Subjective Appraisal of Loudspeaker Directivity for Multichannel Reproduction*

NICK ZACHAROV, *AES Member*

Nokia Research Center, Tampere, Finland

A group of subjective experiments considering the loudspeaker directivity characteristics for five-channel audiovisual reproduction was examined. A large number of tests were performed with a screened and trained listening panel. The effects of listening position, loudspeaker directivity, and program were examined. Frontal and surround loudspeaker directivity characteristics were considered separately. Certain findings suggest that de facto standard approaches may not be optimal in terms of loudspeaker directivity.

0 INTRODUCTION

This is one in a series of papers based on the work conducted within the audio working group PGII of the Eureka Advanced Digital Television Technology (ADTT) project. The group included the following partners: Bang & Olufsen (B&O), British Broadcasting Corporation (BBC), Institut für Rundfunktechnik (IRT), Nokia Research Center (NRC), and Philips Research.

The project has focused on methods for improving television presentations, and the aim of the audio group has been to consider issues relating to 3/2 (three front and two surround channels) and 3/1 (three front and one surround channel) surround sound systems and how they can be optimized for domestic applications.

This paper considers mainly the loudspeaker directivity requirements for a five-channel system, examining the front and surround aspects individually, closely watching the naturalness of the projected presentation.

0.1 Background

Over the years, many different multichannel sound systems have been examined [1] with the primary aim of improving the overall realism of the spatial sound reproduction. As early as 1958 [2] researchers were considering the inclusion of atmospheric surround information in addition to the mono or stereo signal. In more recent times the concepts of quadraphony and ambisonics [3] have considered the overall three-dimensional sound field from a more holistic standpoint. However, none of these systems has gained widespread long-term acceptance.

The advent of the Dolby Prologic and Lucas THX [4]–[6] systems has shown that many households are willing to support the requirements of a four-channel (3/1) loudspeaker system for the benefit of surround sound reproduction. At present, with the advent of near transparent, efficient audio compression algorithms, discrete multichannel reproduction is once again being considered as a viable domestic format. This has now paved the way for the superior characteristics of discrete multichannel reproduction (3/2).

In recent years researchers [7], [8] have considered the requirements for a reproduction system in terms of the number and placement of loudspeakers for optimal spatial reproduction. The conclusions of Theile, Ohgushi, and others have resulted in the general acceptance of a system of five discrete channels, where loudspeakers are placed in the plane of the ear at 0°, ±30°, and ±110° (Fig. 1).

One method of thinking of the three-dimensional sound field can be to consider it to consist of a collection of acoustic cues which indicate the type of acoustic space we are in. Theile [9] has considered these to consist of three main types: precedent cues (directional sound), nonprecedent cues (indirect sound), and environmental cues (atmospheric cues), as illustrated in Fig. 2. Examples of these cues are given in Table 1. For a correct impression of the acoustic space to be reproduced, all of these cues should be in place. Certain research groups [5], [6] have considered that the surround sound cues are dominated by reverberant and environmental cues, which all have somewhat diffuse characteristics.

* Presented at the 102nd Convention of the Audio Engineering Society, Munich, Germany, 1997 March 22–25; under the title "On the Loudspeaker Directivity Considerations for 5.1-Channel Audio-Visual Reproduction: A Subjective Appraisal," revised 1998 January 6.

If this is the case, then this sound field is rather difficult to achieve with a limited number of surround loudspeakers. For this reason, the dipole loudspeaker (acoustic null toward the listening position) has been considered as a viable option to enhance the diffusion of the surround sound field. However, in this way the directional information of the precedent and nonprecedent cues is poorly reproduced. It should also be noted that even environmental cues such as applause or rustling leaves have directional components, though the sound is predominantly randomly incident. The question thus arises, what are the optimal loudspeaker directivity requirements for domestic spatial sound reproduction?

The interaction between auditory and visual perception has been shown to be very complex [10], [11] and has been studied by program makers and researchers alike. These two modalities function independently in the perception of space, but can be combined to enhance perception. For this to occur there must be a strong spatial and temporal coherence between modalities. Incoherence of either property can lead to a rapid degradation of the overall perceived quality (spatial and otherwise). For this reason the discussed experiments have been designed with the inclusion of visual images to maximize the perceived spatial quality [12], [13].

In the domestic environment it is frequently rather difficult to achieve the optimal arrangement suggested in Fig. 1. The most common difference is that the central listening position may not be the only one. Thus an important issue to be considered is suboptimal listener placement and how this is affected by the loudspeaker directivity.

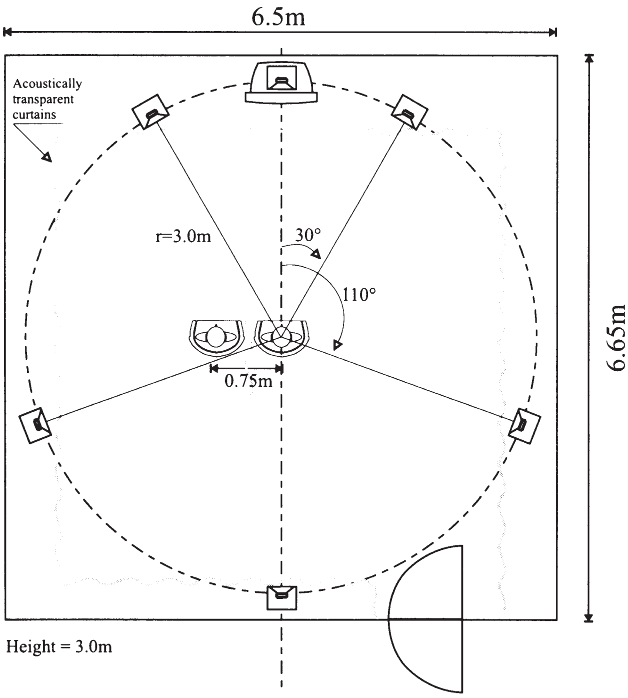

Fig. 1. NRC listening room setup for both experiments.

Methods for evaluating the subjectively perceived quality of both audio and visual images have been tested by many workers in the field [12], [13]. At the present time attempts are being made to standardize such test methods [14], [15], and where suitable, these methods have been applied to the experiments described herein.

A subjective experiment was thus designed to examine the effects of:

1) Loudspeaker directivity in spatial sound reproduction
2) Program material and type
3) Listening position.

These factors were examined in two independent experiments. The first considers the surround loudspeaker directivity requirements, whereas the second focuses on the frontal loudspeaker directivity requirements. All experiments were performed with accompanying images to support the projected spatial presentation.

The null hypothesis H_0 for this study can be stated as: No differences exist between systems in terms of spatial sound reproduction. The alternative hypothesis H_1 would be: Differences do exist between systems in terms of spatial sound reproduction.

This paper is organized in the following manner. Section 1 covers the general aspects of the subjective experimental arrangement. Section 2 is concerned with the objective measurements of the systems under test. Section 3 presents the surround experiment and data analysis and Section 4 looks at the frontal experiment. Section 5 provides conclusions.

1 EXPERIMENTAL DETAILS

In this section general details of the experimental procedure are given. Further details are also provided for the individual experiments. More complete details can also be found in the original version of this paper.[1]

[1] Preprint 4459, presented at the 102nd Convention of the Audio Engineering Society, Munich, Germany, 1997.

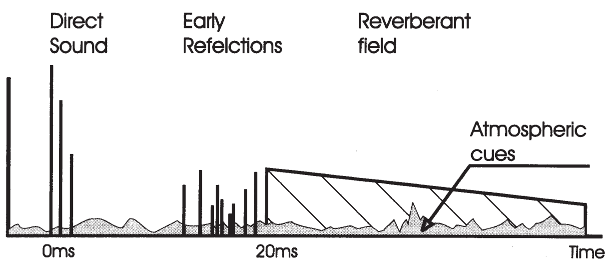

Fig. 2. Spatial cues.

Table 1. Three-dimensional spatial sound cues.

Cue	Sound Classification	Examples
Precedent	Directional sound	Clearly localizable sounds: person speaking, passing car, gun shot
Nonprecedent	Indirect sound	Reverberation, lateral reflections, early reflections
Environmental	Nonreflected atmospheric sound	Applause, wind in trees, air-conditioning

The systems under test can be considered as four transfer function blocks: program, system, room, and listener. For the purposes of this experiment, room and listener remain constant whereas program and system are varied.

1.1 Listening Room

All experiments were performed in the NRC Speech and Audio Systems Laboratory listening room, which was designed in accordance with IEC 268-13 [16]. In addition the room was tested for conformance with ITU-R Rec. BS 1116 [15]. The room satisfies the ITU-R recommendation with the exception that its width is 0.5 m wider than required and the background noise level marginally exceeds NR15 above 3.5 kHz [17].[1] The room is symmetrical about the central axis (see Fig. 1).

1.2 Listening Panel

An initial group of 20 persons was screened audiometrically for normal hearing and the subjects' active interest in listening to or playing music. An Internet web-based questionnaire was created for the selection process [17]. The entire group went through the three-phase training, which consisted of training, induction, and two training experiments.

Following screening, the panel was trained in aspects of spatial sound and surround sound reproduction. They were introduced to the concepts of the different spatial cues and the situations in which these occur in our environment. Structured exposure to different acoustic environments was organized, and the subjects were continuously encouraged to increase their awareness of everyday spatial sound cues.

An induction phase then followed, which consisted of introducing the subjects to the grading scale, questions, and test systems and exposing them to some of the test systems and all of the program items. Groups of four to six persons were trained together and encouraged to discuss what aspects of the spatial sound field they were becoming aware of. Each induction session lasted 2–3 hours.

All 20 subjects then participated in two training experiments: frontal and surround. These were limited to three systems, two programs, and one listening position. All factors were repeated twice. This experiment was used solely for the listener selection process. The questions and attributes tested are discussed in Sections 3.3 and 4.3.

