audio-visual test - implications for hearing aid frequency responses.
A-C Lindblad, M P Haggard and J R Foster

TA107 1983 Clinical measurements of speech reception threshold in noise.
B Hagerman

TA106 1983 Assessment of perceived sound quality in eighteen High Fidelity loudspeakers.
A Gabrielsson, B Lindström and G Elger

TA105 1982 Detection of nonlinear distortion of speech signals by hearing impaired listeners.
A-C Lindblad.

*TA104 1982 Measurement of speech reception threshold - a comparison between two methods.
B Hagerman.

TA103 1981 Problems and methods in judgments of perceived sound quality.
A Gabrielsson.

TA102 1981 Scaling of perceptual dimensions in sound reproduction.
A Gabrielsson, B Lindström.

TA101 1981 Sentences for testing speech intelligibility innoise.
B Hagerman.

TA100 1981 Modulation transfer function and speech transmission index - a method to predict speech intelligibility in Swedish? - The Swedish contribution to the IEC 29B/WG15 multilanguage test.
B Hagerman, A-C Lindblad.

TA99 1980 En metod för subjektiv utvärdering av "kompaktanläggningar".
B Lindström.

* indicates that the TA-report is out of stock, but is available at a cost of SEK 35.

| | | |
|---|---|---|
| TA98 | 1980 | Clinical assessment of perceived sound quality in hearing aids.<br>A Gabrielsson, B Hagerman, C Berg, A Ovegård, L Ånggård. |
| TA97 | 1980 | Statistical analysis of fricative sounds.<br>Å Olofsson. |
| TA96 | 1980 | A method for suppression of periodic noise signals.<br>S-Å Frykholm, Å Olofsson. |
| *TA95 | 1980 | Hörapparaters frekvenskurvor: IEC-coupler, KEMAR, människa.<br>U Olsson. |
| TA94 | 1980 | Some ideas in computerized acoustical impedance measurements.<br>S-Å Frykholm. |
| TA93 | 1979 | Assessment of perceived sound quality in high fidelity sound-reproducing systems.<br>A Gabrielsson, S-Å Frykholm, B Lindström. |
| *TA92 | 1978 | Statistical treatment of data from listening tests on sound-reproducing systems.<br>A Gabrielsson. |
| TA91 | 1978 | Taluppfattbarhet för högtalarsystem vid Arlanda utrikesterminal.<br>B Hagerman, A-C Lindblad. |
| TA90 | 1978 | A generalization of the concept of total harmonic distortion.<br>Å Olofsson. |
| *TA89 | 1978 | Ett programpaket för mätning, syntes och analys av linjära tidskontinuerliga och tidsdiskreta system.<br>Å Olofsson. |
| TA88 | 1978 | Reliability in the determination of speech reception threshold (SRT).<br>B Hagerman. |
| TA87 | 1978 | Detection of nonlinear distortion in telephone systems.<br>A Gabrielsson, A-C Lindblad, B Lindström, Å Olofsson. |
| *TA86 | 1977 | Adjective ratings and dimension analyses of perceived sound quality of headphones.<br>A Gabrielsson, S-Å Frykholm, H Sjögren. |
| *TA85 | 1977 | Adjective ratings and dimension analyses of perceived sound quality of hearing aids. III.<br>A Gabrielsson, H Sjögren. |
| *TA84 | 1977 | Automatic gain control and hearing aids. The influence of different attack and release times on speech intelligibility for hearing impaired with recruitment.<br>T Ahren, S Arlinger, C Holmgren, L Jerlvall, B. Johansson, A-C Lindblad, H Sjögren. |
| *TA83 | 1976 | Detection of amplitude distortion by normal hearing and hearing impaired subjects.<br>A Gabrielsson, P-O Nyberg, H Sjögren, L Svensson. |
| *TA82 | 1976 | Preferred listening levels and perceived sound quality at different sound levels in "high fidelity" sound reproduction.<br>A Gabrielsson, H Sjögren. |
| *TA81 | 1976 | Preliminärrapport slingor.<br>R Ingelstam. |
| TA80 | 1975 | Emulator för kommunikation mellan minidator typ Honeywell 316 och stordator typ IBM 360.<br>Bo Jacobsson. |
| TA79 | 1975 | Mätning av amplitud- och faskurvor för hörapparater med hjälp av minidator.<br>Å Olofsson. |
| TA78 | 1975 | Styr- och registreringsenhet för psykoakustisk mätplats.<br>Å Olofsson, L Persson, H Sjögren. |
| *TA77 | 1975 | Adjective ratings and dimension analyses of perceived sound quality of hearing aids. II.<br>A Gabrielsson, H Sjögren. |
| *TA76 | 1975 | Similarity ratings and dimension analyses of perceived sound quality of hearing aids.<br>A Gabrielsson, H Sjögren. |
| *TA75 | 1974 | Adjective ratings and dimension analyses of perceived sound quality of hearing aids. I.<br>A Gabrielsson, H Sjögren. |
| TA74 | 1974 | Assessment of comfort and discomfort levels for pure tone, a methodological study.<br>A Gabrielsson, B Johansson, Bodil Johansson, A-C Lindblad, L Persson. |
| TA73 | 1973 | Hörapparaternas driftkostnader.<br>B Johansson. |
| TA72 | 1973 | Adjective ratings and dimension analyses of perceived sound quality of sound reproducing systems.<br>A Gabrielsson, U Rosenberg, H Sjögren. |
| TA71A | 1974 | Mätsond för mätning på magnetiska fält inom tonfrekvensområdet.<br>R Ingelstam. |

* indicates that the TA-report is out of stock, but is available at a cost of SEK 35.