Perceptual evaluation of multi-dimensional spatial audio reproduction

Catherine Guastavino
McGill University, Department of Psychology, 1205 Dr. Penfield Avenue, Montreal, Quebec H3A 1B1, Canada

Brian F. G. Katz[a)]
LIMSI-CNRS, Perception Située, BP 133, F91403 Orsay France

(Received 2 September 2003; revised 13 April 2004; accepted 3 May 2004)

Perceptual differences between sound reproduction systems with multiple spatial dimensions have been investigated. Two blind studies were performed using system configurations involving 1-D, 2-D, and 3-D loudspeaker arrays. Various types of source material were used, ranging from urban soundscapes to musical passages. Experiment I consisted in collecting subjects' perceptions in a free-response format to identify relevant criteria for multi-dimensional spatial sound reproduction of complex auditory scenes by means of linguistic analysis. Experiment II utilized both free response and scale judgments for seven parameters derived form Experiment I. Results indicated a strong correlation between the source material (sound scene) and the subjective evaluation of the parameters, making the notion of an ''optimal'' reproduction method difficult for arbitrary source material. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1763973]

PACS numbers: 43.66.Lj, 43.66.Qp, 43.38.Md [MK] Pages: 1105–1115

I. INTRODUCTION

The use of multi-channel audio for the reproduction or simulation of multi-dimensional sound fields is becoming more common in research, artistic performances, home and commercial installations. In the field of psychoacoustic research, the ability to reproduce a multi-dimensional spatial sound field in laboratory conditions is advantageous for the study of auditory perception and cognition in complex sonic environments. A key question concerns the influence of the spatial presentation on a person's perception of various attributes of the reproduced sound field. In particular, how complete (spatially) must the information be for subjects to be ''convinced'' of the reproduction? In addition, are there potentially negative effects linked to providing ''too much'' information, and what if any is the balance? Is there a tradeoff between different perceptual aspects of the reproduced sound scene when more or less spatial information is included?

Sound quality assessment of reproduction methods have traditionally been concerned with non-spatial attributes, concentrating primarily on timbral and distortion issues when assessing the qualities of loudspeakers in monophonic reproductions (e.g., Eisler, 1966; Gabrielsson, Rosenberg, and Sjögren, 1974; Gabrielsson and Sjögren, 1979). Spatial attributes have however been investigated quite extensively in the field of room acoustics (e.g., Beranek, 1962; Schroeder, Gottlob, and Siebrasse, 1974; Kahle, 1995). More recently, the increasing use of multi-channel audio has led researchers to study spatial sound perception in the context of auditory displays (Rumsey, 1998, 2002; Berg and Rumsey, 1999, 2000, 2001, 2002; Zacharov and Koivuniemi, 2001), since spatial attributes are considered an important contributor to overall sound quality of multi-channel systems. The subjective evaluation of spatial features remains however at a very early stage in its development compared with other auditory attributes such as timber or loudness, and the need for a more accurate description of spatial attributes becomes clear to perceptually optimize multi-channel audio systems.

In the present work we examine the results of a set of listening tests in which several spatial loudspeaker configurations were compared, with a variety of source material. Subjects were presented with a reproduction of the same recorded sound scene over different systems. Subjects were asked to evaluate the different configurations using verbal descriptions and value scales. Perceptual evaluations of the different systems as a function of their dependence upon the source material are of particular interest, as the results highlight the fact that there is no single system that is optimal for all conditions.

II. EXPERIMENTAL SYSTEM

A. Recording and reproduction setup

There are various approaches for recording and reproducing spatially distributed audio. The recording industry has developed a wide range of methods over the years starting from 2-channel stereo, to 4-channel quadraphonic, and the current trend of 6-channel 5.1. Various other, often more complicated, systems have been developed for theatrical and performance situations using greater and greater numbers of channels in the recording and/or reproduction. Each system requires its own recording and reproduction technique, these being closely linked.

Our aim in the present work is to investigate the subjective differences regarding the spatial complexity of multi-dimensional audio reproduction. The interest of this work

[a)]Electronic mail: brian.katz@limsi.fr concerns the perceptual effects of spatial presentation and is not intended to be an evaluation of different recording techniques. The method employed in this study was a versatile recording and playback method which consists in recording the sound field with a compact 3-D microphone, containing near-coincident elements. This method, termed Ambisonics (Gerzon, 1977), was chosen as the best suited method for this study, since an Ambisonics recording can be decoded onto a variety of speaker configurations. For each sound scene, a single recording was used and only the spatial presentation of the information varied. In this manner, the effects of recording techniques, multiple microphone placements, and other bias in the technical aspect were minimized.

Ambisonics is an approach to sound field recording and reproduction that decomposes the spatial sound field into spherical harmonics. Currently available 1st order microphones provide four signals: W (zeroth order omnidirectional) and XYZ (1st order components representing the Cartesian axis with figure of 8 directivity patterns). This output result, termed B-format, captures the spatial information of the sound field, resolved into a mono reference signal and left–right, front–back, and up–down information, thus enabling the reproduction of full 3-D information. Reproduction of the sound entails a decoding process from the B-format signal to the array of loudspeakers. The decoding process results in a signal to each loudspeaker being composed of a combination of the spherical harmonics dependent upon the location of the speaker. There are various parameters in the decoding process, but their discussion is beyond the scope of this paper (cf. Gerzon, 1977; Fellgett, 1974; Gaskell, 1979; Daniel, 2000). All recordings used were made with a B-format Soundfield model ST250 microphone and decoded without shelf filtering (Furse, 2003) on an array of Studer A1 speakers and included a JBL 4545C subwoofer.

B. Design of the listening room

A prototype listening room was created for this experiment to test different reproduction methods with the conceptual goal of easing the process of abstraction from the listening room to the original environment. The design of the room can be divided into three parts: the acoustics, the visual, and the reproduction system.

The acoustics of the room were designed to be as dry as possible, given architectural limitations, in order to allow for the reproduction of outdoor soundscapes. The room has a flat frequency response and a reverberation time of <0.05 seconds for frequencies above 200 Hz. Below 200 Hz the reverberation time increased gradually to 0.2 seconds at 40 Hz. The room is acoustically isolated (floated construction) with internal dimensions 2.77×3.24×3.62 m.

The visual design of the room, the most strikingly different aspect as shown in Fig. 1, is based upon a hexagonal shape. The goals were to create a room with minimal reference to the sounds being reproduced or the subject's frame of reference, as well as to ensure that subjects are not visually aware of the test configuration. Other than the point of entrance, there is no Cartesian frame of reference. To further this effect, the room is hexagonal in the vertical plane as well as the horizontal plane, resulting in slanted walls tapering at the floor and ceiling. The visual boundaries of the room are defined using acoustically transparent panels having a neutral gray color, allowing for the structural and acoustic design of the room, as well as all of the loudspeakers, to be hidden from view.

