Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm*

FRANCIS RUMSEY, *AES Fellow*

Institute of Sound Recording, University of Surrey, UK, and School of Music in Piteå, Luleå University of Technology, Sweden

Spatial quality in reproduced sound is a subset of the broad topic of sound quality. In the past it has been studied less rigorously than other aspects of reproduced sound quality, leading to a lack of clarity in standard definitions of subjective attributes. Rigor in the physical measurement of sound signals should be matched by equal rigor in semantics relating to subjective evaluation. A scene-based paradigm for the description and assessment of spatial quality is described, which enables clear distinctions to be made between elements of a reproduced sound scene and will assist in the search for related physical parameters.

1 THE NEED FOR MEANINGFUL SUBJECTIVE EVALUATION OF SPATIAL QUALITY IN SOUND REPRODUCTION

1.1 Introduction

Sound quality is a multifaceted, multidimensional phenomenon. According to Letowski [1], sound *quality* should be differentiated from sound *character*, the former including preferential and emotive responses, but the latter supposed to be purely descriptive. Also according to him, "sound quality is that assessment of auditory image in terms of which the listener can express satisfaction or dissatisfaction with that image. Sound quality can be judged by comparing images produced by several external stimuli or by referencing a perceived image to the concept residing in the listener's memory." Sound character, on the other hand, is value-free and enables judgments to be made that simply represent differences between stimuli. This has strong similarities with Nunally and Bernstein's distinction between sentiments and judgments [2]. One might reasonably suppose that sentiments relating to sound quality are strongly determined by the experience, culture, and conditioning of a subject (and therefore will differ considerably), whereas judgments of well-defined attributes are likely to be more reliable and consistent (depending only on the sensitivity and training of the subject). The degree to which one can generalize about subjective preference or sentiment is not fully known, but it is reasonable to suppose that patterns and trends of preference exist among groups of subjects that conform to similar cultural and educational backgrounds.

Sound quality is typically treated as a composite entity in listening test standards, in the form of a mean opinion score (MOS), which conflates all aspects of sound quality, including preferences and descriptive characteristics, into a single rating. Although such standards do allow for the rating of more distinct attributes, they are rarely used in practice.

The purpose of this paper, though, is not to present a discourse on sound quality in general, but to concentrate on the specific issues of spatial quality and character in sound reproduction systems. The emphasis is on subjective analysis rather than on physical correlates of subjective variables, although some comments are made about the physical factors that have been observed to relate to various subjective attributes.

Work is proceeding in various centers to identify physical measures that relate to subjective spatial attributes, but more time is needed before a complete model emerges. It is important to ensure clarity in the definition of subjective terms before one can establish clear relationships, and it is vital that the experimental rigor expected in the physical measurement of signals be matched by equal rigor in the definitions of subjective terms. This paper, therefore, is mainly about the semantics of spatial quality.

High technical quality or fidelity, it can be argued, may be taken for granted at this point in the history of audio engineering. Although not all audio devices exhibit the highest technical quality, the technical quality of the best sound reproduction available to the consumer exhibits

* Manuscript received 2001 November 13; revised 2002 June 26.

very low levels of distortion, a wide frequency range, a flat frequency response, and low noise, with specifications that match or exceed the limits of human perception. Although improvements may still be made in these domains, the technical quality curve is becoming asymptotic to the ideal, and product development is in a region of diminishing returns.

Spatial quality and character, on the other hand, have some way to go before the curve could be said to be asymptotic to some ideal. For many years sound reproduction has been limited to only two channels, in the majority of applications. So-called binaural audio reproduction (that is, using head-related signals fed independently to the two ears) is capable of high spatial fidelity but has not been widely used commercially. Begault identified a range of challenges that should be overcome before it could be implemented successfully, and some of these are being addressed [3]. The use of head-related spatial audio signals is, however, now growing in virtual reality systems, virtual acoustics, and computer sound reproduction, including systems that reproduce such signals over loudspeakers using crosstalk canceling [4]. Surround sound, or multichannel reproduction involving more than two loudspeakers, is growing in importance and is capable of enhanced spatial quality, compared with two-channel stereo reproduction [5], but still exhibits numerous compromises. Wavefield synthesis [6] allows for accurate sound field reconstruction over a wide listening area but requires a very large number of loudspeakers and advanced signal processing. It is unlikely to be implemented widely in consumer systems for some time to come. Whatever reproduction systems are implemented, it must still be possible to generate source material in creative environments such as recording studios, and convenient methods for spatial image control are still at a relatively crude stage in their development.

The preceding leads to the inevitable conclusion that reliable methods of measuring and subjectively assessing spatial quality are required if reproducing systems, signal processing algorithms, and recording techniques are to be compared reliably. Such methods are also of vital importance in the field of computational auditory scene analysis (CASA) and its partner, virtual reality (VR) [7], in which reliable perceptual descriptors and physical correlates of spatial scene attributes are needed for parametric representation and synthesis.

Much of the extant work on subjective assessment in sound reproduction has concentrated on timbral quality and technical fidelity, tending to place spatial quality at a lower level of priority or to group its attributes under a single heading. This is probably because the focus of such studies has typically been on the quality of loudspeakers (such as in [8], [9]), where other issues were of overriding importance and spatial content could be a distraction from the evaluation of the product under test. Toole, for example, found that listeners were less critical of a loudspeaker when listening in stereo, leading him to prefer monophonic tests for critical evaluations.

This is not to say that spatial quality was unimportant at that time (in 1985 he concluded that "assessments of stereophonic spatial and image qualities were closely related to sound-quality ratings"), but until recently there have been relatively few attempts in the world of reproduced sound to isolate any more detailed spatial attributes than all-encompassing ones, such as spaciousness, spatial impression, sound stage, or stereophonic impression. Those spatial scales that have been used in listening test questionnaires often appear to have been defined by the experimenter, rather than derived from detailed elicitation experiments, and are not known to be universally meaningful or statistically independent of each other.

Spatial quality has been studied in concert hall acoustics, and there is a certain amount that can be learned from these studies in relation to reproduced sound. There are, nevertheless, a number of reasons why reproduced sound may be considered to be different from concert hall acoustics, and may benefit from consideration in its own right. Although many of the features of natural environments and spatial listening may be present in reproduced sound, there are a number of unique properties of each, and the cognitive tasks, context, and concepts involved may be somewhat different, as will be discussed and as already introduced in previous works [10], [11].

1.2 Illusion versus Accuracy

As originally expounded in [5], different applications give rise to different spatial audio quality criteria in reproduced sound. In classical music recording and other recording genres where a natural environment is implied or where a live event is being relayed, it is often said that the aim of high-quality recording and reproduction should be to create as believable an illusion of "being there" as possible. This implies fidelity to a remembered reference in terms of technical quality of reproduction, and also fidelity in terms of spatial quality. Others have suggested that the majority of reproduced sound should be considered as a different experience from natural listening, and that to aim for an accurate reconstruction of a natural sound field is missing the point—consumer entertainment in the home being the aim.

The primary aim of most commercial media production is not true spatial fidelity to some notional original sound field, although a mixing engineer might choose to create spatial cues that are consistent with those experienced in natural environments. In a large number of commercial releases there is no natural environment to imply or recreate, and one is dealing with an artificial creation that has no "natural" reference or perceptual anchor. Here the acoustic environment implied by the recording engineer and producer is a form of acoustic fiction or acoustic art. This is probably what led Nakayama et al., when identifying some subjective dimensions of multichannel reproduction of natural acoustic music recordings back in 1971 [12], to comment in relation to "nonnatural" balances such as pop music:

> Needless to say, the present study is concerned with the multichannel reproduction of music played only in front of the listeners, and proves to be mainly concerned with extending the ambience effect . . . In other types of

four channel reproduction the localizations of image sources are not limited to the front. With regard to the subjective effects of these other types of reproduction, many further problems, those mainly belonging to the realm of art, are to be expected. The optimization of these might require considerably more time to be spent in trial, analysis, and study.