Based on this smaller training experiment, listeners were selected according to their listening reliability. This matter was studied by considering an individual's error variance for a four times repeated rating of the same system–program combination. An analysis of variance (ANOVA) model [18]–[20] was also performed, and subjects were selected based on the magnitude of the loudspeaker F statistics and within-listener error variance, as reported by Bech [21]. The resulting panel consisted of 10 persons (nine male, one female) within the ages of 20 and 33, which was found to be a suitable group size to achieve a 95% confidence level ($p < 0.05$) for these experiments [22].

Based on this selection procedure and training, the panel may be deemed "expert" in accordance with ITU-R Rec. BS 1116 [15].

During the postprocessing of the results from the main experiments, one of the listeners was found to be using the rating scale consistently, but inappropriately. The distribution of that listener's data was clearly different from that of the main group and thus in contradiction to the ANOVA assumption of equal variance between groups. This subject was dropped from subsequent data analysis. Other listeners were checked for reliability and found to be suitable.

1.3 Calibration and Equalization

In line with other studies performed within the ADTT project [23] a B-weighted pink-noise signal was used for objective system calibration, employing a linear, slow average sound pressure level (SPL) measurement. For the purposes of this work, a 110-Hz high-pass filter was applied to the reproduction systems to ensure equal bandwidth. Level was measured with a calibrated Ono Sokki sound level meter, with an upward pointing ½-in (12.8-mm) microphone at the central listening position.

Due to the desire to compare the performance of both 3/2 and 3/1 surround systems, the level calibration of both systems is discussed in detail to ensure a favorable comparison.

1.3.1 3/2 System

The calibration according to ITU-R Rec. BS 1116 for the A-weighted reproduction SPL for each channel is given by

$$L_{\text{ref}} = 85 - 10 \log(n) \pm 0.25 \quad [\text{dB}]$$

where n is the number of reproduction channels in the total setup.

For the five-channel system this yields a calibration level (A-weighted, slow meter) of 78 dB. During the training phase, the calibration level (unweighted, slow meter) was 78 dB in accordance with [23]. However, members of the listening panel objected to the high peak SPLs during program reproduction (peak level, fast meter > 96 dB). For the main experiments the calibration level (unweighted, slow meter) was therefore set to 76 ± 0.2 dB.

1.3.2 3/1 System (Single Surround Loudspeaker)

Due to the desire to compare both 3/2 (five-loudspeaker) and 3/1 (four-loudspeaker) systems in the same experiment, the calibration according to ITU-R Rec. BS 1116 cannot be adhered to for the 3/1 system, as this would yield a different calibration level compared to the 3/2 systems. If the standard calibration were applied, this would result in a difference in the front–back energy ratios of the two configurations. For this reason a calibration of the 3/1 system was made that would ensure that the total energy from both the front and the rear of the systems was equal to that of the 3/2 system and that the total system calibration level was the same. This was found to occur both subjectively and objectively when the calibration level of the 3/1 surround channel was 4 dB higher than that of the individual 3/2 surround channels.

1.3.3 Equalization

Equalization of the systems was considered in detail. Certain researchers suggested equalizing systems to have a flat one-third-octave frequency response when measured using a head and torso simulator at the listening position [24]. This was tested and the responses for each system were obtained at each position (left, center, right, surround left, and surround right). Measurements for left and right ears provided very similar one-third-octave amplitude responses (Fig. 3), and the option to equalize was rejected as unnecessary.

To compensate for any slight timbral differences, the listeners were asked not to grade timbral differences between the systems unless they found that timbral differences also affected spatial reproduction.

1.4 Preparation of Program Material

Audiovisual material for the experiments was obtained from two sources. Commercially available NTSC laserdisc material was transcoded to PAL via a broadcast-quality transcoder. Audio was decoded via a Yamaha DDP-1 audio processor. Original BBC material was down-converted from high-definition to standard-definition PAL (4:3). Image quality was not a factor in these experiments. Some of the recording techniques used for the BBC material are documented in [25].

Material was edited and compiled via the AVID digital audiovisual workstation. The final test material was recorded and time-code synchronized to Betacam SP video and Sony PCM 800 eight-track digital audio tape (48-kHz sample rate). The final test configuration consisted of several 60- to 120-s program items per tape with a 75-s scoring time between items (Fig. 4). Four sets of test tape were made for each of the two experiments, each containing four identical program items, arranged in a Latin square configuration [19]. The use of the tapes was randomized during the experiments.

The eight selected items were chosen carefully to include a broad range of acoustic stimuli, including speech, music, atmospheric effects, reverberation and room effects, outdoor effects, applause, directional effects, and so on. Each item was screened carefully for coherence between auditory and visual cues, particularly for the frontal experiment.

The reproduction system block diagram is shown in Fig. 5.

1.5 Test Items

1.5.1 Frontal Experiment

1) *Wimbledon women's tennis final 1989*, BBC, 1:02:22:00–1:03:47:00: Scene comprising a tennis match with directional effect, applause, speech by umpire and commentators, and outdoor atmosphere.

2) *Tipsy gypsy*, BBC, 1:03:52:00–1:05:31:00: A concert inside the Royal Albert Hall, London, with audience cheering, applause, talking conductor, reverberation and room effects, singing, and orchestral performance.

3) *Braveheart*, Paramount, LV33118-2, side A, 00:07:27:00–00:09:15:00: Mainly an outdoor scene with directional panning effects, wind noise, flowing river, speech, night atmosphere, and some indoor atmosphere.

4) *True Lies*, *China Town*, 20th Century Fox, 8640-80, side A, 00:50:56:00–00:52:20:00: Different scenes, including inside car, in the rain, in Chinese restaurant, with speech, directional effects, panning effects, reverberation, and restaurant atmosphere.

1.5.2 Surround Experiment

1) *Bluebell Railway*, BBC, 1:00:00:00–1:01:58:00:

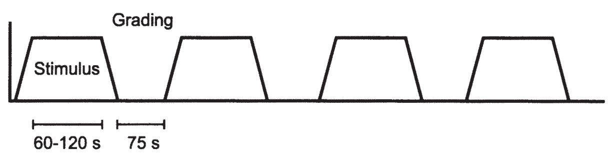

Fig. 4. Program stimulus and grading phase.

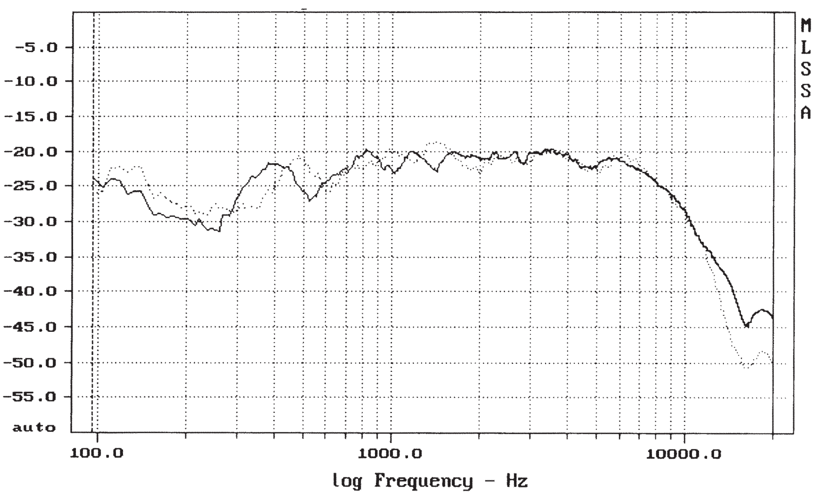

Fig. 3. Head and torso simulator one-third octave measurements of left surround loudspeaker. — — — cardioid; · · · dipole.

Scene consisting of steam train pulling away from station and approaching bridge; contains country atmosphere with directional cues and panning effects.

2) *FA Cup final 1989*, BBC, 1:01:05:00–1:02:17:00: Soccer game with audience applause, cheering, a few directional cues, and commentary.

3) *Lion King, A new day dawns for young Simba*, Walt Disney home video, 2977AS, side A, 00:08:02:00–00:09:30:00: Animated scene within a cave with directional effects, reverberation, some suopporting music, and panning effects. This animated scene was selected to consider how much difficulty subjects have in perceiving project space when the image is very derived.

4) *True Lies, Tango*, 20th Century Fox, 8640-80, side A, 00:11:30:00–00:13:32:00: Initially indoors, with orchestral music, reverberation, and later speech; then outdoors with directional cues, speech, and accompanying music.

1.6 Systems

For the purposes of the experiments, six different types of loudspeakers were used, each having very different directivity characteristics (Figs. 6–13). Only commercially available systems were tested. All of the designs were two-way electrodynamic types, some bass reflex design and others closed box, but this was not an issue due to the high-pass filtering applied.

The systems under consideration include the following types:

1) *Direct radiator*: Consisting of a coaxial 25-mm tweeter and 115-mm woofer. This type of loudspeaker has a reasonable on-axis performance but suffers from phase cancellation effects in the vertical plane (Fig. 7) and exhibits a typical twist in the power response around the crossover frequencies [17]. This design is commonplace in domestic applications and has a very specific radiation characteristic.

2) *Vertical line*: Consisting of a 25-mm tweeter in the center of a line of four 65-mm woofers. Total line length is 500 mm. As with all line sources, this device provides some increase in directivity in the plane of the line (Fig. 10). Perpendicular to this plane (Fig. 11) the directivity has problems similar to those of simple direct-radiator two-way designs.

3) Horizontal line: As vertical line, but oriented horizontally.

4) *Cardioid*: Basically a direct-radiator design with

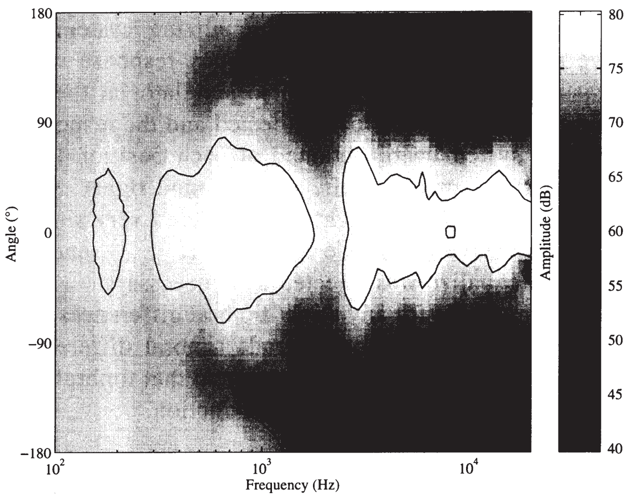

Fig. 6. Direct-radiator horizontal directivity.