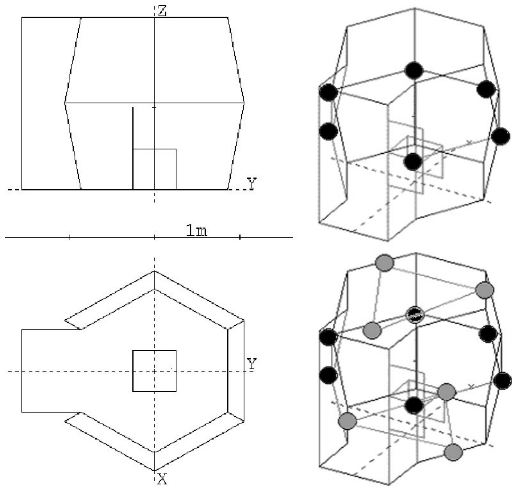

FIG. 1. Listening room visual surfaces indicating 3-D hexagonal structure. Locations for the 2-D and 3-D arrays are shown. The 1-D consists of solely the front pair of loudspeakers to generate a stereo pair. A chair is included for reference.

The reproduction system provides for 13 channels of discrete playback, including a low frequency subwoofer for frequencies below 100 Hz. Small high quality loudspeakers (low frequency roll-off at ∼100 Hz) are suspended on a pipe grid that encircles the room and extends from floor to ceiling behind the visual screens. The subwoofer (flat response to 20 Hz) is placed in one corner of the room. Six speakers are located at seated listening level at the corners of the hexagon. The additional six are placed in two sets, three at ceiling level and three at floor level, corresponding to ±49° on alternating sides of the hexagon. This provides slightly reduced coverage for elevation sounds and full horizontal coverage in the listening plane. The level of the speakers was carefully adjusted to achieve a flat frequency response across the crossover frequency of 100 Hz. The 12 full range speakers were time and level aligned at the center of the listener position.

The result is a room far from the ''standard'' listening room, being in direct contrast to recommendation ITU-R BS.1116-1 for multi-channel sound systems (ITU-R, 1997). The area is one-third the minimum area, the reverberation time is one-half the prescribed value, and the room geometry contradicts the rectangle/trapezium prescription. However, the recommendation only prescribes for a multi-channel loudspeaker array conforming to the 5.1 format. While suitable for evaluating various audio processing techniques, it is not clear that the ''standard'' listing room is suitable for more specific situations such as psychoacoustic testing on individual subjects or more complex sound scenes such as outdoor material, where low reverberation times and abstraction from the listening room are necessary.

III. EXPERIMENT I: URBAN SOUNDSCAPES IN 2D AND 3D

A. Method

27 subjects with normal hearing, aged between 23 and 59 participated in the experiment. They were expert listeners, either studying or working in the field of acoustics. The participants served without pay.

The stimuli were five urban Parisian soundscapes selected from a list of places previously identified as representative of city noises by Maffiolo (1999). Live recordings were used rather than synthesized source material to fully capture complex spatial sound scenes and focus on the ‘‘you are there’’ approach to sound reproduction according to the concept of ecological validity, developed by Gibson (1979). Indeed, the familiarity of the sound material, together with the instructions given to ease the required process of abstraction, enabled the subjects to treat the stimuli with cognitive processes elaborated in real-life situation. The test samples were 45 to 60 seconds long. The B-format files were decoded using the full in-phase decoding scheme without shelf filtering (Furse, 2003). The test configurations were the 2-D (6-channel) and 3-D (12-channel) arrays with and without the subwoofer (x and x.1, following the familiar 5.1 convention). Configurations were equalized in level at the center of the listening position using a reverberant room recording of white noise decoded over each system. The subwoofer channel content was identical between 2-D.1 and 3-D.1 configurations and level matched to provide a flat frequency response over the crossover region.

B. Procedure

Subjects were presented with a reproduction of the same sound scene over four different systems, randomly ordered. Instructions were given to subjects to direct their response strategy towards everyday listening situations, so that they would react, to some extent, as if there were in the actual situation, i.e., in an ecological valid way (Gibson, 1979), rather than in the abstract situation of a laboratory experiment. For each sound example, a free verbalization task and a multiple comparison task were conducted: subjects listened to the four reproduction methods as many times as desired and were asked to freely describe the four versions, choose which one(s) sounded the most like their everyday experiences, and justify their choice. This elicitation method, used in previous studies to investigate the sound quality of complex auditory scenes (Maffiolo, 1999, Dubois, 2000, Guastavino and Cheminée, 2003), was chosen to identify perceptually relevant features without constraining the answers into predefined categories. More specifically, subjects were not instructed to focus on spatial attributes. The nature of the test and the details of the reproduction systems used were not disclosed to the subject prior to the test.

C. Analysis of the verbal data

A semantic analysis was conducted on the spontaneous descriptions of recreated acoustic environments. A total of 512 phrasings were classified in semantic categories emerging from free verbalizations. The verbal data was lemmatized, i.e., inflectional and variant forms of a word were reduced to their lemma: their base form. Synonyms were grouped together, as well as linguistic devices constructed on the same stem (e.g., ‘‘bright,’’ ‘‘brightness’’). Lexical devices belonging to the same semantic field as indicated in a French thesaurus (Péchoin, 1992), were grouped into semantic themes. Semantic themes with fewer than 3 occurrences were excluded from the analysis. Two coders independently combined semantic themes into larger semantic categories relating to presence/immersion, readability of the scene/sense of space, distance to the scene, timber, stability, localization, and hedonic judgments (e.g., ‘‘annoying,’’ ‘‘pleasant’’). Finally, all occurrences in each category were counted.

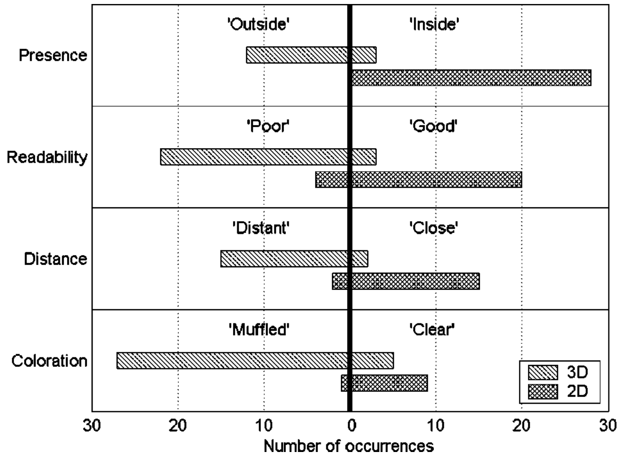

FIG. 2. The number of occurrences of spontaneous descriptions for the reproduction methods within different discriminating categories (2-D and 3-D). Opposing terms are represented on opposite sides of the graphs.

