Perceived sound quality of reproductions with different frequency responses and sound levels

Alf Gabrielsson
Department of Technical Audiology, Karolinska Institute, Stockholm, and Department of Psychology, Uppsala University, Uppsala, Sweden

Björn Hagerman, Tommy Bech-Kristensen, and Göran Lundberg
Department of Technical Audiology, Karolinska Institute, KTH, S-100 44 Stockholm, Sweden

(Received 27 February 1989; accepted for publication 6 April 1990)

Three programs (female voice, jazz music, and pink noise) were reproduced using four different frequency responses and two different sound levels. Fourteen normal hearing subjects listened to the reproductions via earphones and judged the sound quality on seven perceptual scales (loudness, clarity, fullness, spaciousness, brightness, softness/gentleness, and nearness) and a fidelity scale. Significant differences among the reproductions appeared in all scales and could be attributed to the differences in frequency response or sound level or both. Interactions between the reproductions and the programs could be explained by the relations between the spectrum of the programs and the frequency responses used. The results for the noise program were similar to those for the jazz music program.

PACS numbers: 43.66.Lj, 43.88.Md [NFV]

INTRODUCTION

The perceived sound quality of sound-reproducing systems—such as loudspeakers, headphones, and hearing aids—is multidimensional; that is, it is associated with a number of perceptual dimensions. By means of multivariate methods used in experimental psychology, we were able to show that the perceived sound quality can be described in terms of dimensions such as clarity, fullness, brightness versus dullness, sharpness versus softness/gentleness, loudness, spaciousness, nearness, and absence of extraneous sounds (Gabrielsson, 1979a; Gabrielsson and Sjögren, 1979a). Rating scales for these dimensions have been successfully used to provide perceptual descriptions of loudspeakers and other sound-reproducing systems (Gabrielsson and Lindström, 1985; Gabrielsson, 1987; Gabrielsson *et al.*, 1988). Overall evaluations of the systems in terms of fidelity or pleasantness may be regarded as weighted combinations of the separate perceptual dimensions. The weight given to each dimension—that is, how important it is for the overall evaluation—depends on the character of the program to be reproduced and the listener's earlier experiences.

The relations between the perceptual dimensions and various physical properties of the systems are complex and still largely unknown. Among the physical variables the frequency response is often considered as the most important. Its effects on the perceptual dimensions have been explored in a *post hoc* manner by studying the frequency response of the systems receiving different ratings in the respective dimensions and also more directly by experimental manipulation of the frequency response (Gabrielsson *et al.*, 1974; Gabrielsson and Sjögren, 1979a,b; Gabrielsson *et al.*, 1986, 1987, 1988; cf. also similar approaches by Komamura *et al.*, 1977; Kötter, 1968; Staffeldt, 1974; Toole, 1986; Toole and Olive, 1988). The results of our investigations indicate that the frequency response can affect any of the above-mentioned perceptual dimensions. Brightness and sharpness increase (dullness and softness/gentleness decrease) with rising frequency response toward higher frequencies and/or falling response toward lower frequencies (cf. also Stevens and Davis, 1938, pp. 163–166, for density and brightness in sinusoidal tones; Bismarck, 1974). Fullness is favored by a broad frequency range and relatively more emphasis on lower frequencies (cf. Stevens and Davis, 1938, p. 161, for volume in sinusoidal tones). Clarity, spaciousness, and (to some extent) nearness are likewise favored by a broad frequency range, often with a certain emphasis on midhigh to high frequencies. The effects of extraneous sounds, e.g., hiss, may be relieved by reduced response at high frequencies. Generally, the results also depend on the characteristics of the (music or speech) programs that are reproduced. There are thus interactions between the reproduction systems and the programs.

Another important physical factor is the sound level. The available evidence (e.g., Gabrielsson and Sjögren, 1979a) indicates that an increase in sound level will usually increase the perceived fullness, spaciousness, and nearness as well as sharpness and brightness; a decrease in sound level gives the opposite results. Increased sound level may also contribute to increased clarity, although only up to a certain level at which overloading in the auditory system may occur (Kryter, 1970, p. 52). There may be interactions between the sound level, on the one hand, and the frequency response and/or the spectrum of the program on the other hand. For instance, a program reproduced by a system with boosted treble may sound even sharper and brighter if the sound level is increased, while a reproduction with boosted bass will probably sound even duller if the sound level is raised.