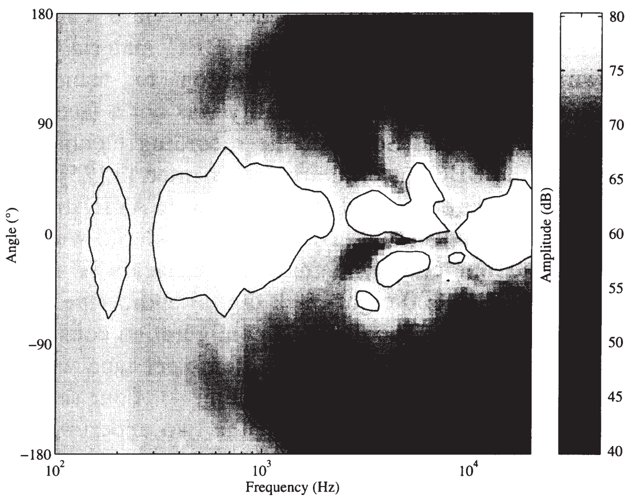

Fig. 7. Direct-radiator vertical directivity.

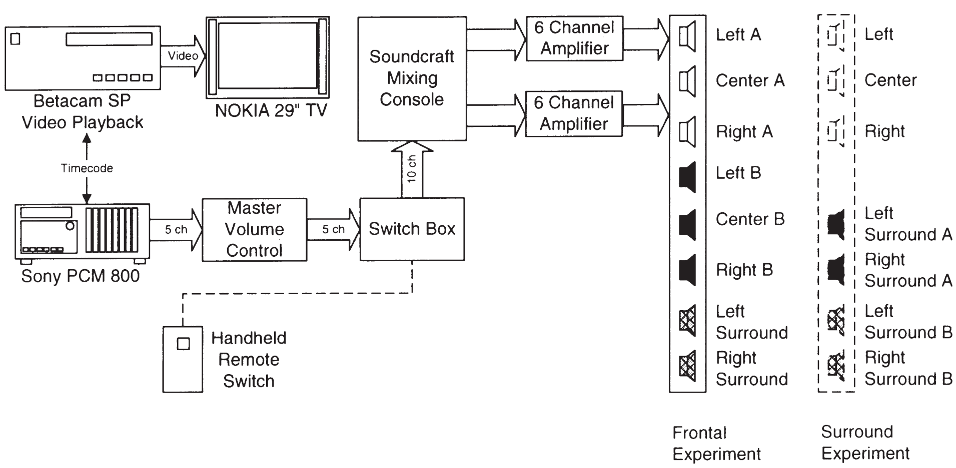

Fig. 5. Electrical audiovisual setup for frontal and surround experiments.

a 20-mm tweeter in a waveguide and a 150-mm woofer. The waveguide design differentiates this from the direct-radiator design, leading to a horizontal polar characteristic that is essentially cardioid in nature. The larger effective radiation area of the waveguide also reduces the twist in the power response [17] near the crossover frequency (3.5 kHz) and improves the vertical directivity characteristic (Figs. 8 and 9).

5) *Pseudo omnidirectional design*: Consisting of a coaxial 20-mm tweeter and a 110-mm woofer. The near

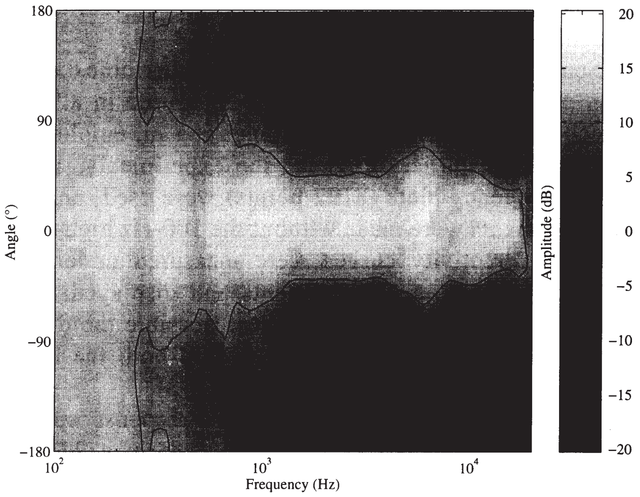

Fig. 8. Cardioid-system horizontal directivity.

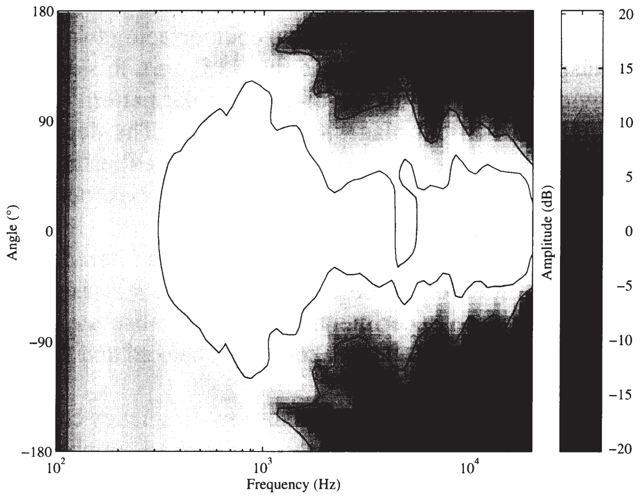

Fig. 11. Vertical-line-source horizontal directivity.

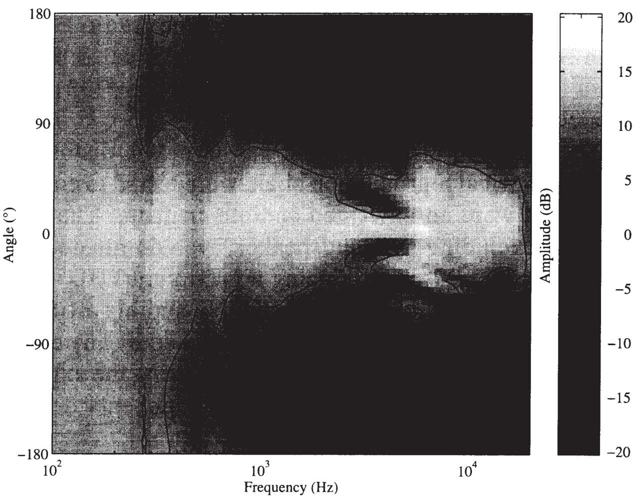

Fig. 9. Cardioid-system vertical directivity.

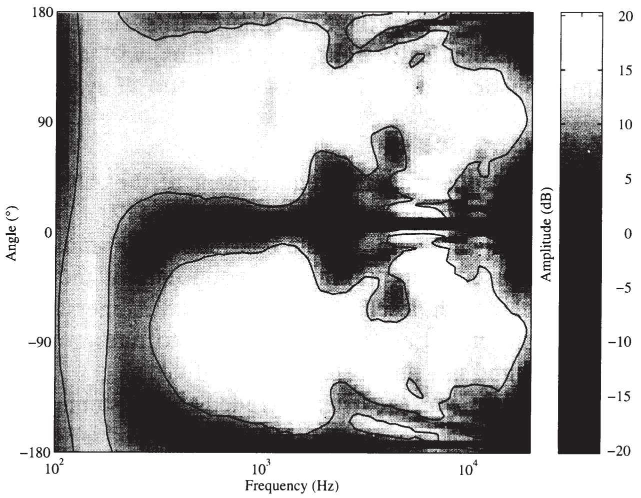

Fig. 12. Dipole-source horizontal directivity.

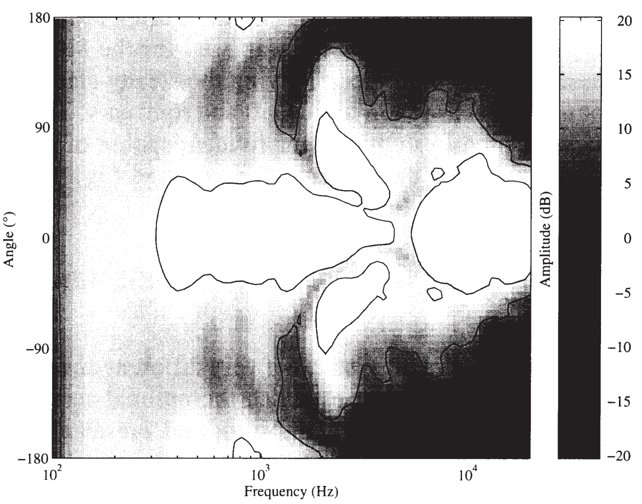

Fig. 10. Vertical-line-source vertical directivity.

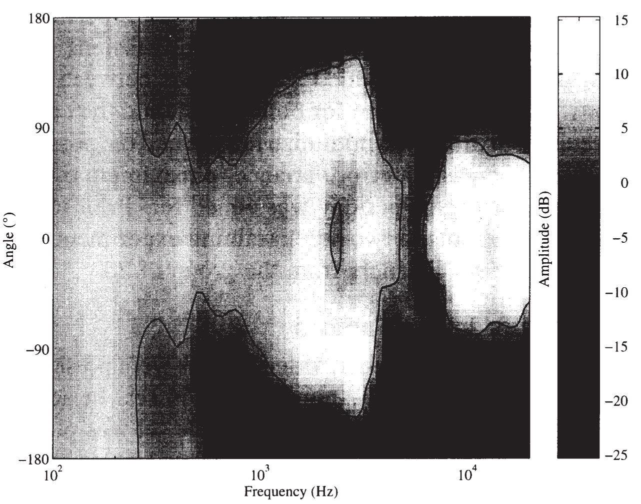

Fig. 13. Pseudo omnidirectional-source horizontal directivity.

omnidirectional radiation characteristic is achieved by directing the loudspeaker elements toward a conical reflector, which then radiates the sound broadly in the horizontal plane. In the vertical plane this loudspeaker does not suffer from phase cancellation at the crossover frequency because of the coincident axis of both elements (Fig. 13).