Even if a reproduced spatial scene is unnatural, unfamiliar, or fictitious, it is possible to compare versions of spatial reproduction (or scene renderings in VR terms), such as might arise from using different recording techniques, forms of signal processing, or reproduction configurations. One can describe their relative quality and/or character in terms of differences in magnitudes of clearly defined attributes. It is also possible to talk in terms of desirable and undesirable, or appropriate and inappropriate, spatial qualities, although this is related closely to preference evaluation, which is a separate matter. One must also bear in mind the possibility for reproduced sound to be "hyperreal," that is, having spatial cues that are exaggerated or not naturally occurring. As virtual environments and augmented reality become more common, our concepts of naturalness may be forced to change—after all, naturalness is mainly related to familiarity.

The ability of spatial sound systems to recreate accurately localized sources is regarded by many as the "holy grail" of stereophonic reproduction, and the evaluation of perceived sound source locations is often the only consideration in subjective experiments. If true identity were possible between recording environment and reproducing environment, in all three dimensions and for all listening positions, then the ability of a recording–processing–reproducing system to render accurate images of all sources (including reflections) would be the only requirement for spatial fidelity. The need for subjective testing would be eliminated as a result, and there would be no need for a discussion such as this. True identity, however, is not currently possible, and may never be, for a variety of practical and technical reasons. Neither is it necessary to render every reflection accurately in order to obtain a perceptually convincing impression of diffuse reverberation, for example, enabling complexity reductions to be made in practical spatial audio rendering systems [13], [14]. Real spatial audio signal chains, from original source to listener, always involve tradeoffs and design compromises of one sort or another, which makes subjective testing and comparison necessary and desirable.

As an interesting aside, it may also be noted that some recent experiments seem to suggest that precise source position rendering in the spatial reproduction of music and other natural signals is not the most important spatial factor governing listener preference. At least two separate studies involving listener preference mapping have found a relatively low correlation between precise localization accuracy and preference ratings [15], [16]. This, however, requires much more study and is likely to be highly context and subject dependent.

Human scene analysis mechanisms have a tendency to group simple stimulus components into meaningful objects in order to make sense of the perceived world [17]. The spatial differences between reproduced sound scenes are typically described by listeners in terms of high-level attributes or constructs, such as scene width and depth, source width, envelopment [18], rather than in analytical terms describing the locations of direct sound and reflections associated with each sound source, as will be discussed. High-level spatial constructs are hard to define and relate to physical quantities, but they are useful "handles" on the subjective reality of individuals and may be excellent "hooks" for parametric analysis and synthesis as well as creative control of artificial spatial environments.

1.3 Product Evaluation or Classical Psychophysics?

A degree of tension may be observed between those whose primary aim is to evaluate products (such as loudspeakers, microphone techniques, signal processing algorithms, audiovisual systems, VR environments) and those whose primary aim is to study human perception mechanisms. The two fields are related, but the aims are different. In classical psychophysics relatively simple stimuli are typically used in experiments that are designed to study the workings of the human brain and its psychological functions. In product evaluation one is less directly concerned with the workings of the human brain, and subjects are used as "quality meters" in order to determine something about the product under test. Letowski [1] classifies these two forms of auditory assessment as subject-oriented and object-oriented, because the former is concerned with gathering information about the listeners themselves and the latter with information about the external world. However, if one considers the listener as just another component in the signal chain from source to receiver, then the distinction between the two paradigms becomes more difficult to justify.

Spatial audio evaluation as discussed here is concerned with the product evaluation, or object-oriented, form of auditory assessment. There is no direct intention to claim greater insight into spatial perception or cognitive processes, although useful insights may arise as offshoots of the argument. This paper is concerned with the development of reliable and valid methods for the evaluation and comparison of products, systems, and techniques that give rise to differing spatial sound quality.

1.4 Reliable Product or Technique Differentiation

In product evaluation experiments one is usually concerned with some form of comparative judgment, either between multiple products or between each product and a reference. Here one needs to develop methods that differentiate reliably and meaningfully between systems, and one looks for attributes or scales upon which such differentiation can be made, as well as suitable program material that highlights these differences.

Whereas in auditory perception experiments one typically uses simple stimuli such as tones and noise, in product evaluation it may be more appropriate to use the sort of program material for which the product will be used, such as music, speech, movie sound, and so forth. The problems

of using such material are numerous and will be discussed in more detail later, but here it is simply asserted that the ecological validity of product evaluations is not easily supported by using tones or noise as program material.

Ecological validity describes the extent to which an experimental situation matches the real-world context and circumstances it is supposed to represent. For example, numerous psychological experiments take place under highly controlled laboratory conditions that may give rise to unrepresentative human responses. Such situations could be considered to have low ecological validity. Ecological validity is similar to external validity, which relates to the validity of experimental results outside the context of the individual experiment. In psychoacoustic experiments there is nearly always a tension between ecological validity and scientific control of variables—the more tightly one controls experimental variables in order to observe individual effects, the less ecologically valid the experiment becomes. There appears to be a form of uncertainty principle at work, in that one can obtain an unambiguous result with high certainty but low ecological validity, or a more uncertain result with higher ecological validity. The more like a real-world situation the experiment becomes, the less easy it is to control all the variables. This tension is strongly evident when one tries to undertake controlled experiments comparing recording techniques.

1.5 Relationships between Spatial Attributes and Preference

As introduced before, attribute judgments and preference rating are different concepts. Letowski chooses to distinguish between global assessment and parametric assessment, the former being close to the concept of a MOS-type evaluation. He divides global assessment into the categories fidelity, naturalness, and pleasantness. He acknowledges that fidelity is a comparative judgment that relates one sound stimulus to another, possibly a reference. Naturalness can be taken as a comparison between the stimulus under evaluation and an internal reference that relates to memories of the characteristics of natural environments. Pleasantness, in his terms, is a form of preference or emotive response that grades the degree of satisfaction with a stimulus. He proposes that sound quality can be broadly divided into the categories of timbral quality and spatial quality. Toole [8], on the other hand, chooses to group ratings of sound system performance into three broad categories: fidelity, pleasantness, and spatial quality. The middle one of these is most clearly a preference attribute whereas the other two are more likely to be purely descriptive.

In [19] similar but not identical categories of responses to those mentioned were identified in a free elicitation experiment that aimed to discover attributes considered relevant by listeners when comparing different modes of spatial reproduction. A form of verbal protocol analysis enabled the grouping of elicited attributes into categories that distinguished between descriptive attributes (supposedly value-free, objective constructs) and emotive/evaluative attributes (similar to the pleasantness category). A distinct group also emerged under the naturalness heading, to some extent confirming Letowski's hypothesis.

An important aspect of spatial attribute evaluation in subjective experiments is the relationship between descriptive attributes and preference. Bech [20] explained how external preference mapping could be used to relate expert-derived descriptive data to naïve subjects' ratings of product preference. Berg and Rumsey [16] and Zacharov and Kuovuniemi [15] also showed how forms of statistical analysis could be used to establish relationships between descriptive terms and preference data from spatial audio experiments. In such a way, product designers and sound designers can begin to discover how certain spatial attributes should be optimized in different contexts in order to give rise to high consumer preference. This possibly simplistic view of reproduced sound as a consumer product such as food or wine, to be optimized according to the preferences of a naïve consumer, deserves careful consideration. Sound products, it might be argued, are "consumed" these days in similar ways to other commodities, rather than being the preserve of an elite band of cognoscenti. Although expert listeners are useful as subjects in the sensitive judgment and discrimination of clearly defined attributes, their preference judgments may not be typical of the average consumer.