Because of such complex interactions and also because of the *post hoc* character of certain results referred to above, a further experiment with systematic manipulation of the frequency response and the sound level was conducted. The purpose was to investigate the effects on the perceptual dimensions of four markedly different frequency responses (flat, boosted at low, midhigh, and high frequencies) and two different sound levels in the reproduction of three programs including speech, music, and pink noise.

I. METHOD

A. Programs

Three programs were used: (1) pink noise (− 3 dB/oct), monophonic recording; (2) female voice reading a fairy-tale in an anechoic chamber, monophonic recording; and (3) jazz music, excerpt from "Ole Miss" by W. C. Handy, performed by The Peoria Jazz Band in an auditorium. Phonograph record: Opus 3, 79-00, Testskiva 1: Perspektiv. The excerpt was copied directly from the stereophonic recording on the master tape, but was played monophonically to the listener.

Each program lasted for about 1 min. The pink noise was chosen to serve as a neutral reference. The female voice in the anechoic chamber has most of its energy below 1 kHz, especially between about 130 and 700 Hz, while the jazz music has a considerably broader frequency range with a boost around 100 Hz (see Fig. 1). The programs were low-passed at 6.7 kHz for reasons explained below.

B. Reproduction system

A tape recorder (Telefunken Magnetophon 28) was used to reproduce the programs, which were then filtered before binaural (diotic) presentation to the listeners through Sony Walkman MDR-E262 earphones. The frequency response of the earphones is shown in Fig. 2. The steep cutoff at about 6 kHz is due to the antialiasing low-pass filter described below.

The filters were implemented as digital FIR filters using a general purpose measuring device (TAMP3) developed in our department. It can be used for measuring the frequency response of linear, time-invariant systems and also for digital filtering of signals in real time. It is equipped with antialiasing filters and fast 12-bit A/D and D/A converters. Various types of input amplifiers, output amplifiers, and attenuators can be connected to the processor. All units are controlled by a personal computer.

The sampling frequency of the digital filter had to be restricted to 20 kHz, and an antialiasing low-pass filter was set to 6.7 kHz. Four different filters were implemented. One filter had a flat response, that is, no filtering at all. The other three filters gave about 9 dB amplification below 200 Hz, around 1 kHz, and around 4 kHz (see Fig. 3). (The lowest filter could not be made symmetrical because of certain limitations in the equipment. Below 100 Hz there is little energy due to the cutoff of the earphones; cf. Fig. 2.) These filters will be referred to as the L (for low), M (midhigh), and H (high) filter, respectively.

The sound levels were set to represent an approximately natural level of the respective program when listened to in the earphones through the filter with flat response. Measured by a coupler according to IEC 711 fitted into the KEMAR manikin, the A-weighted sound level for the pink noise with the flat response was about 68 dB, for the female voice about 56 dB, and for the jazz music about 80 dB. For comparison each program was also presented at a 10 dB lower level.

The filters themselves caused certain changes in the sound level. These effects were different for different programs depending on their spectrum (Fig. 1). For the pink noise there was practically no difference in the A-weighted sound level between the flat response and the L filter, while the M filter gave an increase of 2 dB, and the H filter an increase of 6 dB. For the jazz music there was again no difference in the A-weighted sound level between the flat response and the L filter, while the M filter gave an increase of about 5 dB and the H filter an increase of about 3 dB. For the female voice the L filter increased the A-weighted sound level by

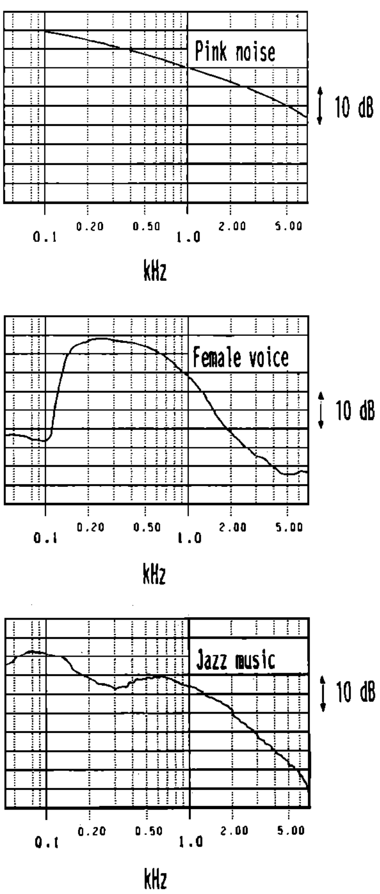

FIG. 1. Long-time average spectrum (LTAS) for the three programs. LTAS was calculated from the autocorrelation function of the program and smoothed across octaves.