6) *Dipole*: Consisting of two counterfacing two-way loudspeakers (two 20 mm and two 115 mm), this design provides a close to "figure-of-eight" polar pattern (horizontal and vertical planes, see Fig. 12). The depth of the null is on the order of 15 dB. In the experimental design the null is aligned with the listening axis (perpendicular to the main lobes).

All loudspeakers were 110-Hz high-pass filtered during the calibration process and experiments to exclude bandwidth as an experimental factor. The absolute electrical phase of the loudspeakers was checked and found to be consistent.

The use of the so-called ".1" or low-frequency energy (LFE) subwoofer signal was considered to support the band-limited main channels. However, due to the variety of program material, both with and without ".1" information, and the complexity of subwoofer calibration [26], this possibility was rejected.

For the purposes of the experiment, all systems were hidden by acoustically transparent cutains, with a maximum acoustic error of 1 dB at 17 kHz for all systems.

1.7 Grading Scale

In accordance with the requirement of the ANOVA model and other similar work in this field [12], [27], a 0–10-point one-decimal-place continuous 100-mm line scale was used for grading. The endpoints of the line were labeled with descriptive adjectives. Listeners were given instructions and trained on how to interpret the questions and use the scale.

For each session, the listener graded each system independently for each program item. This method is sometimes referred to as a hybrid paired-comparison method [12].

1.8 Test Duration

Both experiments were performed over a period of three weeks. During this time six 11-min sessions were run for each person, with 15-min breaks between sessions. This was repeated for both the left and the center listening positions. To minimize fatigue and the possible resulting increase in error variance, a maximum of two sessions per person per day was aimed for. Prior to the main experiments, a two-week training experiment was run to preselect listeners from the group of 20.

1.9 Picture Image Factors

During the experiments, video images were presented via a 100-Hz 29-in (737-mm) television set. In accordance with image viewing standards the set was placed at a viewing distance equal to six times the picture height and the background light level at the screen was <5 lux [14].

2 OBJECTIVE MEASUREMENTS

2.1 System Benchmarking

All of the loudspeakers used in this study were tested in detail to ensure fair matching between units of the same type. Measurements were made in both the anechoic and the reverberant chambers to consider all of the loudspeaker parameters.

Sensitivity, frequency response, and harmonic distortion plus noise (THD + N) were measured in a 7-m^3 anechoic chamber with a 1-W input at 1 m, employing the power amplifiers used for the experiments. Sinewave frequency plots (8192 points) were used, and all measurements were made without the 100-Hz high-pass filters.[1] Directivity curves were measured under the same conditions at 5° increments and displayed after one-third-octave smoothing. Where possible the directivity characteristics were measured in both the horizontal and the vertical plans (Figs. 6–13).

The one-third-octave sound power response curves were measured in accordance with [28] in a 275-m^3 calibrated reverberant chamber, averaged over nine positions. The resultant curves were normalized to 1 kHz. Further details of these measurements can be found in [17].

All measurements were made using the Audio Precision System Two or the MLSSA with a Brüel & Kjær ½-in (128-mm) calibrated microphone (type 4133) and associated electronics.

The amplitude response and the harmonic distortion of the electrical reproduction chain were tested to ensure that there were no unwanted audible artifacts. The distortion, excluding the loudspeaker, was dominated by the power amplifiers, producing a maximum THD + N of 0.07% at 10 W into an 8-Ω load. No audible artifacts (distortion, rattle, or buzz) were introduced by any components within the system. At no time during the experiments were the transmission chain components driven into clipping.

2.2 In-Room Measurements

For the purpose of studying the perceived responses in the listening room, the measurements were taken at the central listening position using the Cortex head and torso simulator. The MLSSA was used for the 800-ms time-domain measurements, giving the steady-state responses for each system for the left and right ears. These data were used to study the amplitude response of each system at the ear drum of the listener (see Fig. 3 for an example).

3 EXPERIMENT 1—SURROUND EXPERIMENT

3.1 Test Systems

For this experiment systems of very different directivities were considered to establish the optimal solution in terms of the realism of the projected presentation, envelopment, and accuracy of directional cues. It was also of interest to consider how poor a single surround loudspeaker would perform with a 3/1 signal compared to 3/2 configurations. This concept also functioned as a method of establishing a lower anchor, as suggested in [15]. The frontal three loudspeakers were kept constant throughout the tests, and the direct-radiator type was selected as a fair midpoint reference system for this purpose. The calibration of the systems was achieved objectively as described previously.

Eight test systems where considered in total. Due to this large number of systems the experiment was split into two blocks of four, so that all system combinations could be tested efficiently (that is, three repeated measures of the same program–system–position per person). The systems tested are shown in Table 2.

3.2 Experimental Procedure

For each session two listeners participated in the listening room at one time, separated by 750 mm and an acoustically transparent curtain. Subjects were asked not to communicate with each other during the experiment. For each session, the listeners compared two systems with the four program items. This method was chosen so that only one pair of loudspeakers was setup at one time to minimize acoustical problems associated with loudspeaker placement, mutual acoustical loading effects, and so forth. The order of the loudspeaker pair presentation was randomized to minimize any block effects. This aspect was later analyzed and found not to be significant within the ANOVA model.

Subjects were encouraged to move their heads during the listening period to improve sound localization. The stimulus and the grading phase for each experiment are illustrated in Fig. 4. Subjects could freely switch between the two test systems with a hand-held remote control (see Fig. 5).

3.3 Questions and Instructions

- *Question 1 (Q1)*: Do you feel enveloped by sound? (0 = not very surrounded, 10 = very enveloped)

  Subjects were asked to consider only the rear sound events in this respect and to consider the degree to which the sound enveloped them, without considering how natural or correct it was. The lower bound (0) was indicative of a mono source and the upper bound (10) of complete envelopment, such as when standing in the rain.
- *Question 2 (Q2)*: How detailed are the directional effects? (0 = unclear or fuzzy, 10 = very distinct)

  Subjects were asked to consider only the rear sound events in this respect and to consider the detail, position, and width of directional cues and how detailed these were. The naturalness of the presentation need not be considered.
- *Question 3 (Q3)*: How natural is the projected presentation? (0 = unnatural, 10 = natural)

  Subjects were asked to consider the spatial naturalness of the whole system, to compare the system to real-life experience, and how true to life the projected presentation was in terms of spatial reproduction.

The questions were partially based on the work reported in [13]. Listeners were instructed to grade systems based on spatial reproduction differences alone. They were asked not to consider any timbral differences unless these were also related to the spatial reproduction.

3.4 Statistical Analysis and Discussion

3.4.1 Analysis

An ANOVA model was used to analyze the data. To ensure that this method was valid, the data were checked thoroughly for the basic ANOVA assumptions, which is discussed in detail elsewhere [17]. On initial inspection the data from both experiments were found to be a slightly skewed normal distribution. This type of problem is typical of psychological rating scales of a bounded nature, as reported in [29]. One method to correct this skewness is to apply a transform to the scale. The "logistic on logistic transform" (LOLT) has been suggested for this purpose. However, once the scale is transformed, the interval scale is no longer linear.

To test for the significance of the skewness, an ANOVA was performed, with the factors program, person, system, position, and sitting (with up to two-way interactions), on both the original and the transformed data, and compared. The transformed data provided a decrease in the residual error of the model, but there was no major change in the F ratios, the significance levels, or their rank order. The ANOVA model is generally considered to be fairly robust to skewness of the normal distribution [18]. Bearing this and the results of the comparison in mind, it was concluded that the skewness will not adversely degrade the model, and the nontransformed data were used for the full analysis. In all models, for both experiments, the significance of the ANOVA model is on the order of $p < 0.001$.

Tests were run for variance between groups, and these were found to be reasonably equal for different populations, as assumed by the ANOVA model.

Table 2. Test systems for surround experiment.

System Label	Frontal System	Surround System	Configuration
1 (block 1)	3 direct radiators	2 direct radiators	3/2
2 (block 1)	3 direct radiators	2 dipole sources	3/2
3 (block 1)	3 direct radiators	2 horizontal-line sources	3/2
4 (block 1)	3 direct radiators	1 omnidirectional source	3/1
5 (block 2)	3 direct radiators	2 cardioid sources	3/2
6 (block 2)	3 direct radiators	2 omnidirectional sources	3/2
7 (block 2)	3 direct radiators	2 vertical-line sources	3/2
8 (block 2)	3 direct radiators	1 dipole source	3/1

3.4.2 Discussion

Having tested the validity of the ANOVA model, a general linear model was used to study the major factors (program, person, system, position). A summary of the ANOVA tables is provided in Table 3, illustrating in rank order the first six significant factors. Further details of the ANOVA analysis may be found in [17]. Results shown in Table 3 indicate that all the main factors are significant to a level of $p < 0.001$ for all questions. Based on these facts, the null hypothesis H_0 for these factors must be rejected, and the alternative hypothesis H_1 considered. Position is a marginal factor for question 2 ($F = 3.584, p < 0.059$), for which the null hypothesis must be accepted. This might be explained by the fact that directional cues are typically very short in duration, and this question is thus a more difficult one to assess.

The expected rank order of significance for questions 1–3 is person, system, program, position. Person is always a significant factor, as different individuals rate systems differently. It is also noted that there are two-way interactions between persons, and again this can be expected as individuals will rate differently for each factor.

Position–system is also found to be a factor that can be explained by the fact that different systems are graded differently, depending on the seating position, which might be associated with the loudspeakers' directional characteristics. Other higher order interactions have not been studied in detail and are generally not significant.

Based on the estimated marginal means presented in Figs. 14–25, the following conclusions have been drawn.