2 WHAT IS A SPATIAL ATTRIBUTE?

Before proceeding much further it is important to discuss exactly what is meant by the term "spatial attribute" in sound quality evaluation. Although the meaning of the phrase may seem obvious to some, it is far from consistently represented in the literature, being open to all sorts of interpretations.

2.1 Meaning, Reliability, and Validity

When planning experiments in the human sciences, one is regularly faced with the concepts of validity and reliability in the definition of scales and attributes. In [21] it was explained that whatever the method used in psychological testing, it must stand up to the normal tests of objectivity, reliability (it should stand up to duplication), validity (measures should be seen to covary with other independent measures of the same construct or, more simply, measures should measure what they purport to be measuring), sensitivity, comparability (comparisons are possible among individuals and groups), and utility (the measure provides information relevant to contemporary theoretical and practical values).

Spatial attributes should be identified that are meaningful, in order of priority; 1) to individual subjects; 2) to a well-defined group of expert subjects forming a listening panel, and that agree upon a set of attributes to be graded; 3) to expert listeners not associated with that listening panel; 4) to independent observers or readers of the results. They should be unambiguous and preferably unidimensional (in other words, they should represent a single perceptual construct). They should enable meaningful and sensitive distinctions to be made between the products or techniques under test, and they should enable repeat-

able judgments. As will be seen, there is considerable room for any of these criteria to remain unfulfilled in subjective experiments on spatial audio reproduction.

2.2 Attributes of Spaces versus Spatial Attributes

A review of the literature relating to spatial quality evaluation in its broadest sense reveals a subtle but crucial division between two different concepts of the spatial attribute. This division, although possibly obvious to those involved, has never been clearly highlighted in the literature. Yet it seems important to this author in establishing clarity about what is to be evaluated and has partly been brought to his attention through the work of Neher, a research student at the Institute of Sound Recording [22]. Put simply, it relates to the distinction between attributes of spaces and spatial attributes. In much of the literature relating to concert hall acoustics or the acoustics of enclosed spaces, the attributes that are used to evaluate "spatial" quality are often parameters that relate to the qualities of the space in question, such as reverberance, warmth, intimacy, and so on. Zacharov and Koivuniemi [10] and Berg [18] review a number of the terms that arise from such studies, and it is clear that only some of them are really what this author would term spatial attributes, which could be related to the evaluation of sound reproduction. The most useful spatial terms that arise repeatedly in different forms in such experiments can be classed as source width and envelopment or spatial impression. (A more detailed discussion of these terms follows.)

In [23] we attempted a definition of spatial impression as the "the auditory perception of the location, dimensions, and other physical parameters of a sound source and the acoustic environment in which the source is located." This definition is not entirely satisfactory, though. In [19] we have also described the search for valid spatial attributes as being primarily concerned with "the three-dimensional nature of sound sources and their environments," which is possibly closer to the mark. Both these attempts at definitions of what is meant by a spatial attribute imply that we are concerned with those perceptual constructs that relate to directionality, size (height), depth, and width of reproduced sources, groups of sources, and acoustical environments. In other words we are concerned with describing and evaluating the three-dimensional characteristics of the components of a spatial audio scene that is reproduced using loudspeakers or headphones. This scene-based approach to spatial attribute definition is expanded upon in Section 3.

The following is an example of the conceptual difference between spatial attributes as defined in this paper and attributes of spaces as discussed by some other authors. The extensive research carried out primarily at IRCAM, resulting in the *Spatialisateur (Spat)* software package and partially incorporated into the MPEG-4 spatial audio scene description language (for example, [24], [25]), resulted in a number of perceptual parameters for "spatializing" reproduced audio scenes. They enable salient perceptual features of natural acoustical spaces to be isolated and controlled. Most of these parameters affect the acoustical characteristics of the modeled space and are only indirectly related to the spatial attributes of sound reproduction as defined before. In other words, there is rarely a direct mapping from these "virtual acoustics" parameters to what this author would call spatial attributes:

Group I: Source-related attributes and corresponding objective criteria
- Source presence: energy of direct sound and early room effect
- Source warmth: variation of early energy with frequency
- Source brilliance: variation of early energy with frequency
- Room presence: energy of late room effect
- Running reverberance: early decay time
- Envelopment: energy of early room effect relative to direct sound.

Group II: Room-related attributes and corresponding objective criteria
- Late reverberance: late decay time
- Heaviness: variation of decay time with frequency
- Liveness: variation of decay time with frequency.

Two things are interesting about these perceptual parameters from *Spat*: first that they are grouped into source- and room-related attributes (which corresponds broadly with our requirements) and second that envelopment is really the only parameter that comes close to our definition of a spatial attribute. Changes in the envelopment parameter during our informal trials appeared mainly to give rise to what we would have called changes in source width, as well as changes in timbral characteristics, when reproduced using the 3/2 stereo rendering mode. This difference suggests that there may also be issues of linguistic interpretation to consider as well as conceptual differences. The relationship between the *Spat* perceptual parameters and the examples of unidimensional spatial attributes given in the following, such as source width, environment width, and source distance, is not straightforward, suggesting a radically different conception of spatial quality.

2.3 Spatial Attributes in Existing Listening Test Standards and Earlier Work on Reproduced Sound Quality

Listening test standards such as those devised by the ITU have typically concentrated on the mean opinion score (MOS) or basic audio quality judgment that is taken to include all aspects of sound quality. Optionally, ITU-R BS.1116 [26] proposes that one can grade the following spatial attributes (with their definitions):

- *Stereophonic image quality (two-channel systems)*: attribute is related to differences between the reference and the object in terms of sound image locations and sensations of depth and reality of the audio event
- *Front image quality (multichannel systems)*: attribute is related to the localization of the frontal sound sources; it includes stereophonic image quality and losses of definition
- *Impression of surround quality (multichannel systems)*:

attribute is related to spatial impression, ambience, or special directional surround effects.

Clearly these terms are multidimensional. Although the author has had some success in using the second of these in experiments on surround sound [27], the term "impression of surround quality" (even when interpreted as simply spatial impression) was found to be too variable in its interpretation by subjects. They found it impossible to distinguish between the multiple dimensions contained in spatial impression and were confused between quality and quantity of the same.

Some of the most comprehensive studies involving subjective testing of loudspeakers were conducted by Toole in the early 1980s [8]. Here he was primarily concerned with evaluating sound quality, using scales based on the work of Gabrielsson and Sjögren (for example, [9]), but he also needed to evaluate spatial quality in some cases. In such cases he used scales that he admits were not as rigorously defined as those for other aspects of quality, but they seemed to embrace most listener comments in a pilot test. These were (with this author's comments in parentheses):

- *Definition of sound images* (stability, focus, source separation)
- *Continuity of sound stage* (a form of width homogeneity relating to the even distribution of sources across the sound stage)
- *Width of sound stage* (related to the width between outer sources on the sound stage, not including reverberation)
- *Impression of distance/depth* (the definition suggests it is in fact depth that is meant, as discussed further in Section 3.1.2)
- *Abnormal effects* (unusual or unnatural spatial effects such as phasiness)
- *Reproduction of ambiance, spaciousness, and reverberation*
- *Perspective* (graded from "you are there" through "they are here" to "artificial/contrived").