D. Results

The results of the comparison test show a strong preference for the 2-D configurations over other methods. Total results for the ‘‘naturalness’’ selection for the four reproduction setups were 62(2D), 45(2-D.1), 42(3-D), and 20(3-D.1). The number of occurrences for each reproduction method within discriminating semantic categories, namely presence, readability, distance, and coloration is presented in Fig. 2. It is interesting to note that nonspatial attributes were spontaneously evoked, although only the spatial presentation varied. The 2-D configurations (2-D and 2-D.1) were spontaneously described in the open questionnaires as very enveloping, spatially well defined, and providing a good sense of immersion in the scene, equating to a high degree of presence. Furthermore, subjects mentioned that the sound field reproduced by the 2-D systems sounded close to them. The 3-D configurations (3-D and 3-D.1) on the other hand were described as poorly enveloping and sounding farther away from the listener. Subjects indicated that space was poorly defined and indistinct. Regarding timber, the sound field recreated by the 3-D configurations was described as ‘‘muffled.’’

The 2-D.1 and 3-D.1 configurations were described as rich and too rich in low frequencies (31 and 39 occurrences, respectively), but were chosen for realism in the traffic noise

TABLE I. Details of the six soundscapes used in Experiment II.

Name	Description	Recording position
Train	Announcement on a train (small enveloping scene)	Seated within train car
Market	Walk in an open-air market (many sources at various distances)	Walking head height
Symphony	String orchestra (position close to conductor)	Above and behind conductor position
Organ	Organ music in a very large reverberant cathedral.	Well into the reverberant field
Traffic	Urban traffic noise (many sources at various distances and levels)	Corner of intersection
Improvised music	Modern improvisational music with organ, percussions, and wind instruments in a large reverberant cathedral (same as organ).	Well into the reverberant field

recordings. A further analysis of the comments suggested different ratings or different sound examples, depending on the relevance of the low frequency information in the scene (meaningful in traffic noise to recreate the rumbling of heavy vehicles, meaningless in pedestrian areas where no low frequency events are ''expected'' to occur regardless if it is actually present or not).

No distinction between the different configurations could be established on the basis of descriptions of stability of the image, localization, or hedonic judgments. The distinction between 2-D and 3-D configurations relies mainly on spatial attributes, but in an unexpected way. The 2-D systems provide a better feeling of presence and spatial definition and a closer image than the 3-D systems.

Relevant criteria for the perceptive evaluation of complex soundscapes were identified by considering semantic categories with the greatest number of occurrences. Six parameters were derived from the linguistic analysis: readability, presence, distance, localization, coloration and stability of the image. Experiment II was designed to evaluate multichannel spatial reproductions along these parameters on a wider range of auditory scenes.

IV. EXPERIMENT II: VARIOUS SOUNDSCAPES IN 1D, 2D, AND 3D

A. Method

26 subjects with normal hearing, aged between 23 and 62 participated in the experiment. They were expert listeners, either studying or working in the field of acoustics. All the participants served without pay.

The stimuli were recordings of six different soundscape excerpts as described in Table I, providing a wide variety of scenes. The decoding configuration was slightly altered, following observations and comments obtained from Experiment I, such that the B-format recordings were decoded using a 60% in-phase decoding scheme (comparable to a hyper-cardioid directivity pattern) without shelf filtering. This decoding option was seen as an improvement over the configuration in Experiment I as it provided the best compromise between localization of sources and sensitivity to listening position in preliminary listening tests. In addition, the low frequency level was adjusted to better compensate for the response of the microphone. The subwoofer channel content was identical between all three configurations. The test samples were 13 to 36 seconds long, and the subjects could listen to them as many times as desired.

The test configurations were 1-D (2.1), 2-D (6.1), and 3-D (12.1) arrays, all equalized in level at the center of the listening position. The subwoofer was included in all sound samples as it has been shown in Experiment I that the low frequency channel contributes to realism. Subjects were asked to compare perceptual differences between the three randomly ordered versions.

It should be noted that the 1-D, or stereo, configuration was not a simple 2.1 channel system. Due to the very low reverberation time in the listening room, the acoustics were deemed too dry for standard stereo. To present stereo in a more typical and favorable condition, a virtual listening room was utilized. The concept for this approach was to create a computer model of a good listening room (following a LEDE design with diffusion) using CATT-Acoustic, a geometric room acoustic simulation software. The virtual room had a mid-frequency reverberation time of 0.2 sec. The stereo speakers were placed in the model at the correct locations relative to the listener and 10 hyper-cardioid microphones were placed at the positions of the remaining speakers, pointing away from the listener. The predicted impulse responses from the virtual microphones were convolved with the

TABLE II. Perceptual parameters with extreme values as presented (translated from the original French) in Experiment II.

Parameter	Additional description	Left limit (−)	Right limit (+)
Readability ⟨Lisibilité⟩	Spatial definition, readability of the scene	Well defined	Poorly defined
Presence ⟨Présence⟩	Sense of ''being there,'' feeling of being	Inside	Outside
Distance ⟨Distance⟩	The auditory scene sounds ...	Close	Distant
Localization ⟨Localisation⟩	Localization of the sources/precision of the image	Precise	Indistinct
Coloration ⟨Coloration⟩	Spectral coloration/timber	Muffled	Clear
Stability ⟨Stabilité⟩	Stability/sensitivity to head movements	Stable	Unstable

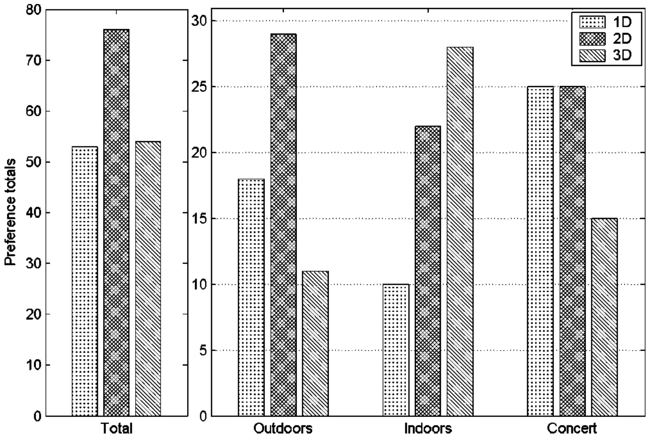

FIG. 3. Naturalness responses for the 3 types of soundscape.

B-format 2-channel decoded signal, thus creating a 12.1 channel simulation of a 2.1 system reproduced in a good listening room. This method avoided the use of B-format synthesis for the room, maintaining the separation between systems. No negative effects were reported, and the system was described as a very natural stereo reproduction.

B. Procedure

For each sound example, subjects were asked to listen to the three reproduction methods, freely describe the three versions, choose which version(s) sounded the most like their everyday experiences and justify their choice, as in Experiment I. Following this, the six parameters (readability, presence, distance, localization, coloration, and stability of the image) were presented, in random order, with slider bars corresponding to each of the three samples for comparative judgments. An optional open questionnaire for each also existed for comments or explanations of perceptions. The semantic scales for this test, and their extreme values as presented on the slider scales, were derived from the spontaneous descriptions collected in Experiment I. These are listed in Table II.

C. Results

1. Naturalness

General results of Experiment II, as shown in Fig. 3, show a subjective impression of a more ''realistic'' or ''natural'' representation of the soundscape using the 2-D system versus the other systems. A more detailed analysis shows that the subjective ratings depend heavily on the soundscape.