Some broad conclusion can be drawn from the data for all questions. First it is found that in general highly directional loudspeakers (cardioid and line sources) are preferred for all questions and program items, and independent of position. It is argued that this is due to an increase in the direct-to-reverberant sound ratio in the room, which leads to a lower excitation of the reproduction. This might be considered desirable in terms of the room's independence of reproduction, but this aspect has not been studied further in the present work. Furthermore, the directive source provides directional information that exists within precedent, nonprecedent, and environmental cues that are essential for many sounds to be reproduced correctly.

Both vertical line sources and cardioid designs are found to be superior for all questions. This is explained by the fact that they provide relatively high directivity while also ensuring that both seating positions have good coverage. The horizontal line is slightly inferior, even though it has the same power response as the vertical line. It is suggested that this is due to the inferior amplitude response at the off-axis listening position because of the directivity in the horizontal plane.

The 3/1 designs rate very poorly in all cases, though the single dipole design is found comparable to the poorer 3/2 configurations (Figs. 15, 19, and 23). This configuration is generally not widely accepted, but it is

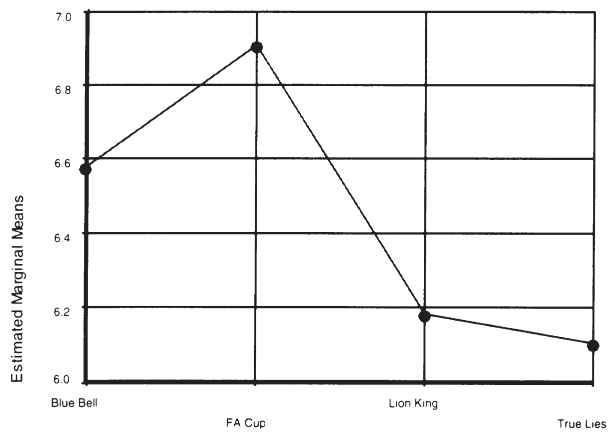

Fig. 14. Mean values of program for question 1 (surround), averaged across position and system.

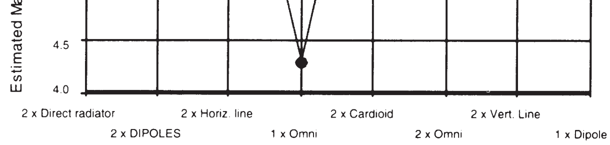

Fig. 15. Mean values of system for question 1 (surround), averaged across position and program.

Table 3. Summary of ANOVA tables for surround experiments in rank order of significance.

Question 1	Question 2	Question 3
Model ($F = 9.419, p < 0.000$)	Model ($F = 8.418, p < 0.000$)	Model ($F = 8.826, p < 0.000$)
Person ($F = 253.464, p < 0.000$)	Person ($F = 243.304, p < 0.000$)	Person ($F = 244.308, p < 0.000$)
System ($F = 171.921, p < 0.000$)	System ($F = 164.626, p < 0.000$)	System ($F = 95.715, p < 0.000$)
Program ($F = 51.528, p < 0.000$)	Program ($F = 24.370, p < 0.000$)	Program ($F = 48.831, p < 0.000$)
Person–program ($F = 167.457, p < 0.000$)	Person–program ($F = 12.648, p < 0.000$)	Person–program ($F = 19.380, p < 0.000$)
Person–position ($F = 10.796, p < 0.000$)	Person–position ($F = 7.913, p < 0.000$)	Position ($F = 13.359, p < 0.000$)
Position ($F = 9.210, p < 0.002$)	Position ($F = 4.820, p < 0.000$)	Person–system ($F = 11.723, p < 0.000$)

interesting to see that the single dipole is not considered far inferior to the standard 3/2 configurations (such as the direct radiator). Furthermore, this rating is significantly higher than that for the single omnidirectional source.

It is noted that coded material is rated inferior to discretely recorded material. This is not only due to the coding aspects, but also to the program content and the film production methods. To make a fair comparison of the coding scheme, further studies would have to be made with identical material, coded and uncoded. Figs. 16, 20, and 24 illustrate averages across programs as a representation of overall program-independent performance.

The off-axis listening position was found to provide generally higher ratings for most systems and questions. This was initially a puzzling result, but has now been

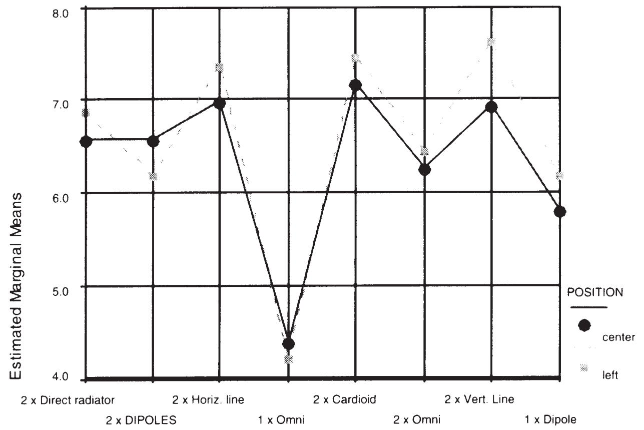

Fig. 16. Mean values of system for question 1 (surround), averaged across program.

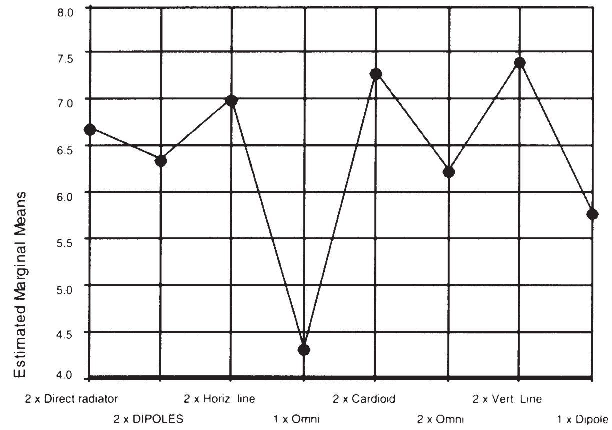

Fig. 19. Mean values of system for question 2 (surround), averaged across position and program.

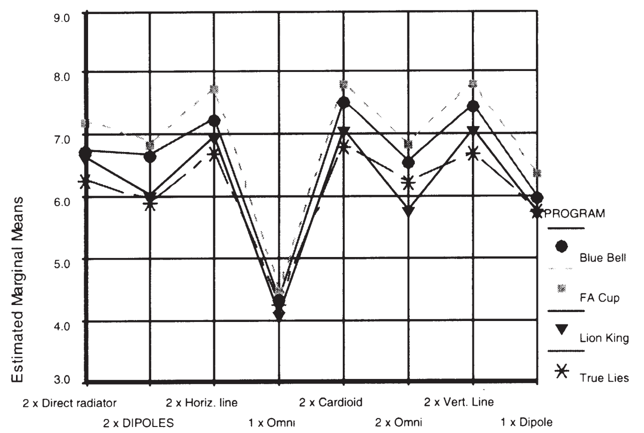

Fig. 17. Mean values of system for question 1 (surround), averaged across position.

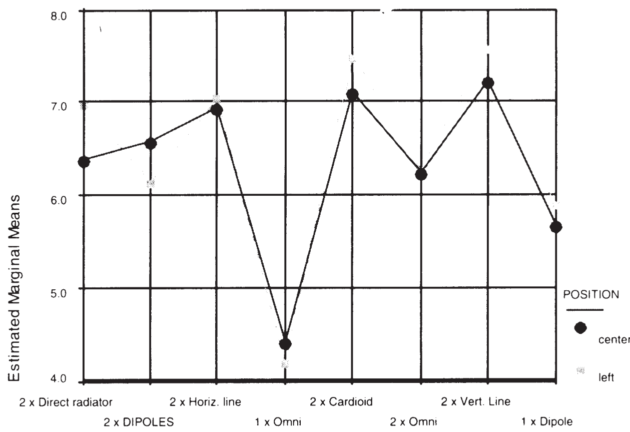

Fig. 20. Mean values of system for question 2 (surround), averaged across program.

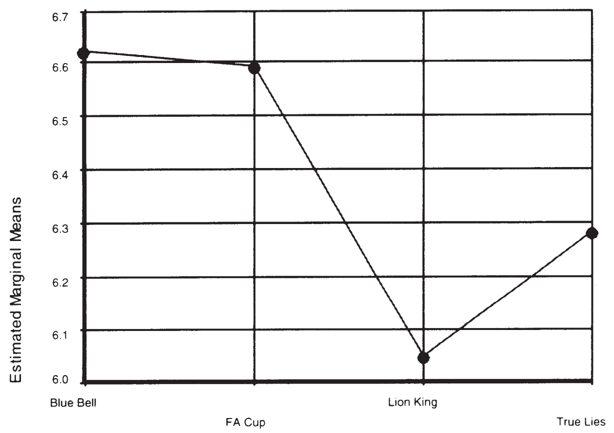

Fig. 18. Mean values of program for question 2 (surround), averaged across position and system.

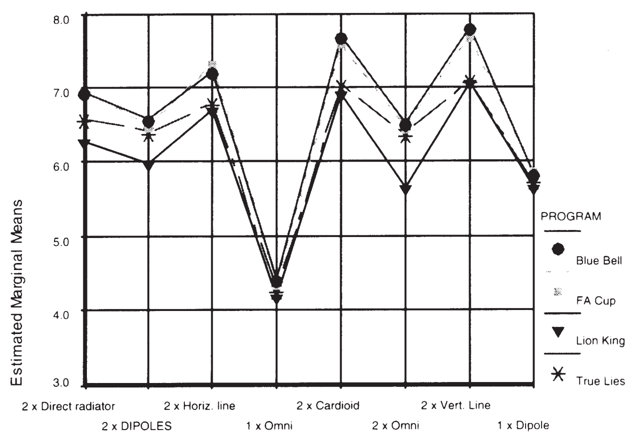

Fig. 21. Mean values of system for question 2 (surround), averaged across position.

explained by the symmetry of the room. Due to the acoustic and geometric symmetry of the room, there exist a large number of pressure standing-wave nodes across the width of the room. This type of phenomenon will provide less lower frequency energy at this position compared to the off-axis position and may lead to a lack of spaciousness [30]. This type of symmetry does not occur often in domestic environments, and it is expected that the result would not be repeated in a typical domestic listening environment.