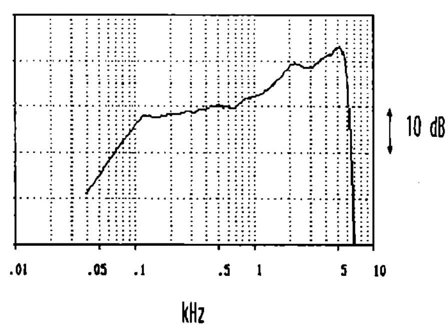

FIG. 2. Frequency response of the earphone measured on a manikin (KEMAR) equipped with an ear simulator according to IEC 711.

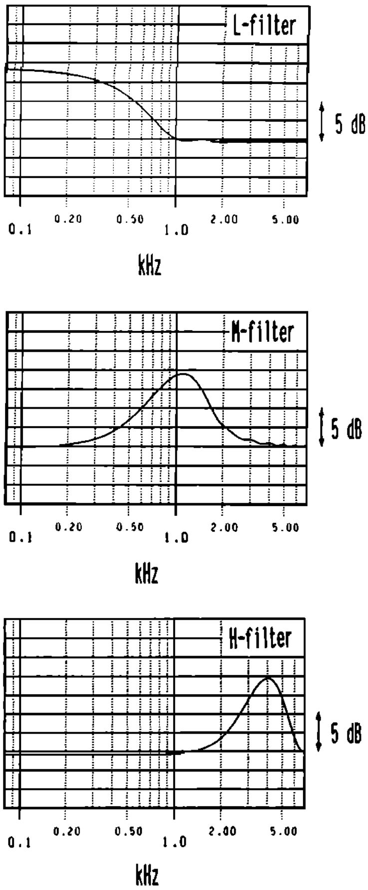

FIG. 3. Frequency responses of the three filters.

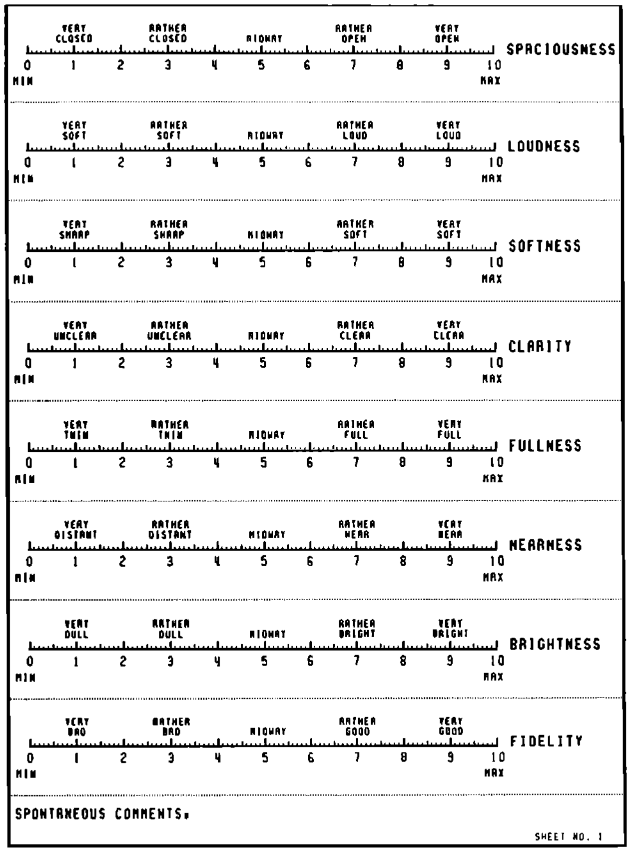

FIG. 4. Example of the response form (translated from Swedish).

about 3 dB, the M filter about 2 dB, and the H filter about 1 dB. These effects have to be considered in the interpretation of the results.