For concert scenes, where clarity and precise localization of the instruments would be expected, the 1-D and 2-D systems were equally selected. We believe that modern listening habits also accounts for the choice of the 1-D system, as many subjects often listen to music on a stereo set-up and are thus inclined to choose this familiar configuration as seeming ''natural.'' For complex outdoor environments, where the sounds are expected to be surrounding at the level of the listener, but also with precise locations of the numerous sources, the 2-D system was selected, confirming the results of Experiment I on the reproduction of urban soundscapes. For indoor environments, where the sounds are expected to be surrounding and coming from above (announcement in the train, organ in the cathedral), the 3-D system was selected. For this grouping analysis, the Organ example was classified in the ''Indoors'' category rather than ''Concert'' due to the fact that the verbal data suggest that subjects paid greater attention to the room effect of the church than to the musical content. It should be stated that the Organ recording was made in the far reverberant field of the instrument, as noted in Table I.

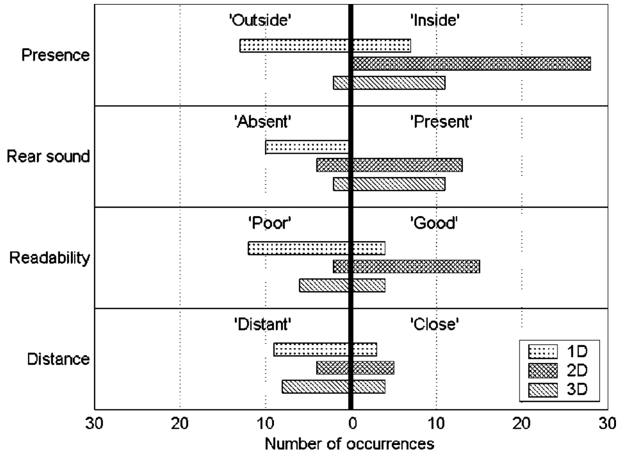

FIG. 4. The number of occurrences of spontaneous descriptions for the reproduction methods within different discriminating categories (1-D, 2-D, and 3-D). Opposing terms are represented on opposite sides of the graphs.

2. Analysis of the verbal data

A total of 453 phrasings were classified in semantic categories emerging from free verbalizations using the same linguistic analysis as in Experiment I. Free descriptions were classified in semantic categories relating to the spatial distribution of the sound, presence, realism, readability, spectral balance, and localization. The number of occurrences for each reproduction method within discriminating categories, namely presence, readability, rear sound, and distance is presented in Fig. 4.

As regards perceptive evaluation, two major distinctions were established. The first one distinguishes the 1-D array from the 2-D and 3-D arrays on the basis of spatial distribution of sound. The 2-D and 3-D configurations were described as providing sound all around the listener, including behind and above the listener, as opposed to the 1-D configuration, which was spontaneously described as frontal. The second distinction, isolating the 2-D set-up, was observed on the basis of presence. The 2-D configuration was described as providing the most immersive environment.

3. Interaction between semantic scales

Cross-correlations were computed for every possible pair of variables over all ambiances. Results indicated a correlation between readability and localization for all three reproduction methods ($r^2=0.28$, $p=0.01$) as well as between presence and distance ($r^2=0.23$, $p=0.02$). The analysis of verbal comments confirmed these interactions: 14 comments indicated that an immersive scene sounds close, and 6 comments indicated that sources can easily be located in a spa-

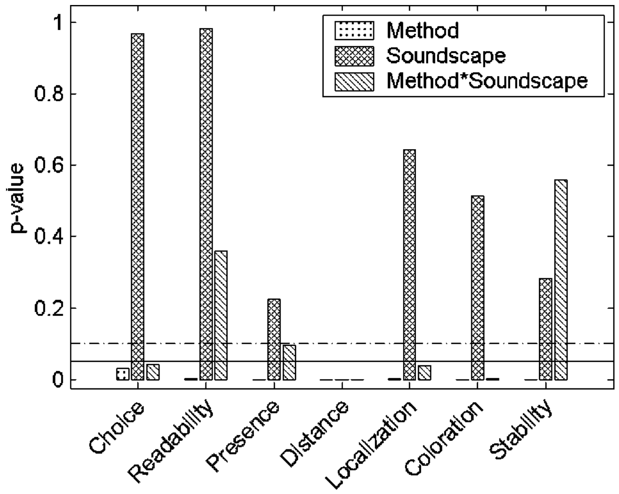

FIG. 5. ANOVA results (p-value) for a multivariate analysis of subjective parameters. Significant effects, evident from low p-values, of the method were observed for all variables : choice [$F(2,25)=3.58$], readability [$F(2,25)=6.0$], presence [$F(2,25)=162.7$], distance [$F(2,25)=43.6$], localization [$F(2,25)=6.5$], coloration [$F(2,25)=16.7$] and stability [$F(2,25)=41.67$]. A significant effect of soundscape was observed on distance ($F=6.5$). Significant effects of method*soundscape were observed for choice ($F=1.9$), distance ($F=8.2$), localization ($F=2.0$) and coloration ($F=3.2$).

tially well defined environment. The verbal data further suggested an interaction between distance and coloration, with 12 comments associating ‘‘muffled’’ with ‘‘distant,’’ or ‘‘clear’’ with ‘‘close.’’

4. ANOVA

A three (reproduction methods) by six (sound samples) by seven (variables) ANOVA on the ratings was calculated. The Green–Greenhouse correction was used for a violation of the sphericity assumption. The main effect of the reproduction method on the seven variables was significant ($p<0.05$) for all sound samples. Figure 5 presents the F and p values with regards to the relevance of method, soundscape, and the combination of method*soundscape for each variable.

The results show that the responses to all variables are strongly linked to the reproduction method. Aside from the distance parameter, all responses were invariant with regards to soundscape. Finally, there was an evident correlation between choice of the most ‘‘natural’’ method and the specific soundscape. This correlation was also seen for three other parameters: coloration, localization, and distance. To a lesser extent this correlation existed for presence. Significant effects of method and method*soundscape were observed on both spatial and timbral attributes. Gabrielsson, Rosenberg, and Sjögren (1974) also found a significant effect of both method and soundscape and a significant interaction between the most ‘‘true-to-nature’’ reproduction (monophonic reproduction on different loudspeakers) and sound samples (different music sections).

Post-hoc analyses for the present study were conducted using Bonferroni’s comparison tests. Concerning the binary variable of choice, results indicated a strong tendency ($p=0.07$) for subjects to select the 2-D set-up rather than the other two configurations (1D and 3D). Similarly, the 2-D array was evaluated as providing a higher degree of readability, i.e., a more readable presentation of the sound scene, than the other two ($p=0.05$).

Results concerning the variables of presence and distance confirmed the counter-intuitive subjective judgments observed from the verbal data in the first experiment. Indeed, the 2-D set-up was again considered as more immersive and producing a closer auditory scene than the 3-D array ($p=0.01$ for presence and $p=0.05$ for distance). But the sound field recreated by the 1-D configuration was judged even less immersive and farther away ($p=0.01$ and $p=0.05$, respectively). Concerning localization and coloration, the 3-D reproduction was perceived as indistinct and muffled in comparison to the 1-D and 2-D reproductions, which were described as clearer and more precise ($p=0.05$ for localization and $p=0.01$ for coloration). Finally, the auditory recreation by the 1-D configuration was evaluated as more stable than the other two when the listeners moved away from the sweet spot ($p=0.01$).