In terms of the specific questions, the following conclusions were drawn.

Considering envelopment (question 1), the FA cup final program was found to give the highest ratings, possibly due to the surrounding atmospheric nature of the content (Fig. 14). This item was constantly graded high for all questions, even though detailed directional cues were missing. Having said this, it is well known that applause and cheering contain a lot of detailed directional information. This information is not well reproduced with a low-directivity system, that is, two pseudo omnidirectional or dipole sources. It is noted that even for directional information (question 2), this item is very well suited and that subjects still tend to prefer the directional sources (Figs. 19 and 20).

Specific to the naturalness of the projected presentation, subjects also rate highly the more directional systems, particularly the cardioid and vertical-line sources. The Lion King item is rated poorly in this respect as there are few directional surround cues in this item and error variances for this item were significantly higher than for all other items.

4 EXPERIMENT 2—FRONTAL EXPERIMENT

4.1 Test Systems

The purpose of this experiment was to establish the optimal directivity of the frontal three loudspeakers. For this purpose four different directivity systems were used. Only 3/2 systems were considered and the surround loudspeakers were kept constant, using the direct-radiator type. The factor under consideration included the level of correlation between sound and picture image (question 1), naturalness of projected presentation, and acoustic space. For the purposes of this experiment the systems shown in Table 4 were tested.

The system with a single dipole center channel (system 2) was mainly introduced as a method of creating a lower anchor, though this type of system would have little application in practice. Once again three repeated measurements were taken for each program–system–position combination per person.

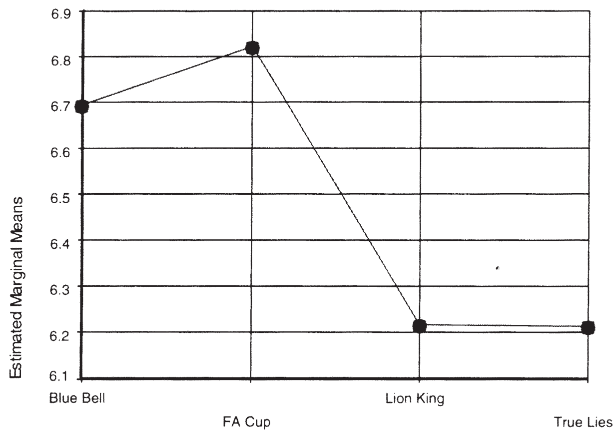

Fig. 22. Mean values of program for question 3 (surround), averaged across position and program.

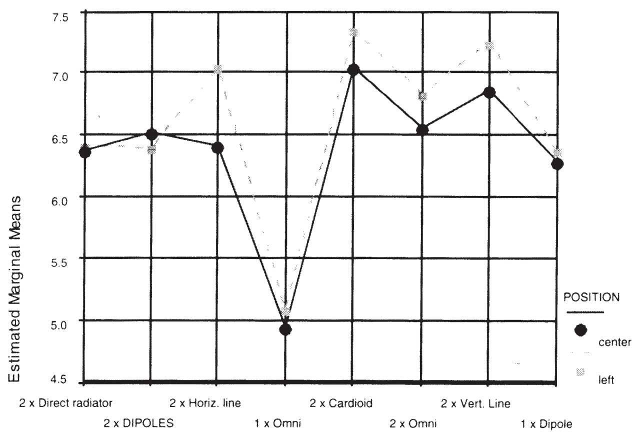

Fig. 24. Mean values of system for question 3 (surround), averaged across programs.

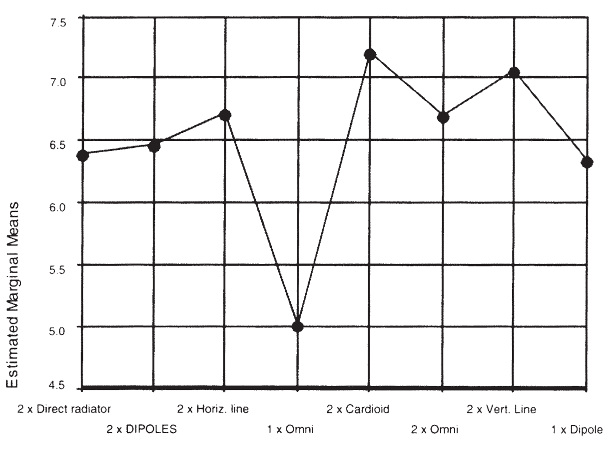

Fig. 23. Mean values of system for question 3 (surround), averaged across position and program.

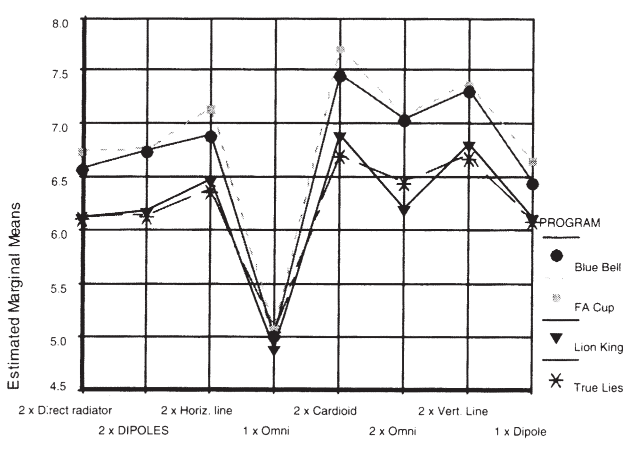

Fig. 25. Mean values of system for question 3 (surround), averaged across position.

4.2 Questions and Instructions

• *Question 1 (Q1)*: How well coordinated are sound and picture? (0 = uncoordinated, 10 = well coordinated)

Subjects were asked to focus on the frontal sound events only and to consider the correlation between sound and picture images in terms of image width, position, and motion. A grade of 0 could be considered if sound and picture image were completely uncorrelated.

• *Question 2 (Q2)*: How well do you sense acoustic space or changes thereof? (0 = poorly, 10 = clearly)

Subjects were asked to consider this question in terms of the whole system and to judge the correctness of the acoustic space suggested by the picture image and changes in acoustic space, if any. A mono sound source provides a poor sense of acoustic space (0).

• *Question 3 (Q3):* How natural is the projected presentation? (0 = unnatural, 10 = natural)

Subjects were asked to consider the spatial naturalness of the whole system, to compare the system to real-life experience, and to consider how true to life the projected presentation is in terms of spatial reproduction.

4.3 Statistical Analysis and Discussion

An ANOVA model was used to analyze the data as described for experiment 1. A type IV sum of squares method was applied to overcome the few data points missing due to experimental error. The factors person, program, position, and system were once again considered. Results for this analysis are presented in Figs. 26–37. A summary of the first six significant factors,

Table 4. Test systems for frontal experiment.

System Label	Frontal System	Surround System	Configuration
1	3 cardioid sources	2 direct radiators	3/2
2	2 cardioid sources (L & R) + 1 dipole source (center)	2 direct radiators	3/2
3	3 horizontal-line sources	2 direct radiators	3/2
4	3 direct radiators	2 direct radiators	3/2

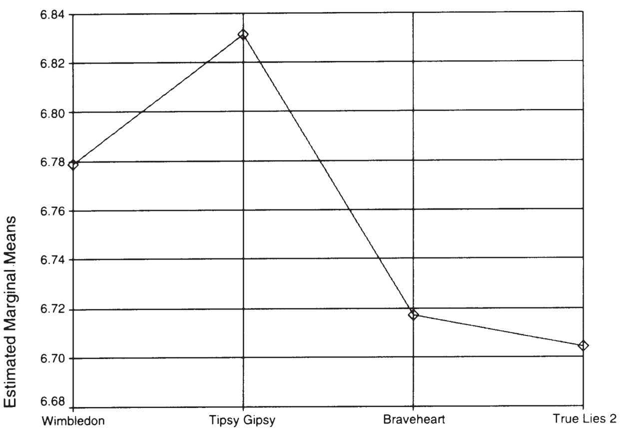

Fig. 26. Mean values of program for question 1 (frontal), averaged across position and program.

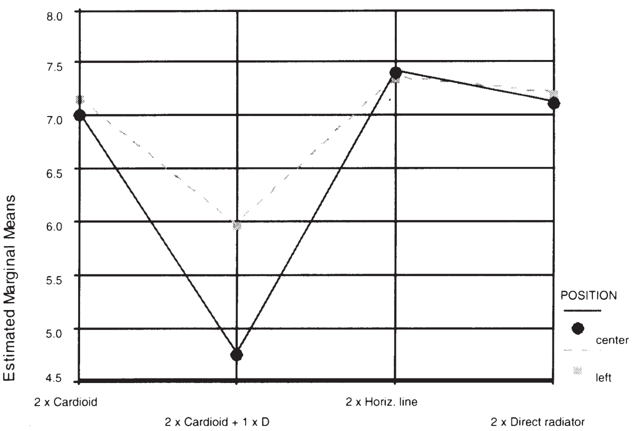

Fig. 28. Mean values of system for question 1 (frontal), averaged across program.

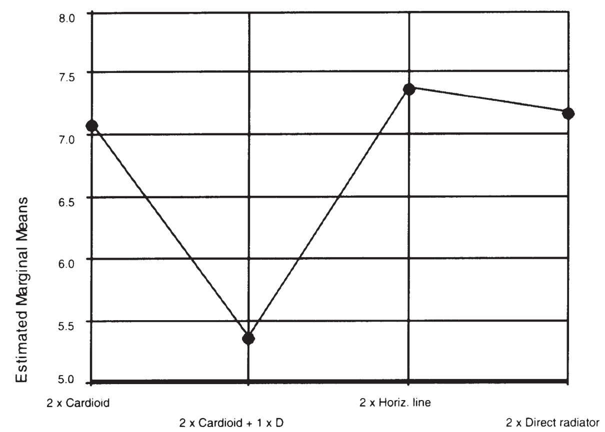

Fig. 27. Mean values of system for question 1 (frontal), averaged across position and program.