Most of these attributes are global spatial characteristics, as will be explained in Section 3, and a number of them include more than one perceptual construct. Listeners found these scales to be useful in evaluating loudspeaker spatial quality in Toole's experiments.

IEC 60268 [28] defines three factors under the heading "overall spatial quality":

- *Image localization*: perceived spatial location of a reproduced sound source. The image may be well defined or blurred.
- *Image stability*: perceived location of the reproduced sound source, may change with pitch, loudness, or timbre. It may also change as a function of listener position, head rotation, or other normal movements. If these effects are small, the image will be stable.
- *Width homogeneity*: stereophonic image should be distributed uniformly between loudspeakers.

The first of these is somewhat unclear as it is not certain what "reproduced sound source" is, whether a single source or the location of all sources or reverberation. The definition implies that the attribute is related to image focus—in other words, the degree of "locatedness" of phantom images. An earlier version of IEC 268-13 (essentially the same standard but in the old numbering system) also proposed some scales relating to spatial attributes:

- Spaciousness (closed–spacious)
- Distance (distant–near)
- Location of sources (unstable–stable)

EBU 562-3 [29] also suggests attributes that may be useful in the multidimensional evaluation of spatial reproduction, based on Japanese experiments involving multichannel sound for HDTV:

- Apparent sound stage width
- Surround effect
- Apparent room size
- Horizontal and vertical localization
- Naturalness
- Sense of reality
- Agreeableness

A small number of other experiments involving spatial quality attributes in reproduced sound were reviewed in [30], the main conclusion being that the majority of attributes used were multidimensional and often unclear in their interpretation.

2.4 Relationship of Perceived Attributes to Source Material

The spatial attributes of importance are strongly dictated by the nature of the source material and the context or task in question. First it is only valid to talk about the static spatial attributes of a reproduced sound scene when it is relatively consistent and unchanging; otherwise one really needs to talk in terms of a dynamic description of components in that scene as they change. Second the choice of source material, as is well known from tests on low-bit-rate codecs [31], can easily dictate the results of an experiment, and should be chosen to reveal or highlight the attributes in question. Third, so-called demand characteristics of subjects can influence their perception of spatial attributes in sound reproduction. In other words, the subjects may have certain expectations of the spatial structure in the scene that is presented, based upon their experience and education, especially when that scene is of a familiar nature such as an orchestra or a string quartet. They may therefore communicate what they expect rather than what they actually perceive. This issue can be considered important if one is concerned with describing the absolute spatial characteristics of a scene, or when attempting to study human perception of reproduced sound scenes, but is less of an issue when attempting to conduct product evaluations where the judgments are mainly comparative.

A fourth issue, relating to source material, is that of complexity in the reproduced scene. Simple scenes consisting of a single source in an anechoic environment are

simple to control and simple to describe, making the subjective task very easy for a listener. Virtually the only judgments of relevance in such an evaluation are of source location and source size or extent. The next step up from this is a single source in a reflective environment, which can give rise to attributes such as source width, source focus, depth, distance, envelopment, and spaciousness (distance may be considered an aspect of source location, but absolute distance is hard to judge accurately in anechoic environments [32]). Such simple stimuli are often considered important when trying to establish relationships between physical variables and subjective parameters, as one can just about control all the variables and be clear about which subjective factors are affected. As soon as one introduced typical audio program material, involving multiple sources in different locations, coupled with room reflections or artificial effects, the scene becomes complex and more difficult to evaluate. Questions about attributes such as source width become possibly ambiguous (which source, and whether you mean the width of the whole image/scene or individual sources within it). Yet such complex source material is exactly the type of material that is important to use in the type of product evaluations that are in question here. If experiments are to have high ecological validity and enable the evaluation of the full range of problems and effects that can arise in spatial audio reproduction, then one cannot always be restricted to using simple source material. So it is important to develop a library of subjective terms with clear meanings, and to adopt a hierarchical structure of attributes in a "scene-based" spatial evaluation language, as introduced in the next section. Here graphical evaluation languages such as introduced in [33] may become more relevant and useful.

2.5 Spatial Attributes in Concert Hall Acoustics

This discussion would not be complete without mentioning the extensive work that has been carried out on spatial quality in natural acoustics, primarily in relation to concert hall design. As summarized by Morimoto [34], there is a long-established understanding in natural acoustics that auditory spatial impression consists of two primary dimensions, apparent source width (ASW) and listener envelopment (LEV). Considerable work has been undertaken to isolate the physical factors that affect these two subjective variables. Griesinger [35] has also proposed components of spatial impression based upon a concept of background and foreground auditory streaming, which helpfully separate source- and environment-related streams and depend on the temporal structure of sounds.

It is not the intention, in this paper, to attempt to analyze or criticize that literature in any detail, as it is well covered elsewhere. Neither is it intended to suggest that the attributes defined therein are irrelevant in listening tests or subjective experiments on reproduced sound. However, there is sufficient evidence to persuade this author that reproduced sound and synthetic auditory scene creation can give rise to subjective attributes either not encountered or not considered relevant in natural acoustics (see, for example, [10], [18]). ASW and LEV are not found to be sufficient on their own to describe the spatial sensations arising when comparing different forms of sound reproduction in a way that is meaningful to listeners when evaluating multisource, ecologically valid source material. For example, they say nothing about depth or distance, image skew, and so forth.

3 A "SCENE-BASED" APPROACH TO SPATIAL QUALITY EVALUATION

In order to address the need to evaluate complex reproduced source material that has high ecological validity, it is proposed that spatial audio reproduction characteristics should be evaluated subjectively according to a "scene-based" paradigm. This requires that the elements of the reproduced scene be grouped according to their function within the scene, at levels appropriate to the task. The concept of auditory scenes is not novel, but there is little evidence that the concept has been applied rigorously to the issue of spatial subjective assessment. This paradigm is primarily concerned with descriptive attributes, rather than with preference-related or naturalness constructs. It is also concerned, in the first instance, with scenes that are nominally static, although the paradigm might be extended to dynamic scenes in the future. In the examples given here the paradigm is considered in a two-dimensional form that excludes height, but it could easily be extended to include this dimension.

A basic and somewhat abstract example is shown in Fig. 1. Here a number of individual sources are located within a reflective acoustic environment. These could be instruments in a band or ensemble, for example. Typically, in recorded sound these are panned or otherwise located at points within the scene, giving rise to a stereophonic image that is perceived as having an overall width spanning the distance between the outer limits of the sources within the scene. The sources making up that image might be grouped together cognitively by the listener as an entity that could be labeled "ensemble." The macro scene element labeled "ensemble" may be perceived as having certain spatial attributes such as lateral location, width, depth, and distance (the distinction between depth and distance is considered important and will be discussed later). There may need to be a number of levels of ensemble width if it is necessary in a particular context to describe the characteristics of groups within groups, the largest ensemble of all being all the sources in the scene. In addition, the individual sources could themselves be perceived as having lateral location, width, distance, and possibly depth.

One can also extend this paradigm to include the acoustic environment in which the sources are located. Results from previous experiments [36] indicate that subjects can distinguish clearly between source- and environment-related attributes, enabling them to judge characteristics such as room width and room size independently of each other, and independently of source attributes. To take this one step further, global judgments may be made of the entire scene.

So there is an argument for grouping spatial attributes into micro and macro attributes, the micro attributes

describing the features of individual elements within a scene, and the macro attributes describing the scene as a whole, or groupings of elements within it. The concept is Russian doll–like, with the scene containing an environment (usually a collection of reflections and diffuse reverberation), within which are groups of sources, within which are individual sources. The reason this is considered important is to avoid the confusions that have been observed in subjective experiments with which the author is familiar. These confusions arise out of the use of complex source material coupled with a lack of clarity in the definition of the subjective attributes and the scene elements to which they relate. If subjects are to be trained to identify and grade these attributes reliably, then clarity in definition is required. Such clarity will also aid the establishment of clearer relationships between physical variables and subjective attributes.