The main effect of soundscape was significant for the perceived distance only. Along this variable, the three reproduction methods were ranked from ‘‘close’’ to ‘‘far’’ in the (2-D|3-D|1-D) order for all ambiances but the Organ, for which the order (1-D|2-D|3-D) was observed. An analysis of the verbal comments indicate that the 1-D set-up recreated a more direct frontal sound of the organ and less reverberated room effect than the 2-D and 3-D set-up, thus making the listener feel closer to the instrument.

5. Variations between sound scenes

To further examine the effect of soundscape on the various parameters, a statistical summary of the judgments is presented in Fig. 6, showing the responses for coloration, presence, and distance for each soundscape separately. While all recordings were made with the same microphone model (and all but the Symphony excerpt were made with the same physical microphone) and processed in an identical manner, there is a noticeable variation in coloration judgments between soundscapes. This indicates a potential bias in subjective evaluations of coloration as the responses are based upon the expectations of signal content, and not necessarily on the actual content. This effect is further complicated by the variations between method*soundscape which indicate that the spatial distribution of timbral information, and its expected distribution, is linked to coloration judgments.

The presence parameter judgments show the clear distinction between the methods, regardless of soundscape. Finally, the distance parameter shows the same general trend between soundscapes, but the distribution dependence on soundscape is interesting. For example, there is little variation for the Organ sample, while for the Improvised music excerpt there is a strong variation. This is most interesting as the microphone placement is identical and the Improvised music excerpt contains the same organ (with additional instruments at the same location), though playing a modern improvisational piece. For the Organ piece, verbal comments indicate that all subjects expected a large reverberant space, namely a church, whereas for the Improvised music excerpt,

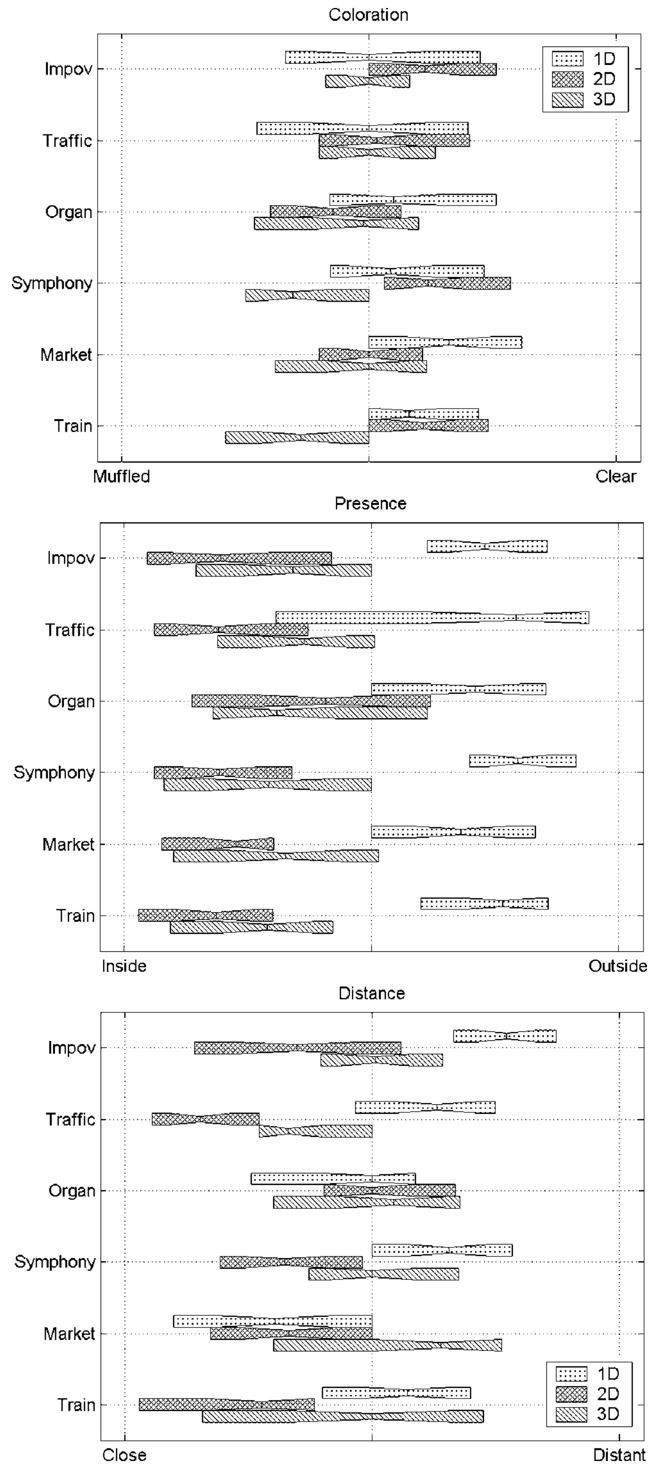

FIG. 6. A summary of slider parameter responses for the parameters ‘‘Coloration,’’ ‘‘Distance,’’ and ‘‘Presence’’ for each soundscape and the three different spatial presentation methods. Data shown as whisker plots spanning from the lower to upper quartiles, with the narrowest point identifying the median. The effect of soundscape on perceptive judgments is clear, as indicated in the ANOVA analysis.

no specific architectural configuration was expected, as it could have been recorded in different places (concert hall, studio), at various distances from the instruments. This indicates that a change in style and content of the audio information can affect the perceived distance of the events. Gabrielsson (1979) stated that interactions between parameters (and/or method) and sound material are due to ‘‘physical interactions,’’ i.e., differences in the physical properties of the sound samples. The present results suggest that such interactions can also be attributed to cognitive attributes, such as semantic content of the sound samples and subjects’ expectations.

6. Principal component analysis

The subjective ratings obtained using the sliders provide information regarding the perceptual differences in the various methods according to the six parameter questions. In an attempt to reduce the complexity of the data space the Principal Component Analysis (PCA) reduction method is used. This technique is highly suitable for reducing the dimensions of a complex space into a smaller number of orthogonal dimensions, which are composed of a linear combination of the initial parameters. PCA analysis is commonly used in psychoacoustics to investigate sound quality attributes (Kahle, 1995; Susini, McAdams, and Winsberg, 1999), and has previously been used in particular to study the perceptual evaluation of sound reproduction systems (e.g., Eisler, 1966; Gabrielsson and Sjögren, 1979; Zacharov and Koivuniemi, 2001).

The PCA analysis on the slider dataset presents an orthogonal data space as described in Table III. Using the PCA projection, 74% of the variance of the responses can be explained using the first three components, and 84% with the first four. Projections of the subjective responses to the six parameters into the space defined by the first four components of the PCA are presented in Fig. 7 (see Table II for +/− direction definitions of parameter vectors). From this analysis, it is possible to examine perceptual differences between the three spatial reproduction schemes.