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

C. Subjects

Fourteen subjects, seven males and seven females, age 22–34 years, participated. None of them had any experience in this type of experiment. All of them were tested for normal hearing (less than 20-dB hearing loss, 250–8000 Hz, ISO 389).

D. Response variables

The reproductions were rated on eight scales. Seven of them refer to perceptual dimensions: loudness (Swedish: ljudstyrka), fullness (fyllighet), brightness (ljushet), softness/gentleness (mjukhet), nearness (närhet), spaciousness (rymdkänsla), and clarity (tydlighet). The eighth scale required an overall evaluation of each reproduction in terms of its fidelity. All scales were graded from 10 (maximum) to 0 (minimum) and with definitions for 9, 7, 5, 3, and 1 as seen in Fig. 4. Decimals were included, since many subjects in earlier investigations used decimals in their ratings (Gabrielsson and Lindström, 1985; Gabrielsson *et al.*, 1988). Definitions and further explanations were given in the instructions (see the Appendix).

E. Design and procedure

There were 24 conditions (three programs × four filters × two sound levels), and they were rated twice by each subject on all eight scales (however, the noise program was rated on seven scales omitting the fidelity scale). The presentation order of the stimuli was randomized differently for each subject. The order of the rating scales on the response form (Fig. 4) was also randomized differently for each subject. After introducing the subject to the situation and trying out the earphones, the instructions (Appendix) were given, followed by 12 practice trials. Then the main experiment including 48 trials (24 stimuli × 2 trials per stimulus) was conducted with a break in the middle. After that, the subject answered some questions related to the experiment.

The position of small earphones is critical for the resulting frequency response at the eardrum. In order to check that the earphones were placed in a similar way for the two parts of the main experiment (before and after the break), a broadband signal was used, and the resulting response was measured by a Diaphon probe microphone inserted into the ear canal behind the earphone. The microphone response was fed into TAMP3 for analysis of the frequency response. On the whole, these frequency responses were fairly similar in both parts of the experiment. Between 200 and 3000 Hz, the differences were usually less than 2 dB. A dip was often found around 150 Hz, somewhat varying in position and size. Differences larger than 5 dB were found around 4–5 kHz in a few cases. The variance of the repeated ratings for each stimulus (cf. Sec. II below) was of the same order as for a group of corresponding subjects in an earlier experiment with supra-aural headphones (Gabrielsson *et al.*, 1988).

The total time required for each subject was about 2.5 h. Beside the necessary time for the instructions, the practice trials, and the break, much time was spent in fitting the probe microphone and the earphone. The actual listening time was about 50 min.

F. Data treatment

The subjects' ratings were subjected to analysis of variance, separately for each scale. This was done both for each individual subject (sources of variance: filters, sound levels, and programs; fixed model) and over all subjects (sources: filters, sound levels, programs, and subjects; mixed model). One-tailed t tests were used to test specific hypotheses, derived from our earlier investigations, concerning the effects of different filters. For general principles concerning analysis of variance and related questions, see Winer (1971) or Kirk (1982). For application in listening tests, see Gabrielsson (1979b).

II. RESULTS

A. Reliability of ratings

The *intraindividual reliability* was studied by means of 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 is the reliability. The median value for MSw over all 14 subjects for each scale appears in Table I. (The median was chosen rather than the arithmetic mean because of one extremely deviating subject.) MSw is clearly lowest for the loudness scale (0.53), which is the most familiar dimension. For the other scales MSw varies between 1.27 and 1.64. These values are about the same as for another group of unselected subjects with normal hearing described in Gabrielsson *et al.* (1988), but are higher (that is, the reliability is worse) than for subjects selected for experience of listening to high-fidelity sound reproduction (Gabrielsson and Lindström, 1985). Another indication of good reliability is the occurrence of significant F tests (at 0.05 level or lower) for the different experimental variables. Out of the 14 subjects, typically at least half of them had significant differences among the various filtered reproductions in each scale.

The *interindividual reliability* (the agreement between the subjects) was estimated by means of the r_b index (Winer, 1971, p. 283; Gabrielsson, 1979b). Its maximum value is 1.00; the higher the value, the better the reliability. As seen in Table I, the reliability is again highest for loudness (0.98), but is generally high (0.84–0.95) for the other scales as well.