3.1 Examples of Macro and Micro Attributes

3.1.1 Width

In this section it is proposed that, subjectively, there are at least three different types of width attribute, listed from micro to macro: individual source width, ensemble width, and environment width. There may also be a fourth, termed scene width, although this will depend on the context. These are shown in Fig. 2.

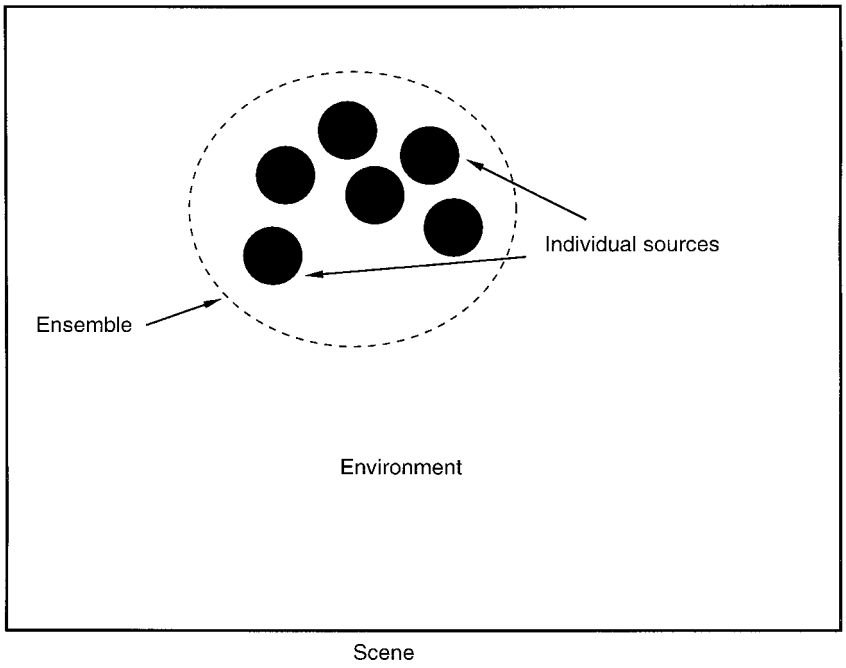

Fig. 1. Scene elements.

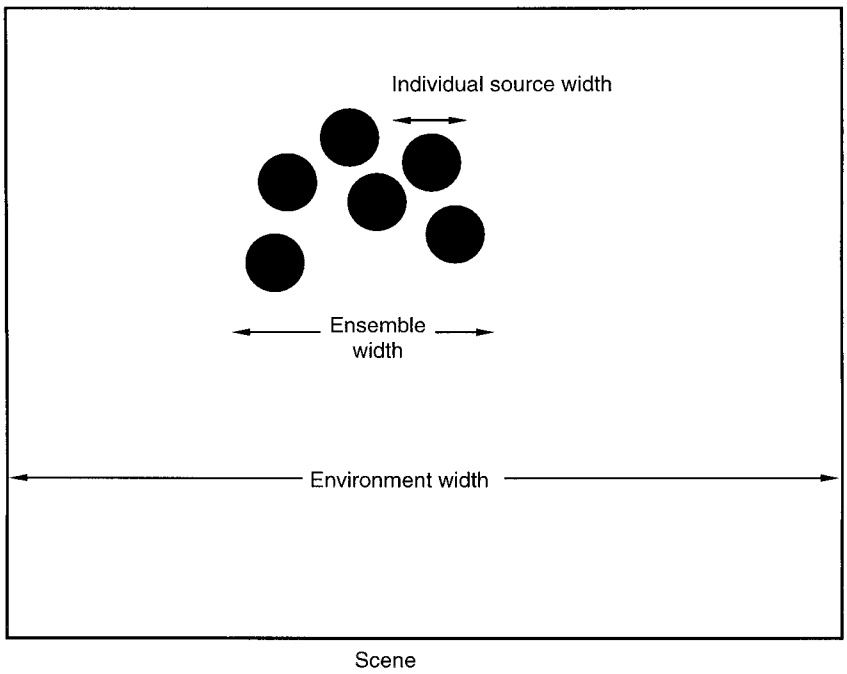

Fig. 2. Width attributes.

From concert hall acoustics research we are told that the phenomenon of apparent source width, or ASW, is dependent on the level, direction, frequency content, and structure of early reflections associated with a sound source, affecting the degree of interaural cross correlation (DICC) [37]. The perceptual stream labeled "source" may also be isolated and used to gain greater insight into the signal components that affect perceived source width [38], [39]. The effect of source broadening is observed, depending on the nature of these reflections, and has been associated with positive listener responses in such contexts. (It is not proved that the same positive connotations of large source width are present in judgments of reproduced sound, but some evidence was noted that precise source location accuracy is not of paramount importance for listener preference.)

Individual sources can appear to be made wider in sound reproduction by spreading or divergence controls, which divide energy between loudspeakers, as well as by the addition of artificial reflections. For the sake of clarity in the structure of the attributes proposed here, this type of width will be referred to as individual source width (ISW) to stress the fact that it refers to the perceived lateral extent of single sources.

This individual source phenomenon may be related to the degree of locatedness that a single source can be said to possess. (Locatedness, as described by Blauert [40], is the degree to which an auditory event can be said to be clearly in a particular location.) When a source has a small ISW, it is also likely to have high locatedness (it is easy to locate and appears to resemble a point source), whereas when it has a large ISW, it is more likely to have poor locatedness (it appears to be very large, possibly rather diffuse and difficult to locate). Listeners sometimes prefer to use terms such as poor image focus rather than large individual source width, but here they are describing the global characteristics of reproduced sound scenes in which all the sources appear to be fuzzy and difficult to localize. It is not clear, however, that a high source diffuseness is exactly congruent with large perceived width, or that these are identical attributes. (One could conceive, for example, of a large source with clearly defined boundaries that was also easy to localize.) A clear relationship was noticed, however [36], in experiments using different modes of spatial sound reproduction, where the attributes localization (defined in this case as the ease with which the direction of a source could be pinpointed) and source width were found to be negatively correlated.

The macro entity we call an ensemble is a group of sources that has a common cognitive label (orchestra, band, or string section). (The term ensemble used here has musical connotations but is intended to mean any group of sources that can legitimately be grouped together as a "macro object.") Reproduced stereo images of multiple sources (such as an ensemble or orchestra) have width by virtue of the amplitude and/or time differences between the loudspeaker channels arising from each source in the ensemble. Such width can be varied by altering these relationships using panpots, MS (midside) processing, or by controlling the relative amplitudes and timings of the signals from different instruments at an array of recording microphones [5]. (It is often the case, for example, that different microphone techniques, stereo processing algorithms, or reproduction arrangements have the effect of narrowing or widening the perceived width of groups of sources within the overall scene.) Clearly the perception and physical correlates of this width attribute are different from those for individual source width, because it is not primarily dependent upon early reflections, DICC, or divergence control, as in the case of ISW. It specifically excludes the apparent width of the environment within which the ensemble is housed (which may be perceived differently). Here, for the sake of clarity, this new type of width will be defined as ensemble width because it relates specifically to the perceived width of a group of sources which together are cognitively labeled an ensemble.