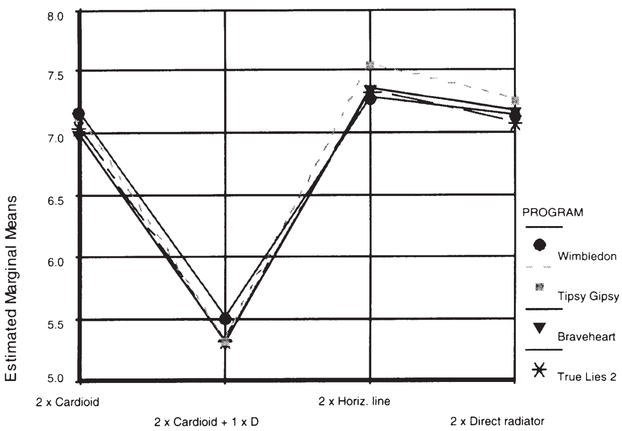

Fig. 29. Mean values of system for question 1 (frontal), averaged across position.

in rank order of the full ANOVA tables presented in [17] is given in Table 5.

For the frontal experiment (questions 2 and 3) the factors person, system, and program were found to be significant ($p < 0.001$). Based on these facts, the null hypothesis H_0 for these factors must be rejected, and the alternative hypothesis H_1 must be considered. Position was not a significant factor ($p > 0.1$), and so means can be averaged across this factor. This implies that position cannot be stated as having a meaningful effect on the subjects' rating of systems. Program was not found to be a significant factor ($F = 1.137, p > 0.333$) for question 1, while person and system remained significant.

In all cases the mixed system (two cardioids + one dipole) was graded significantly lower than the other systems

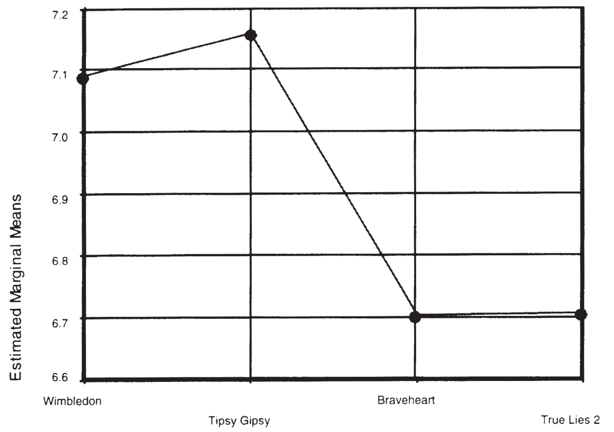

Fig. 30. Mean values of program for question 2 (frontal), averaged across position and system.

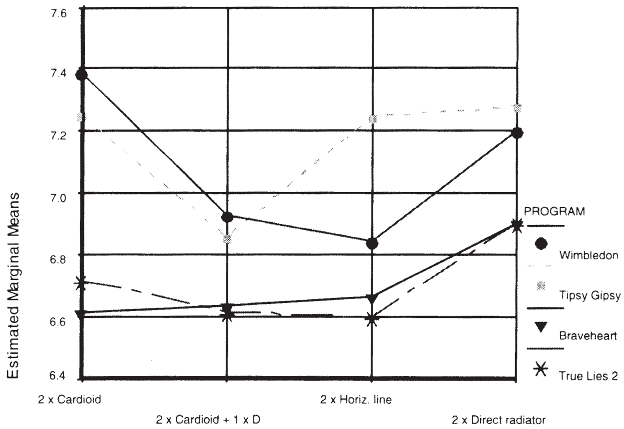

Fig. 33. Mean values of system for question 2 (frontal), averaged across position.

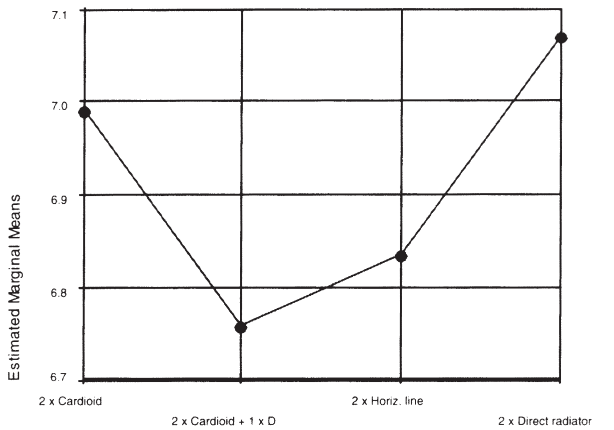

Fig. 31. Mean values of system for question 2 (frontal), averaged across position and program.

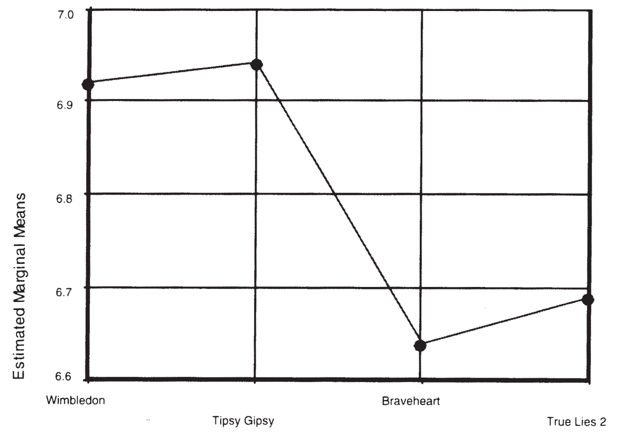

Fig. 34. Mean values of program for question 3 (frontal), averaged across position and system.

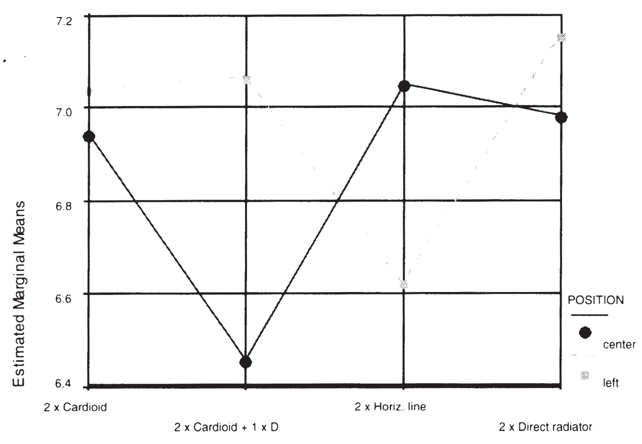

Fig. 32. Mean values of system for question 2 (frontal), averaged across program.

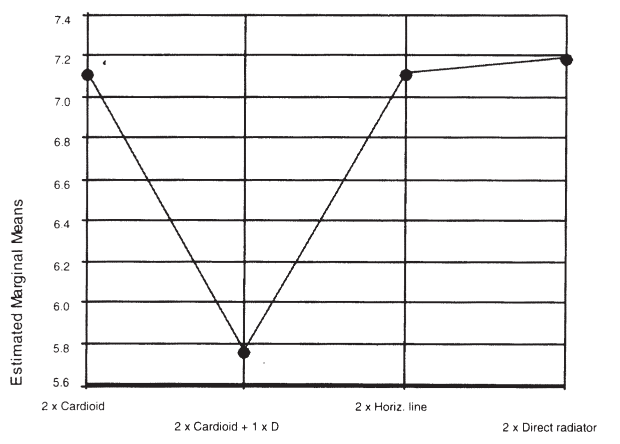

Fig. 35. Mean values of system for question 3 (frontal), averaged across position and program.

and served its purpose as a lower anchor. This also indicates that systems providing a low direct-to-reverberant energy ratio may be undesirable in frontal systems.

For questions 1 and 3 little variation was found between systems other than for mixed system 2 (Figs. 27 and 35). For the other three systems the ratings are essentially equal.

Once again it can be seen that the coded material is graded lower than the discretely recorded material, but the different program contents must be considered.

Specific to the questions, system has a relatively low F statistic ($F = 5.562, p < 0.001$) for question 2, below that of program ($F = 14.089$, $p < 0.000$), which is rather surprising. No significant variations between systems can be seen (Fig. 31). Subjects had commented during the experiment that the acoustic space question was somewhat difficult to interpret, and this may have lead to rather diverse ratings, as seen in Figs. 32 and 33.

It would appear that there are only minimal differences between systems in terms of sound and picture image coordination (question 1) and naturalness of projected presentation (question 3). In terms of directional cues and correlation to the picture, the horizontal line is found marginally superior to the other systems (Figs. 28 and 36).

Overall it could be concluded that for all feasible systems considered (low-anchor system excluded) there are only marginal differences in the ratings for all questions and material types. If anything, a higher directivity design appears to be slightly superior.

5 CONCLUSIONS

From these experiments the following conclusions can be drawn.

- Using a selected and trained listening panel, it is possible to obtain high statistical significance with a relatively small listening group of nine.
- Animated program material was found to be problematic as the content is too artificial for the subjects to consider it natural. This type of material should be avoided in audiovisual subjective tests.

5.1 Surround Experiment

- Higher directivity loudspeaker systems provide higher mean ratings for all questions and program items. This is explained by the fact that higher directivity loudspeakers provide a lower excitation in the reproduction room, leading to superior independence of the reproduction on the room. Furthermore, essential directional information in all types of cues is well reproduced with these sources. These findings suggest a different approach to the de facto standard methods, which support less the directional dipole source.
- Both cardioid and vertical-line sources provide superior ratings for on- and off-axis listening positions.
- In general subjects gave higher mean ratings for the off-axis seating position. This result has been associated with the loss in low-frequency energy and thus

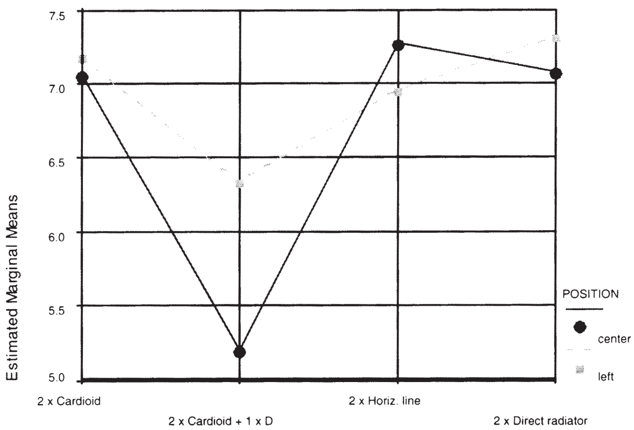

Fig. 36. Mean values of system for question 3 (frontal), averaged across program.