TABLE I. Median value across subjects for MSw and value of the r_b index for each rating scale.

	MSw	r_b
Loudness	0.53	0.98
Clarity	1.40	0.91
Fullness	1.37	0.86
Spaciousness	1.64	0.91
Brightness	1.27	0.92
Softness	1.27	0.95
Nearness	1.48	0.93
Fidelity	1.29	0.84

B. Effects of sound levels

The effects of the different sound levels can be seen in Fig. 5, which shows the average ratings across all subjects in each scale for the different reproductions. In loudness the ratings at the high sound level are 1.5–3 units higher than for the corresponding cases at the low level. The difference is highly significant [$F(1,13) = 140$, $p < 0.001$]. The high sound level also gives better clarity [$F(1,13) = 19.8$, $p < 0.001$], more fullness [$F(1,13) = 17.0$, $p < 0.01$], more spaciousness [$F(1,13) = 35.2$, $p < 0.001$], more nearness [$F(1,13) = 95$, $p < 0.001$], and better fidelity [$F(1,13) = 7.6$, $p < 0.025$], but less softness/gentleness [$F(1,13) = 47.2$, $p < 0.001$], than the low sound level. In brightness there was no statistically significant difference between the two levels.

There is also a significant difference in rated loudness among the programs [$F(2,26) = 9.9$, $p < 0.001$], and a significant program×level interaction [$F(2,26) = 18$, $p < 0.001$]. The meaning of these results is clear from Fig. 5. At the high level the voice is rated lower in loudness than the other programs, while there is practically no difference among the programs at the lower level. The perceived loudness is thus reduced more for the noise and jazz programs than for the voice, when the sound level is lowered. The fact that the noise is rated almost as high in loudness as the jazz music, although there is a considerable difference between their sound levels, is probably due to the continuous and "irritating" character of the noise.

The effects of the different programs on the other perceptual scales are usually self-evident and will not be further commented.

C. Effects of filters

The effects of the different filters are discussed separately for each scale.

1. Loudness

There is a significant difference among the various filters [$F(3,39) = 20$, $p < 0.001$] and a significant filter×level interaction [$F(3,39) = 6.2$, $p < 0.01$]. As seen in Fig. 5 for the high level, the L, M, and H filters give an increase in loudness in comparison with the flat response, which may be expected since these filters usually also produce an increase of the overall sound level (cf. Sec. I B). An exception is the H filter for the voice program; but since the voice has most of its energy below 1000 Hz, it is not much affected by the H filter (cf. Figs. 1 and 3). At the low level there are similar but less-pronounced tendencies.

The fact that the filters usually give an increased overall sound level in comparison with the flat response has to be considered when interpretating the results in the remaining scales. If the results for the different filters are in the same direction as could be expected for a higher sound level, the