The projection plane defined by PCA1×PCA2 shows a clear separation between (1-D) and (2-D and 3-D) presentations. The 1-D is more ‘‘distant’’ and ‘‘outside’’ while being more ‘‘stable,’’ as shown by the apparent data clustering separations along the +distance, +presence, and −stability vectors. In addition, 1-D and 2-D are more ‘‘clear’’ in coloration than 3-D. Finally, 3-D is more ‘‘indistinct’’ and ‘‘poorly defined’’ in reference to the 1-D and 2-D presentations. The projection planes defined by PCA1×PCA3 and PCA2×PCA3 show similar tendencies. There is an evident correlation between ‘‘localization’’ and ‘‘readability’’ in both planes. The projection plane defined by PCA1×PCA4 shows a separation of ‘‘localization’’ and ‘‘readability’’ with 1-D being more ‘‘poorly defined’’ than 2-D and 3-D, with localization becoming more precise when going from 1-D to 3-D to 2-D.

To summarize, there is a noticeable difference between 1-D and (2-D and 3-D) in terms of perceived distance, presence, and stability. The judgments for the 3-D representation fall between the 1-D and 2-D method values for all parameters but coloration.

Similar PCA analysis studies have been performed which showed correlations between Sense of space, Sense of depth, and Sense of movement, and with these three attributes loading positively Preference (Zacharov and Koivuniemi, 2001). It was also found that Penetration and timbral Emphasis were negatively correlated to Preference. It is un-

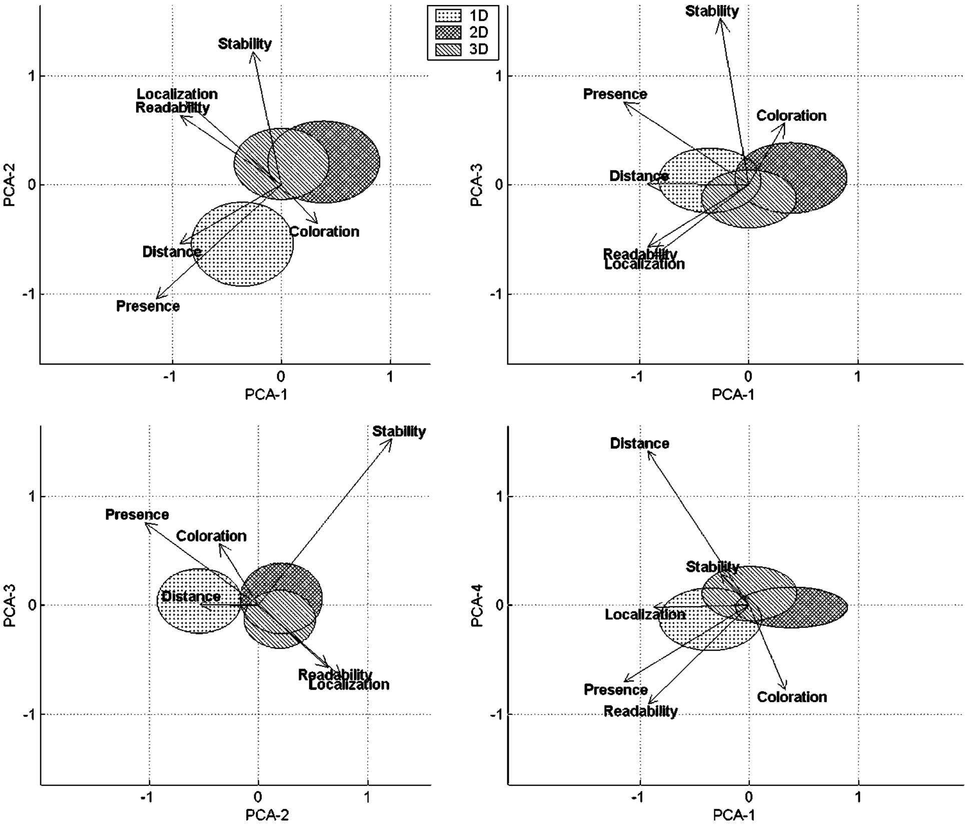

FIG. 7. Projection of subjective parameter judgments onto a PCA reduced space. Plates show the three orthogonal views for the space created by the first 3 PCA components and the projection of the space created by the 1st and 4th component. The projected data is represented by an ellipse spanning the spread of the data (using interquartile range to provide a robust estimate). The projections of the subjective parameters used in the construction of the PCA space are also presented (lengths multiplied by a factor of 2 to improve legibility). Arrows are in the direction of the $right(+)$ limit of the slider scale, as presented in Table II.

clear from the citation the exact meanings of these parameters, but we have noted that the original term for Penetration (Pistävyys) can also be translated from the original Finnish as Piercing, and may therefore help explain the evaluation of this parameter toward a negative judgment.

D. Discussion

1. Spatial attributes

Traditionally, quantifying perceptual attributes involves rigorous subject training to minimize differences among subjects and to identify small differences between parametrized stimuli. However, in the absence of clearly identified subjective dimensions for spatial sound perception, a free exploratory approach was considered more appropriate to allow subjects to define their own attributes rather than impose predefined factors of interest. An experimental protocol was designed to elicit relevant features by analyzing spontaneous verbal descriptions without constraining the answers into categories predefined by the experimenter. Interestingly, most of the semantic scales derived from the analysis of spontaneous descriptions are similar to those tested in other spatial sound reproduction studies (Berg and Rumsey, 1999; Zacharov and Koivuniemi, 2001) although using different sound material, multi-channel configurations, and methodology.

Berg and Rumsey (1999) used the Repertory Grid Technique developed by Kelly (1955) to elicit a structure of perceptual features from free verbal descriptions of perceived similarity and dissimilarity between various spatial reproduction systems. Four perceptual attributes relating to spatial features were identified: naturalness (authenticity, feeling of presence), source localization (width and lateral positioning), envelopment (positioning of the sound field relative to the subjects), and depth (ability to perceive different distances to the sources). These attributes seem to be related to the parameters of naturalness, localization, presence, and readability derived from Experiment I. Berg and Rumsey (2001) further validated these attributes with a new group of subjects listening to new stimuli. These results were also extended from stimuli differing in modes of reproduction to stimuli recorded with different surround sound microphones tech-

TABLE III. Principal Component Analysis data reduction results for slider parameters. Data indicates the linear weighting components of each parameter in constructing the new orthogonal data space. Values are also presented indicating the percentage of variation in the data which can be explained by each principal component. The major contributions for each component are indicated with an*.

	Coloration	Presence	Readability	Localization	Stability	Distance	% Explained
PCA-1	0.16	−0.58*	−0.46*	−0.44*	−0.13	−0.47*	34.7
PCA-2	−0.18	−0.52*	0.32	0.38	0.61*	−0.27	26.7
PCA-3	0.28	0.38	−0.28	−0.33	0.76*	0.01	12.5
PCA-4	−0.39	−0.35	−0.45*	−0.01	0.14	0.71*	10.2
PCA-5	0.82*	−0.22	−0.11	0.44*	−0.05	0.28	9.1
PCA-6	0.21	−0.28	0.62*	−0.60*	0.03	0.36	6.8

niques (Berg and Rumsey, 2002). As regards interactions between attributes, the strongest correlation was observed between naturalness and presence (Berg and Rumsey, 2001), in agreement with our findings.