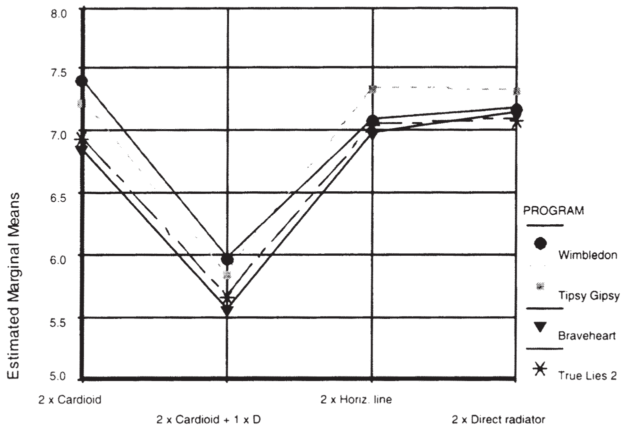

Fig. 37. Mean values of system for question 3 (frontal), averaged across position.

Table 5. Summary of ANOVA tables for frontal experiments in rank order of significance.

Question 1	Question 2	Question 3
Model ($F = 133.460, p < 0.000$)	Model ($F = 166.886, p < 0.000$)	Model ($F = 155.910, p < 0.000$)
System ($F = 159.087, p < 0.000$)	Person ($F = 126.406, p < 0.000$)	Person ($F = 137.296, p < 0.000$)
Person ($F = 107.538, p < 0.000$)	Program ($F = 14.089, p < 0.000$)	System ($F = 108.362, p < 0.000$)
Person–system ($F = 13.924$, $p < 0.000$)	Person–system ($F = 8.184, p < 0.000$)	Position–system ($F = 10.085$, $p < 0.000$)
Position–system ($F = 4.019$, $p < 0.008$)	System ($F = 5.562, p < 0.001$)	Person–system ($F = 9.065$, $p < 0.000$)
Person–position–system ($F = 3.716, p < 0.000$)	Person–position ($F = 4.159$ $p < 0.000$)	Program ($F = 5.712, p < 0.000$)
Position ($F = 2.403, p < 0.122$)	Position–system ($F = 4.119$, $p < 0.007$)	Person–position–system ($F = 3.220, p < 0.000$)

spaciousness at the central listening position due to standing-wave nodes at the center of the symmetrical room.
- In general the 3/1 (single-loudspeaker) reproduction system is far inferior to any 3/2 system, confirming previous studies in this field.

5.2 Frontal Experiment

- Little difference was found between the three systems under consideration (direct radiator, horizontal line source, and cardioid).
- The horizontal-line source is found to be marginally superior in terms of directional cues and picture image correlation.
- Lower directivity designs, providing low direct-to-reverberant energy ratios, are found to be undesirable.
- The listening position is not a significant factor in these frontal experiments (with systems offering three front loudspeakers), which supports other work in this field.

6 FURTHER WORK

The work in this field is far from complete, and this study only highlights some of the interesting issues and some of the areas in which further research could be performed.

Clearly, for a full understanding of the interrelationship between loudspeaker directivity and the quality of spatial sound reproduction a more precise set of subjective attributes should be created. It is suggested that a thorough study be made in this field, possibly using multidimensional scaling methods, to establish orthogonal unidimensional attributes. Once accurate attributes are in place, a more complete set of subjective tests could be performed with a better defined set of directivity systems, in different rooms, and with an extended set of program items.

A natural progression of this work could be to establish objective measures for the quality of spatial sound reproduction. To achieve this, a study of other room acoustic measures for spatial sound quality, such as interaural cross correlation, might be of interest. Based on the subjective data obtained in this work, it is possible to make further objective measurements and perform a correlation analysis on these data to find the most suitable results.

7 ACKNOWLEDGMENT

The Nokia Research Center is acknowledged for providing the dynamic environment for creative research. Tekes (Technology Development Centre of Finland) provided financial support of the study, which was part of the Eureka ADTT project. The author wishes to thank Dr. Søren Bech for his continuous support, endless patience, and stimulating discussion, as well as all the members of the ADTT project group for their continuous comments, discussion, and support throughout the work. He thanks his colleagues at NRC for their general assistance, and the members of the listening panel for their patience, participation, and good hearing. Mikko Suonio provided the listening test scheduling system and Rob Woods gave advice on applied statistical analysis.

8 REFERENCES

[1] G. Steinke, "Surround Sound—The New Phase: An Overview," presented at the 100th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 44, p. 651 (1996 July/Aug.), preprint 4286.

[2] H. Buttenberg, "Wege zum echten Klangerlebnis," *Radio und Fernsehen (Berlin)*, vol. 7, pp. 289–292 (1958).

[3] M. A. Gerzon, "Surround Sound Psycho-Acoustics," *Wireless World*, vol. 80, pp. 483–486 (1974).

[4] http://www.thx.com/thx/hth_audio_art.html.

[5] http://www.thx.com/thx/hth_dolby_art.html.

[6] http://www.thx.com/thx/hthx51.html.

[7] G. Ohgushi, K. Tsujimoto, S. Komiyama, K. Kurozumi, and J. Ujihara, "Subjective Evaluation of Multichannel Stereophony for HDTV," presented at the 81st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 34, p. 1024 (1986 Dec.), preprint 2363.

[8] G. Theile, "HDTV Sound Systems: How Many Channels?," in *Proc. AES 9th International Conf. "Television Sound Today and Tomorrow"* (Detroit, MI, 1991 Feb. 1–2), pp. 217–232.

[9] G. Theile,"Multichannel Sound Systems in the Home," IRT Kolloquium 95, Institu für Rundfunktechnik (1995).

[10] D. R. Perrott, "Auditory and Visual Localization: Two Modalities, One World," in *Proc. AES 12th International Conf.*, "The Perception of Reproduced Sound." (Copenhagen, Denmark, 1993 June 28–30).

[11] B. E. Stein and M. A. Meredith, *The Merging of the Senses* (MIT Press, Cambridge, MA, 1993), pp. 148–156.

[12] S. Bech, V. Hansen, and W. Woszczyk, "Interaction between Audio-Visual Factors in a Home Theater System: Experimental Results," presented at the 99th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 43, p. 1089 (1995 Dec.), preprint 4096.

[13] W. Woszczyk, S. Bech, and V. Hansen, "Interaction between Audio-Visual Factors in a Home Theater System: Definition of Subjective Attributes," presented at the 99th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 43, p. 1089 (1995 Dec.), preprint 4133.

[14] ITU-R Rec. BS 775-1, "Multichannel Stereophonic Sound System with and without Accompanying Picture," International Telecommunication Union, Geneva, Switzerland (1994).

[15] ITU-R Rec. BS 1116, "Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems," International Telecommunication Union, Geneva, Switzerland (1994).

[16] IEC 268-13, "Sound System Equipment," pt. 13,"Listening Tests on Loudspeakers," International Electrotechnical Commission, Geneva, Switzerland (1985).

[17] N. Zacharov, "Subjective Testing of Loudspeaker Directivity for Multichannel Audio," M.Sc. thesis, Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology (1997).

[18] Statistical Package for the Social Sciences (SPSS), "Training, Building ANOVA Models," SPSS (1995).

[19] G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters, An Introduction to Design, Data Analysis and Model Building* (Wiley, New York, 1978).

[20] A. Gabrielsson, "Statistical Treatment of Data for Listening Tests on Sound Reproduction Systems," Rep. TA 92, Department of Technical Audiology, Karolinska Institute, Sweden (1979).

[21] S. Bech, "Selection and Training of Subjects for Listening Tests on Sound-Reproducing Equipment," *J. Audio Eng. Soc.*, vol. 40, pp. 590–610 (1992 July/Aug.).

[22] H. Staffeld, "Choice of Sample Size in Listening Tests," presented at the 83rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 35, p. 1052 (1987 Dec.), no preprint.

[23] S. Bech, "Calibration of Relative Level Differences of a Domestic Multichannel Sound Reproduction System," *J. Audio Eng. Soc.*, this issue, pp. 304–313.

[24] H. Staffeld and E. Rasmussen, "The Subjectively Perceived Frequency Response in Small and Medium Sized Rooms," presented at the 123rd SMPTE Technical Conference (1981 Oct.).

[25] D. G. Kirby, N. A. F. Cutmore, and J. A. Fletcher, "Program Origination of Five-Channel Surround Sound," *J. Audio Eng. Soc.*, this issue, pp. 323–330.

[26] N. Zacharov, S. Bech, and D. Meares, "The Use of Subwoofers in the Context of Surround Sound Program Reproduction," *J. Audio Eng. Soc.*, this issue, pp. 276–287.

[27] S. E. Olive, P. L. Schuck, S. L. Sally, and M. E. Bonneville, "The Effects of Loudspeaker Placement on Listener Preference Ratings," *J. Audio Eng. Soc.*, vol. 42, pp. 651–669 (1994 Sept.).

[28] ISO 3741, "Acoustic Determination of Sound Power Levels of Noise Sources—Precision Method for Broad-Band Sources in Reverberant Rooms," 2nd ed., International Standards Organization, Geneva, Switzerland (1988).

[29] J. Allnatt, Transmitted Picture Assessment (Wiley, New York, 1983), pp. 22–40.

[30] L. Beranek, "Concert and Opera Hall—How They Sound," Acoustical Society of America (1996).

The biography for Nick Zacharov is published on page 286 of this issue.