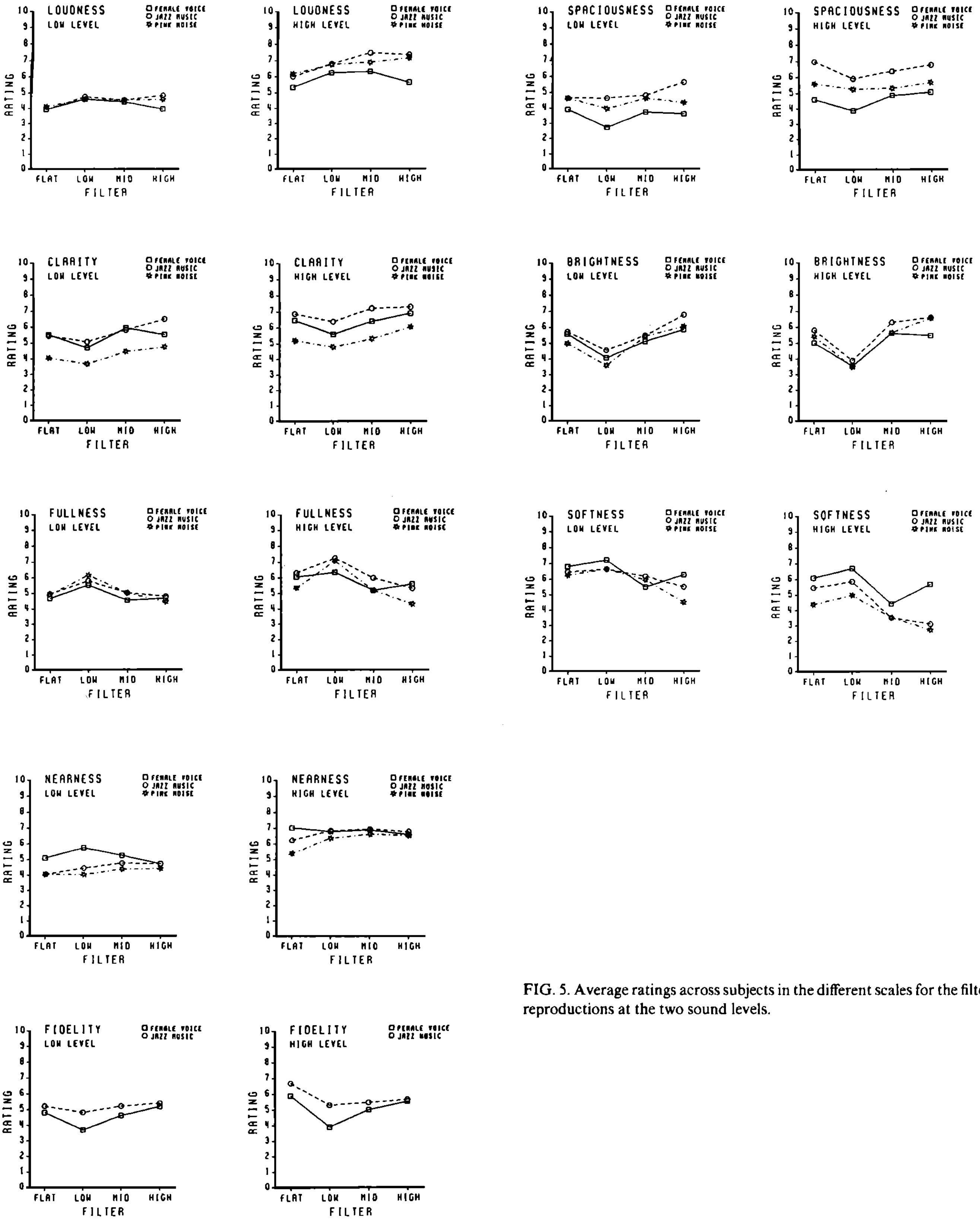

FIG. 5. Average ratings across subjects in the different scales for the filtered reproductions at the two sound levels.

situation is ambiguous: The effect may be due to the respective frequency response, or to increased sound level, or to a combination of both. However, if the results go in the opposite direction, or if a difference in sound level has no expected effect, as for brightness, the effect can be attributed to the respective frequency response.

2. Clarity

The average rating for the L filter across all programs and both sound levels is 5.0, which is significantly lower than the corresponding value for the flat response, 5.6 [$t(39) = 2.4, p < 0.025$]. Further the average rating for the

H filter (6.2) is higher than for the flat response [$t(39) = 2.4, p < 0.025$]. The difference between the M filter and the flat response is not significant. The better clarity for the H filter may be due to its frequency response or to an increase of the overall sound level, or to both. However, the worse clarity for the L filter cannot be attributed to an increased overall sound level (this would give better clarity), but depends on the increased response at low frequencies.

3. Fullness

The L filter gives more fullness than the flat response [$t(39) = 2.65, p < 0.01$]. There is no difference between the M filter and the flat response. There is a tendency, just short of statistical significance, for the H filter to give less fullness than the flat response at the higher sound level, especially for the noise and jazz programs. The last-mentioned effect cannot be due to an increased overall sound level associated with the H filter, since this would increase fullness. The reason is probably that the H filter makes the lower frequencies, which contribute most to fullness, less important. For the voice program the effect of the H filter is small, since most of its energy lies below the range of the H filter. The increased fullness for the L filter may depend on its frequency response or on an increase of the overall sound level, or on both. However, since the increase in overall sound level is larger with the H and M filters than for the L filter (cf. Sec. I B), and since there is no difference or even a decrease in fullness for the H and M filters, the increased fullness for the L filter must be an effect of its frequency response.