Environment or room width is yet another specific attribute, and experiments have shown that it is both separately perceivable by subjects, distinguishable from room size (which can be judged even in mono [36]), and separately controllable in terms of the physical parameters of the sound field [41]. It is derived from a cognitively separate information stream to foreground information that represents individual sources. (It appears to be dependent on the interaural decorrelation and time difference fluctuations of decaying reverberation tails. This supports Griesinger's concept of background spatial impression (BSI) [42] and depends on the ability of source material to reveal background reverberation in the gaps between notes of music or phonemes of speech.) Environment width seems to be related to a perception of the reverberant sound within the reproduced space and (under the definitions proposed here) is dependent on the ability to experience a sense of presence (see Section 3.1.3). It relates to the difference between the auditory sensation of a wide space and that of a narrow space.

In our experiments a sense of large environment width has, not surprisingly, gone hand in hand with the perception of well-externalized reverberation (perception of reverberation outside the head). This is only a relevant attribute when a separate reverberant environment is implied and perceived, such as in the majority of natural music recordings made in reverberant spaces. It is closely related to what others have called spaciousness and may have some things in common with LEV, but this will be discussed in more detail later. It may be less relevant when using program material such as pop music, where effects added to essentially dry sources may not imply the location of sources within a fixed space.

The fourth width category, here termed scene width, is proposed as a global spatial attribute (the highest level macro attribute) that describes the apparent width of the entire scene, including the reflective environment. The chances are that this will usually be the same as the environment width, as this is likely to be larger than any of the other components, but one can allow for situations in which this might not be the case. For example, in certain artificially constructed or "hyper-real" scenes, sound objects might be able to be placed outside the implied environment, or the environmental cues might be extremely

narrow. Such situations are hypothesized by Begault in [43]. Table 1 summarizes these different levels of width attribute. The width referred to is always the perceived with rather than the physical width of original sources.

An interesting question arises occasionally about what happens when a source or a group of sources in a reproduced sound environment are made so wide or diffuse that they apparently become enveloping (see Fig. 3). In other words, at what point does the attribute we call source width become another one called envelopment? (The correct answer is probably, "when subjects say that it does.") Interestingly Morimoto independently also makes a similar observation in [34], where he notes it could be argued that there is only one spatial impression dimension and that the difference between ASW and LEV might only be a matter of degree, depending on the size of the object. In surround sound reproduction this phenomenon is more easily encountered than in two-channel stereo, owing to the presence of loudspeakers to the sides or rear of the listener and the possibility for sources to be panned or spread all around the listener. This highlights the difficulty of a precise unidimensional definition of such attributes and is discussed further in Section 3.1.3.

3.1.2 Depth and Distance

Depth and distance attributes might initially appear to be the same, but here it is argued that they should be evaluated separately because they are different psychological constructs and are at different levels in our scene-based hierarchy of attributes. Again one may need to distinguish between sources and groups of sources, and between sources and environment.

Referring to the diagram in Fig. 4, it will be seen that source distance is considered to be the perceived range between a listener and a reproduced source. Depth on the other hand is related to the sense of perspective in the reproduced scene as a whole, and refers to the ability to perceive a scene that recedes from the listener, as opposed to a flat sound image. It is sometimes possible, for example, to judge source distance in mono (by listening to the direct-to-reverberant sound ratio and the relative loudnesses of the sources), but mono reproduction (it may be argued) gives little or no sense of spatial depth.

It is possible that individual sources may be perceived as having depth (as shown in Fig. 4). This has so far proved an elusive concept in subjective experiments and subjects do not often report perceiving it, although Martens reports its relevance during tests on low-frequency decorrelation in [44], and Berg and Rumsey [45] have noted subjects describing a contrast between curved and flat sources, which seems similar. (Martens found that subjects *drew* representations of source depth in graphical responses, but only rated individual source distance in a scaling experiment.) Groups of sources, or ensembles, may more readily by perceived as having depth that relates to the perceived front–back dimension of an ensemble, although this does not appear to be a prominent perception in formal and informal experiments conducted to date. Elicitation experiments so far conducted do not seem to have revealed a separate construct of environment depth, although it may yet come to light. By far the stronger perception seems to be environment width, for reasons not yet explained, although it may have to do with the concept of construct masking, in which strong perceptual constructs have a tendency to dominate the overall judgment, thereby hiding weaker ones.

Table 2 summarizes the different proposed levels of distance and depth attributes.

Table 1. Proposed definitions of width attributes in reproduced sound (see also Fig. 2).

Attribute	Construct Definition
Individual source width	Width of individual source(s) within a scene
Ensemble width	Overall width of a defined group of sources (may be all the sources in the scene if required)
Environment width	Broadness of (reflective) environment within which individual sources are located
Scene width	Composite or global width of entire scene

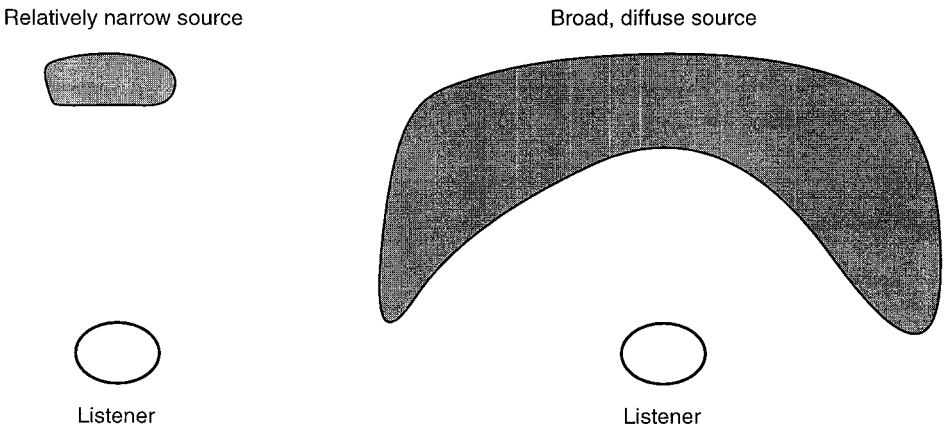

Fig. 3. (a) Narrow source. (b) Source perceived as very wide, wrapped around listener, and diffuse may be considered enveloping.

3.1.3 Envelopment, Spaciousness, and Spatial Impression

Envelopment, spaciousness, and spatial impression are terms that seem to result in the most varied interpretation in the literature. They are harder to conceive of than dimensional quantities such as width, depth, or distance, as they are not perceived directly as linear quantities but more as semiabstract and multidimensional impressions. They are all terms that relate in some way to the degree of immersion in the sound field experienced by the listener or to a global description of the scene, and are presented here as a different class of attributes to the dimensional attributes proposed in the previous sections. The earlier group might be termed dimensional attributes, whereas this group might be termed immersion attributes.

In colloquial terms, it is important for the future of this field that "everyone is singing from the same hymn sheet," a task that is extremely hard and sometimes impossible when dealing across cultures and languages. Anyone who attempts to wrestle with the semantics of these terms is to some extent asking for trouble, but it seems important that it be done. It is also quite likely that each individual using these terms will think that everyone else understands the same thing by them, but the literature is full of subtly different interpretations.

Spatial impression has typically been used as a form of "cover all" term, describing one or more spatial sensations. It is not very helpful in practice, as it is not well defined and different people interpret it in different ways, so it is dispensed with as a useful unidimensional sensation. Barron and Marshall [46] originally discussed two forms of spatial impression, one related to diffuse reverberation and the other to lateral reflections. The former resulted in the sensation of being inside a room and was accompanied by a sense of distance from the source, whereas the latter appeared to give rise to a form of source-related envelopment involving sensations apparently close to the listener. Rather, as proposed in Section 3.1.1, they suggested that as the source width increases because of increasing levels of lateral reflections, the sensation becomes enveloping (a form of individual source envelopment in the terms of this paper). These sensations have been clarified over the years in the concert hall acoustics literature, leading to the relatively clear definitions of ASW and LEV, as introduced earlier.