The parameters used in this study can also be compared to the 12 attributes elicited by Zacharov and Koivuniemi (2001) through guided discussion as follows: Sense of space, Sense of depth, Sense of directions, Sense of movement (all four similar to ‘‘readability’’), Penetration (or piercing, as a negative quality), Distance to events (‘‘distance’’), Broadness (similar to ‘‘localization’’), Naturalness (the ‘‘choice’’ parameter), and four timbral attributes (‘‘coloration’’) Richness, Emphasis, Tone color, and Hardness.

It is encouraging to note that a certain consensus begins to emerge in the field of spatial sound reproduction for perceptual attributes relating to spatial features, although the semantics of these terms vary across languages and may give rise to different interpretations (for a review of terminology and meanings of spatial attributes, cf. Rumsey, 2002). However, results suggest that these attributes are not independent dimensions as interactions between factors were observed in the present experiments as well as in other sound quality evaluation studies (Gabrielsson, 1979; Susini, McAdams, and Winsberg, 1999; Zacharov and Koivuniemi, 2001). The diversity of spontaneous descriptions of the systems and the interdependency between perceptual attributes suggest that sound quality is a complex concept aggregating various physical properties (spatial and spectral) and semantic features such as judgments of pleasantness.

2. Overall quality

Results of the linguistic exploration of free responses suggest that presence and readability play an important role in the evaluation of the overall sound quality of reproduction methods. Furthermore, the most frequently selected configuration was evaluated as providing a significantly stronger feeling of presence and better readability of the sound scene. However, a significant interaction was observed between choice of the reproduction method and soundscapes. Logistic regression procedures have been tried to model the choice as a function of the parameters, but the weights differ significantly between different methods and soundscapes, further suggesting that the selection of a universally optimal reproduction method remains difficult

V. CONCLUSION

The approach presented here brings together methodological tools derived from psycholinguistics and statistical analyses to investigate spatial quality for reproduced sound. In Experiment I, relevant criteria for sound quality were identified by means of linguistic analysis of spontaneous verbal descriptions. This exploratory study of verbal descriptors resulted in six parameters: presence, coloration, readability, timber, localization, and stability of the image. In Experiment II, three configurations (1-D, 2-D, and 3-D loudspeaker arrays) were evaluated using scale judgments and free responses along these parameters on a wider range of auditory scenes. These results of the statistical analysis are in agreement with the analysis of the verbal data and help provide a clear method for interpreting the perceptual variations of the reproduction systems.

Results of the perceptive evaluation can be summarized as follows. The 1-D (traditional 2-channel stereo) configuration was characterized as providing precise localization in a frontal image, stable with regards to head shifting, but distant from the listener and spatially poorly defined. The 2-D configuration (a periphonic horizontal 6-channel circular array), on the other hand, was judged as providing a very immersive and spatially well defined environment, but less stable relative to head shifting. The judgments for the 3-D configuration (a 12-channel spherical array) interestingly fell between the 1-D and 2-D method values for all parameters but coloration and localization. The 3-D configuration was characterized by a salient ‘‘muffled’’ coloration and a poor localization.

As regards sound quality, results suggest that presence and readability make a strong contribution to overall sound quality of reproduction methods. However, the selection of a universally optimal reproduction method remains difficult, as naturalness depends highly on the sound material. Indeed, the 3-D configuration appeared to be more adapted to indoor environments, the 2-D configuration to outdoor environments, and the 1-D configuration to frontal musical scenes, though the choice of the 1-D for musical scene can possibly be attributed to it resembling a home listening environment and not necessarily the live performance environment. Furthermore, interactions between parameters were observed, consistent with other perceptual evaluation studies.

In similar experiments, Guastavino (2003) observed that the choice of reproduction methods differed for different groups of subjects. Several recordings of indoor and outdoor material were carried out using simultaneously a Soundfield microphone, binaural microphones on a dummy head, and a set-up of five noncoincident microphones. A multiple comparison task was carried out on three groups of subjects: sound engineers, acousticians, and nonexperts. When asked to select which recording sounded more like their everyday experiences, audio engineers gave greater attention to the localization and precision of the sources, whereas the other two groups based their selection on presence and spatial distribution of sound. Similarly in the present study, a conflict was observed between precise localization (with the 1-D configuration) and presence (with the 2-D configuration), leading to different choice strategy among subjects. Similar differences were already observed by Gabrielsson (1979) for monophonic reproduction. When comparing various reproductions for similarity, experts based their judgment on ‘‘brightness’’ rather than ‘‘loudness,’’ while nonexperts tended to do the opposite. Furthermore, the reproduction method must be well suited for the tasks of the listening test. Guastavino, Katz, Polack, Levitin, and Dubois (submitted) showed that stereophonic reproduction was ecologically valid for source identification tasks, but not for processing complex auditory scenes in a global manner. It was further shown that a multichannel reproduction was necessary to enable subjects to process urban soundscapes in laboratory conditions as they would in real life situation.

Most relevant to the general notion of sound quality is the observed gap between ‘‘objective’’ physical accuracy and ‘‘subjective’’ perceived naturalness. Indeed for most auditory scenes, the 2-D configurations were judged by the participants as more natural and realistic than the 3-D configurations although spatially incomplete, thus indicating potentially negative effects linked to providing ‘‘too much’’ information. These findings underline the difference between illusion and accuracy pointed out by Rumsey (2002): the illusion of ‘‘being there’’ is not necessarily related to true spatial fidelity. This counter-intuitive observation, from a physical point of view, indicates the importance of considering subjective psychological attributes in the evaluation of perceived sound quality. Furthermore, the lack of preference for 3-D configurations could be explained by the unfamiliarity with 3-D audio reproduction, although the natural world is always present in 3D. As surround sound systems become more common, 2-D audio reproduction systems may sound more familiar and thus more ‘‘natural’’ than 3-D configurations, which are not widely used. The results reported here suggest a shift from physical descriptions to cognitive ones in exploratory studies, to identify relevant perceptual features and better understand how acoustic phenomena are perceived and cognitively processed before addressing physical parameters in more controlled experiments [for a further discussion on this point, see Dubois (2000) and Guastavino and Cheminée (2003)].

Together, these findings underline the fact that different applications give rise to different sound quality criteria. The appropriate choice of a reproduction method must take into account the type of sound samples (music, indoor, or outdoor material), the type of application (task, entertainment), and even the expertise of the audience.