4. Spaciousness

Spaciousness is significantly lower for the L filter than for the flat response across programs and sound levels [$t(39) = 1.72, p < 0.05$]. (However, there is no difference for the jazz program at the lower sound level.) This effect is due to the frequency response of the L filter, since a possibly increased overall level would lead to more spaciousness.

5. Brightness

Since there was no difference in brightness between the two sound levels, any difference between the filters is due to different frequency responses. The L filter reduces brightness throughout, and the difference between the flat response and the L filter across programs and sound levels is highly significant [$t(39) = 6.04, p < 0.0005$]. The H filter increases brightness, and the difference between the H filter and the flat response is also highly significant [$t(39) = 3.04, p < 0.005$]. There is no significant difference between the M filter and the flat response, but there is a tendency toward higher brightness for the M filter at the higher sound level.

The difference between the H filter and the flat response is much smaller for the voice program than for the other programs. The reason is that the H filter does not influence the voice program very much, since its effect lies mainly outside the spectrum of the voice program.

6. Softness/gentleness

Although the difference between the L filter and the flat response does not quite reach statistical significance [$t(39) = 1.08, p < 0.15$], the data in Fig. 5 indicate an increase in softness for the L filter, which then must be due to its frequency response (since a possible increase in sound level would reduce softness). The M filter gives less softness (= more sharpness) than the flat response [$t(39) = 2.74, p < 0.005$], and so does the H filter [$t(39) = 3.33, p < 0.005$]. These effects may be due to the frequency responses of the filters or to the accompanying increase in overall sound level, or to both.

There is a significant level$\times$filter interaction [$F(3,39) = 5.2, p < 0.01$], meaning that the decrease of softness with the M and H filters is more pronounced at the higher level than at the lower (see Fig. 5). In our own experience, the noise and the jazz music sound sharp and irritating, when reproduced by the H filter at the high level. There is also a significant filter$\times$program interaction [$F(6,78) = 6.5, p < 0.001$]. While the noise and jazz music sounds sharpest with the H filter, the voice program sounds sharpest with the M filter (see Fig. 5). The H filter cannot contribute very much to sharpness in the voice program, since it only affects frequencies above the main part of the voice spectrum.

7. Nearness

The noise and jazz programs tend to sound nearer for all three filters, especially at the higher level, whereas this does not hold for the voice program. The difference between the M filter and the flat response across programs and levels is close to significance [$t(39) = 1.64, p < 0.10$]. This may be due to its frequency response or the concomitant increase of the sound level or both.

8. Fidelity

As seen in Fig. 5, the fidelity is consistently worst for the L filter, and the difference between the flat response and the L filter is highly significant [$t(39) = 3.84, p < 0.0005$]. However, there is also an interaction between sound levels and filters [$F(3,39) = 4.7, p < 0.01$]. The differences among the filters are much more evident at the higher, natural sound level than at the lower level. At the lower level the only clear result is that the L filter is worse than the others. However, at the higher level the L, M, and H filters are all worse than the flat response. Since fidelity thus becomes worse with those filters, despite their accompanying increase of the overall sound level, these results are due to the frequency responses of the respective filters.

Earlier results indicate that a slight or moderate emphasis on midhigh to high frequencies is favorable to fidelity (Gabrielsson and Sjögren, 1979a; Gabrielsson *et al.*, 1988). However, the present M and H filters represent too pronounced boosts (about 9 dB) and make fidelity worse.

III. DISCUSSION

The manipulations of the sound level and the frequency response affected all perceptual dimensions included here.

The results mainly agree with what could be expected from our earlier investigations (see the Introduction). The higher, natural sound level provided better clarity, more fullness, spaciousness, and nearness—but less softness/gentleness—as well as better fidelity than the 10-dB reduced level. There was no difference with regard to brightness. Use of the L filter resulted in more fullness and possibly softness/gentleness, but less clarity, spaciousness, and brightness (= more dullness), and worse fidelity than with the flat response. These effects can all be attributed to the frequency response of the L filter. The M filter gave less softness (= more sharpness) than the flat response, which may be due to its frequency response or to an accompanying increase in the overall sound level, or to both. The M filter also gave worse fidelity at the higher, natural sound level than the flat response. This effect is due to its frequency response. The H filter resulted in better clarity and more brightness, but less softness (= more sharpness) and possibly fullness, and worse fidelity at the higher, natural sound level than the flat response. The effects on brightness, fullness, and fidelity are due to its frequency response, while the effects on clarity and softness may depend on its frequency response or on the accompanying increase of the overall sound level, or on both.