LEV is related to the subjective impression of being immersed in the reverberant sound in a hall and was found to be related to late, lateral reflected energy in concert hall acoustics, as examined by Bradley and Soulodre [47]. Morimoto, however, along with other Japanese colleagues, tends to refer to LEV as the degree of fullness of

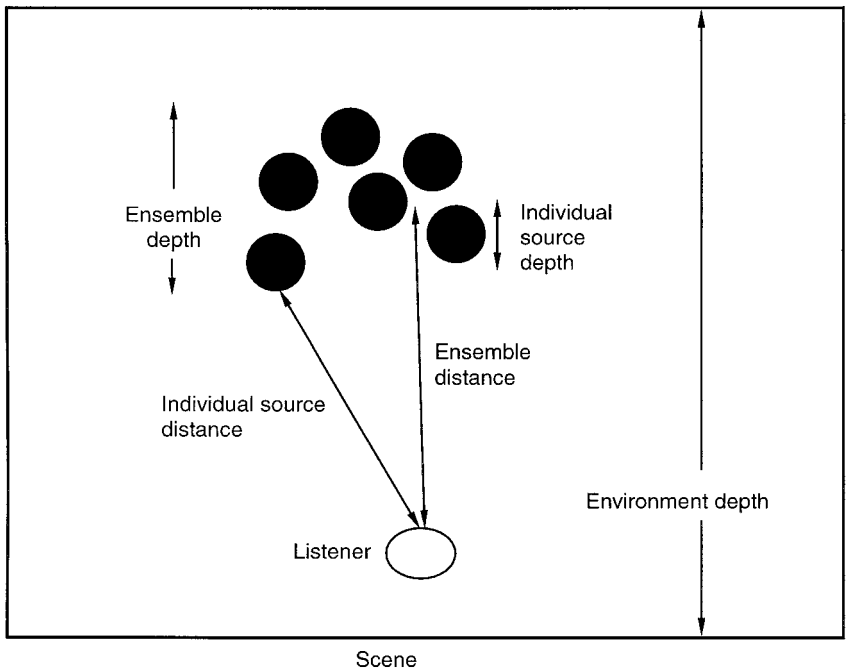

Fig. 4. Depth and distance attributes.

Table 2. Proposed definitions of distance and depth attributes in reproduced sound (see also Fig. 4).

Attribute	Construct Definition
Individual source distance	Distance from listener to perceived location of a source
Ensemble distance	Distance from listener to perceived midpoint of an ensemble
Individual source depth	Depth of individual source within a scene
Ensemble depth	Depth of a group of sources
Environment depth	Depth of (reflective) environment within which sources are located
Scene depth	Composite or global depth of entire scene, including environment

sound images around the listener, excluding the precedent sound image composing ASW [34]. Is fullness the same thing? His diagrams and writing suggest that he is definitely talking about immersion in reverberation.

As mentioned in Section 3.1.1, subjects often use the term envelopment when they are surrounded by a number of dry sources in surround sound reproduction, and they sometimes even do so when a single source is so broadly spread and diffuse as to "wrap around" the subject and appear enveloping. This sensation is almost certainly not a property of late reflected sound, as the sources in question can be dry and direct, so it cannot be considered to be LEV in the traditional sense. This scenario rarely arises in concert hall acoustics, as the listener is rarely placed in the middle of the orchestra. If they were, they would probably claim to be enveloped by sound or "inside the music," but this would not conform to the received definition of LEV, although it might have something to do with Morimoto's sense of "fullness of sound images around the listener." So a new term is needed for this type of envelopment.

Spaciousness may relate to a variety of scene elements and implies a sense of being inside a spacious environment. Letowski [1] defined spaciousness as "that attribute of auditory image in terms of which the listener judges the distribution of sound sources and the size of acoustical space. Spaciousness, he said, "enables the listener to judge that two sounds, which have, but do not have to have, the same pitch, loudness, duration, and timbre, are arriving from different directions." In his terms, then, it is also a multidimensional concept that refers to any spatial context in which the source direction can be determined and where the size of the space can be judged. He subdivides spaciousness in his MURAL (*mu*ltilevel auditor*r*y *a*ssessment *l*anguage) as shown in Fig. 5.

In [39] Griesinger differentiates between spatial impression and spaciousness. Here he refers to spatial impression with examples that imply the sense of being present within any enclosed space, whereas spaciousness is reserved for the experience of large reverberant spaces. This is useful, as it is close to our need for a dimensional judgment of some scene element—in fact spaciousness is here very similar to the definition of environment width and depth given in the preceding.

As mentioned in the introduction to this section, for the purposes of this discussion it is convenient to separate subjective attributes relating to environment dimensions (width, depth, height) from immersion attributes (such as envelopment). Subjective impressions of large and small environments were already dealt with in the previous two sections with terms such as environment width and environment depth. These specifically refer to subjective sensations of the dimensions of the space around a subject arising from a background stream of reverberant information, rather than being related to sources.

We will propose a new attribute named "presence," defined as the sense of being inside an (enclosed) space. This implies that the subject is able to sense the boundaries of the space around him or her. In other words, subjects feel present within the space rather than absent from the space. (This concept is supported by subjective data from elicitation experiments in which subjects have described their experience of different spatial modes of reproduced sound as "outside the event" and "in a corridor outside" in opposition to being "in the center of the sound" [21].) Presence, as defined here, is primarily related to environmental, contextual, or background cues.

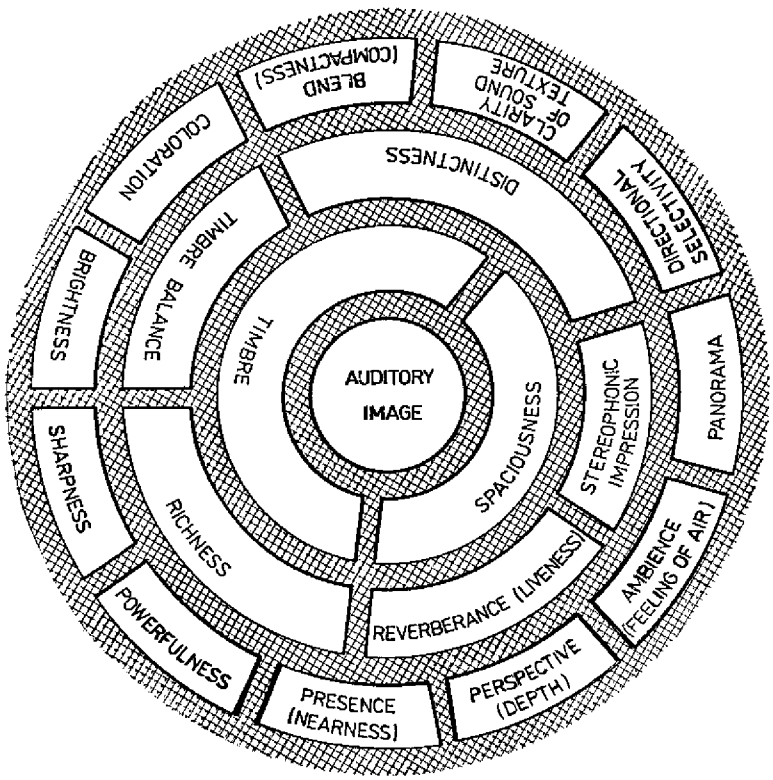

Fig. 5. Letowski's MURAL (from [1]).