ACKNOWLEDGMENTS

The authors would like to thank the following for contributing to the soundscape material: Jeff Silberman for providing the symphony recording; Christophe d’Alessandro and the Eglise Sainte-Elisabeth, Paris, for providing the material and location for the Organ and Improvisation Pieces; and IRCAM for providing the Soundfield microphone. Thanks also to Holger Strauss for the command line audio utility ASIOPLAY which allowed for the playing of multichannel audio from within MatLab. This work was funded by the French Ministries of Environment (equipment) and Research (funding for C. Guastavino’s doctoral work conducted at the Université Paris 6) and a Chateaubriand Fellowship (B. Katz), Office for Science and Technology of the Embassy of France in the United States. The writing of this paper was supported by a CFI/VRQ grant attributed to Daniel Levitin (McGill University). The authors would like to thank Danièle Dubois and Jacques Poitevineau for fruitful discussions during the research and Daniel Levitin for helpful comments during revisions of this manuscript. Experimental work performed while at the Laboratoire d’Acoustique Musicale (Université Paris 6, CNRS, Ministère de la culture), 11 rue de Lourmel, 75015 Paris, France.

Beranek, L. L. (**1962**). *Music, Acoustics and Architecture* (Wiley, New York).

Berg, J., and Rumsey, F. (**1999**). ‘‘Spatial attribute identification and scaling by repertory grid technique and other methods,’’ *Proceedings of the 16th AES International Conference on Spatial Sound Reproduction*, Audio Eng. Soc.

Berg, J., and Rumsey, F. (**2000**). ‘‘Correlation between emotive, descriptive and naturalness attributes in subjective data relating to spatial sound reproduction,’’ *Proceedings of the 109th AES Convention*, Los Angeles, 22–25 September, preprint 5206, Audio Eng. Soc.

Berg, J., and Rumsey, F. (**2001**). ‘‘Verification and correlation of attributes used for describing the spatial quality of reproduced sound,’’ presented at the *AES 19th International Conference: Surround Sound Techniques, Technology and Perception*, 21-4 June, Schloss Elmau, Germany, Audio Eng. Soc.

Berg, J., and Rumsey, F. (**2002**). ‘‘Validity of selected spatial attributes in the evaluation of 5-channel microphone techniques,’’ presented at the *112th AES Convention*, Munich, Germany, 11–14 May, Paper 5593, Audio Eng. Soc.

Daniel, J. (**2000**). ‘‘Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia (Acoustic field representation, application to the transmission and the reproduction of complex sound environments in a multimedia context),’’ Ph.D. dissertation, Université Paris 6.

Dubois, D. (**2000**). ‘‘Categories as Acts of Meaning: The Case of Categories in Olfaction and Audition,’’ Cogn. Sci. Qu. **1**, 35–68.

Eisler, H. (**1966**). ‘‘Measurement of Perceived Acoustic Quality of Sound-Reproducing Systems by Means of Factor Analysis,’’ J. Acoust. Soc. Am. **39(3)**, 484–492.

Fellgett, P. B. (**1974**). ‘‘Ambisonic reproduction of directionality in surround sound systems,’’ Nature (London) **252**, 534–538.

Furse, Richard W. E. (**2003**). MN Audio Library, http://www.muse.demon.co.uk/ (site last visited 26-August-03).

Gabrielsson, A. (**1979**). ‘‘Dimension analysis of perceived quality of sound reproduction systems,’’ Scand. J. Psychol. **20**, 159–169.

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

Gabrielsson, A., Rosenberg, U., and Sjögren, H. (**1974**). ‘‘Judgments and dimension analysis of perceived sound quality of sound-reproducing systems,’’ J. Acoust. Soc. Am. **55**, 854–861.

Gabrielsson, A., Rosenberg, U., and Sjögren, H. (**1974**). ‘‘Judgments and dimension analysis of perceived sound quality of sound-reproducing systems,’’ J. Acoust. Soc. Am. **55**, 854–861.

Gaskell, P. S. (**1979**). ‘‘Spherical harmonic analysis and some applications to surround sound,’’ BBC Research Department Report No. BBC RD 1979/25.

Gerzon, M. A. (**1977**). ‘‘Design of ambisonic decoders for multi speaker surround sound,’’ presented at the 58th AES Convention, New York.

Gibson, J. J. (**1979**). *The Ecological Approach to Visual Perception* Houghton Mifflin, Boston, MA

Guastavino, C. (**2003**). ‘‘Étude sémantique et acoustique de la perception des basses fréquences dans l’environnemet sonore urbain (Semantic and acoustic approaches to low frequency perception),’’ Ph.D., dissertation, Université Paris 6.

Guastavino, C., and Cheminée, P. (**2003**). ‘‘Conceptualisations en langue, représentations cognitives et validité écologique: une approche psycholinguistique de la perception des basses fréquences (Cognitive and linguistic representations and ecological validity: A psycholinguistic approach to low frequency perception),’’ Psychol. Française **48**(4), 91–101.

Guastavino, C., Katz, B., Polack, J-D., Levitin, D., and Dubois, D. (under review), ‘‘Ecological validity of sound reproduction systems,’’ Acustica.

ITU-R, Recommendation BS.1116-1 (**1997**). ‘‘Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems,’’ International Telecommunications Union Radiocommunication Assembly.

Kahle, E. (**1995**). ‘‘Validation d’un modèle objectif de caractérisation de la qualité acoustique dans un ensemble de salle de concerts et d’opéras (Validation of an objective model for characterizing the acoustic quality of a set of concerts hall and opera houses),’’ Ph.D. dissertation, Université du Maine, Le Mans.

Kelly, G. (**1955**). *The Psychology of Personal Construct* (Norton, New York).

Maffiolo, V. (**1999**). ‘‘De la caractérisation sémantique et acoustique de la qualité sonore de l’environnement sonore urbain (Acoustic and semantic characterization of the sound quality of urban environments),'' Ph.D. dissertation, Université du Maine, Le Mans.

Péchoin, D. (**1992**). *Thésaurus Larousse; des Idées aux Mots, des Mots aux Idées* (Larousse, Paris).

Rumsey, F. (**1998**). ''Subjective assessment of the spatial attributes of reproduced sound,'' *Proceedings of the AES 15th International Conference on Audio, Acoustics and Small Space*, Audio Eng. Soc., pp. 122–135.

Rumsey, F. (**2002**). ''Spatial quality evaluation for reproduced sound: terminology, meaning and a scene-based paradigm,'' J. Audio Eng. Soc. **50**, 651–666.

Schroeder, M. R., Gottlob, D., and Siebrasse, K. F. (**1974**). ''Comparative study of European concert halls,'' J. Acoust. Soc. Am. **56**, 1195–1201.

Susini, P., McAdams, S., and Winsberg, S. (**1999**). ''A multidimensional technique for sound quality assessment,'' Acustica **85**, 650–656.

Zacharov, N., and Huopaniemi, J. (**1999**). ''Results of a round robin subjective evaluation of virtual home theatre sound systems,'' *Proceedings of the AES 107th International Convention*, New York, 24–27 September, Audio Eng. Soc.

Zacharov, N., and Koivuniemi, K. (**2001**). ''Audio descriptive analysis & mapping of spatial sound displays,'' *Proceedings of the 2001 International Conference on Auditory Display*, Espoo, Finland, July 29–August 1, 2001.