There was often an interaction between filters and programs such that the effect of the H filter was different for the voice program than for the other programs, probably due to the restricted frequency range of the voice program. There were also interactions with the sound levels, meaning that the effects of the filters and/or the difference among the programs were more pronounced at the higher sound level than at the lower; see the results for loudness, fullness, softness, and fidelity in Fig. 5.

Interestingly, the results for the "neutral" noise program are very similar to those for the jazz music program as seen in Fig. 5. Those two programs are also rather similar in their long-term average spectrum.

ACKNOWLEDGMENTS

We wish to express our gratitude to Ove Till and to our subjects. This work was supported by The Bank of Sweden Tercentenary Foundation and the Swedish Ministry of Health and Social Affairs, Commission for Social Research.

APPENDIX

1. Instructions

You are going to listen to various reproductions of speech, music, and noise through earphones. Your task is to judge the sound quality of the different reproductions by means of the scales 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.

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).

Fidelity: Judge how similar the reproduction is to the original sound. 10 = perfect fidelity, 9 = very good, 7 = rather good, and so on. (This scale is not used for the noise.)

There is a new response form for each reproduction. Mark your judgment on each scale by a straight vertical line. Do your ratings on each scale without looking at the other scales or earlier response forms. There are no right or wrong answers. It is solely *your* opinion about the sound that should be decisive.

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

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

Gabrielsson, A. (**1979b**). "Statistical treatment of data from listening tests on sound-reproducing systems," Tech. Audiol. Rep. No. 92 (Karolinska Institute, Stockholm).

Gabrielsson, A. (**1987**). "Planning of listening tests: Listener and experimental variables," in *Perception of Reproduced Sound*, edited by S. Bech and O. J. Pedersen (Technical University of Denmark, Copenhagen), pp. 51–60.

Gabrielsson, A. and Lindström, B. (**1985**). "Perceived sound quality of high-fidelity loudspeakers," J. Audio Eng. Soc. **33**, 33–53.

Gabrielsson, A., Lindström, B., and Till, O. (**1986**). "Loudspeaker frequency response and perceived sound quality: Measurements in listening room," Tech. Audiol. Rep. No. 114 (Karolinska Institute, Stockholm).

Gabrielsson, A., Lindström, B., and Till, O. (**1987**). "Loudspeaker frequency response and perceived sound quality: Comparison between measurements in listening room, anechoic room and reverberation room," Tech. Audiol. Rep. No. 115 (Karolinska Institute, Stockholm).

Gabrielsson, A., Rosenberg, U., and Sjögren, H. (**1974**). "Judgments and dimension analyses of perceived sound quality of sound-reproducing systems," J. Acoust. Soc. Am. **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," J. Speech Hear. Res. **31**, 166–177.

Gabrielsson, A., and Sjögren, H. (**1979a**). "Perceived sound quality of sound-reproducing systems," J. Acoust. Soc. Am. **65**, 1019–1033.

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

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

Komamura, M., Tsuruta, K., and Yoshida, M. (**1977**). "Correlation between subjective and objective data for loudspeakers," J. Acoust. Soc. Jpn. **33**, 103–115.

Kryter, K. D. (**1970**). *The Effects of Noise on Man* (Academic, New York).

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

Staffeldt, H. (**1974**). "Correlation between subjective and objective data for quality loudspeakers," J. Audio Eng. Soc. **22**, 402–415.

Stevens, S. S., and Davis, H. (1938). *Hearing: Its Psychology and Physiology* (Wiley, New York).
Toole, F. E. (1986). "Loudspeaker measurements and their relationship to listener preferences: Part 2," J. Audio Eng. Soc. **34**, 323–348.
Toole, F. E., and Olive, S. E. (1988). "The modification of timbre by resonances: Perception and measurement," J. Audio Eng. Soc. **36**, 122–142.
Winer, B. J. (1971). *Statistical Principles in Experimental Design* (McGraw–Hill, New York), 2nd ed.