One should not rule out the possibility that sensations of presence might be experienced in outdoor environments, where numerous low-level dry sources merge to create a background ambiance (hence the parentheses around "enclosed"), but here we are primarily considering reverberant environments. An important criterion for presence is hypothesized to be an awareness of background-stream sound energy arriving from many directions.

Envelopment, on the other hand, must be subdivided into environmental envelopment and source-related envelopment, the former being similar to LEV in concert halls and the latter to envelopment by one or more dry or direct foreground sound sources. This is summarized in Table 3.

These definitions give rise to a number of observations. First, presence and environmental envelopment are not necessarily the same, although they may be closely related. The former is a prerequisite for the latter. Once a subject feels to be inside the space, they are able to judge concepts such as environment width and depth as defined before, and they can be enveloped to varying degrees by reverberant sound. This hypothesis is supported by the work of Berg, to be discussed later. Second, sources (and groups of sources) can be enveloping. (This point concurs with Griesinger's view that individual sources can be enveloping, and in his writing related to the interaction between continuous sound sources and reflected energy he refers to this as continuous spatial impression (CSI [47].) Third, the physical mechanism for each of these types of effect is different.

The mechanisms for these effects are still being studied. Individual source envelopment can be caused in sound reproduction by effects similar to Griesinger's CSI or by the artificial and very wide spreading of dry sources by variable panning devices such as Gerzon's stereo image spreading circuit [48]. Ensemble source envelopment is caused by panning numerous dry sources to locations that together surround the listener (similar to the concert hall concept of a listener placed in the middle of an orchestra). Environmental envelopment appears to be related to the background information stream, in reproductions of natural spaces being dependent on the level and directional distribution of late, diffuse reverberant energy, similar to the concert hall LEV.

Data from the experiment described in [36] have been analyzed in greater detail by Berg [49], lending some support to the paradigm and distinctions suggested earlier, at least in the context of that experiment. Here a subset of the subjective ratings given by listeners when comparing different modes of spatial sound reproduction was analyzed by factor analysis for the following attributes:

- Presence (psc)
- Envelopment (env)
- Room width (rwd)
- Room size (rsz)
- Room level (rlv).

Fig. 6 shows the factor loadings of these attributes when two factors were extracted and subjected to varimax rotation. One possible interpretation of the factors is that factor 2 represents a sense of presence in the reproduced environment (being strongly loaded, not surprisingly, for the presence attribute). Factor 1 represents an ability to judge aspects of the reproduced environment such as room size and reverberant level. Other interesting observations from this analysis are that room size and reverberant level do not appear to require a strong sense of presence to judge them, whereas envelopment requires a strong sense of presence (although it may be the other way around). Also the ability to judge room width (in this paper's terms,

Table 3. Proposed definitions of immersion attributes in reproduced sound.

Attribute	Construct Definition
Individual source envelopment	Sense of being enveloped by a single sound source
Ensemble source envelopment	Sense of being enveloped by a group of sound sources
Environmental envelopment	Sense of being being enveloped by reverberant or environmental (background stream) sound
Presence	Sense of being inside an (enclosed) space or scene

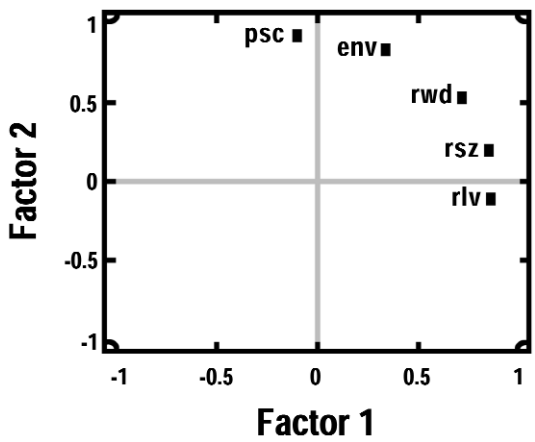

Fig. 6. Factor loadings for a selection of environment-related attributes.

environment width) requires a degree of presence. The factor 2 (presence) attributes appeared to be strongly dependent on surround modes of reproduction, whereas a number of the room acoustics attributes loading factor 1 could be judged even using mono reproduction.

3.2 Miscellaneous Spatial Attributes

A number of further attributes or characteristics may be important in the evaluation of spatial audio, not all of which fit cleanly into the aforementioned scene-based paradigm, but are nonetheless relevant. Without them one would not have a complete description of spatial quality, and most of them relate to some form of spatial distortion of the global scene compared with a reference scene rendering. Some examples of these are defined in Table 4, with an attempt to place them at an appropriate level in the scene-based model.

An accurate evaluation of all of these constructs still does not enable one to differentiate between natural spatial characteristics and unnatural ones. For example, phasiness or phase reversal in stereophonic signals can lead to a strong sense of unnaturalness, as can a simple left–right reversal of a scene. The analysis of precisely what constitutes naturalness, though, is a separate topic and will be considered at another time.

It is not proposed that all of these and the aforementioned spatial attributes should be used in every listening experiment, but simply that similar clarity of definition should be employed. Attributes should be chosen for an evaluation based on the task and context in question.

4 CONCLUSION

In the foregoing paper the need for reliable, preferably unidimensional, spatial attributes has been justified more broadly within the context of sound quality. Existing standards and previous work in the field have been reviewed and the spatial attributes therein defined have been found insufficient in various respects. In order to ensure clarity in semantics concerning spatial attributes for the subjective evaluation of ecologically valid source material, a novel scene-based paradigm has been proposed. This separates descriptions of sources, groups of sources, environments, and global scene parameters. It also separates attributes into a dimensional group and an immersion group. It is currently based on the evaluation of static characteristics, but could be extended to dynamic scenes in the future. The paradigm is regarded as ongoing work, and is based on results so far obtained from formal and informal listening experiments, and on literature-based observations of the author and colleagues. It is presented as a contribution to the debate rather than a definitive account of completed work.

5 ACKNOWLEDGMENT

The author wishes to thank research students and staff at the Institute of Sound Recording and the School of Music in Piteå for numerous interesting discussions, listening sessions, and results that have led to the formation of these ideas: Jan Berg, Tim Brookes, Dave Fisher, Natanya Ford, Douglas McKinnie, Russell Mason, David Murphy, Amber Naqvi, Tobias Neher, and Slawek Zielinski. In particular, the author wishes to thank Jan Berg, Gilbert Soulodre, Russell Mason, Slawek Zielinski, David Murphy, Bill Martens, and the review panel of this *Journal* for comments on the manuscript that helped to clarify a number of important concepts.

6 REFERENCES

[1] T. Letowski, “Sound Quality Assessment: Cardinal Concepts,” presented at the 87th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 37, p. 1062 (1989 Dec.), preprint 2825.

[2] J. Nunally and I. Bernstein, *Psychometric Theory*, 3rd ed. (McGraw-Hill, New York, 1994).

[3] D. R. Begault, “Challenges to the Successful Implementation of 3-D Sound,” *J. Audio Eng. Soc. (Engineering Reports*), vol. 39, pp. 864–870 (1991 Nov.).

[4] J. Huopaniemi, “Virtual Acoustics and 3D Sound in Multimedia Signal Processing,” Ph.D. thesis, Rep. 53, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Helsinki, Finland (1999).

[5] F. Rumsey, *Spatial Audio* (Focal Press, Oxford and Boston, 2001).

[6] A. Berkhout, D. de Vries, and P. Vogel, “Acoustic Control by Wave Field Synthesis,” *J. Acoust. Soc. Am.*, vol. 93, pp. 2764–2778 (1993).

[7] J. Blauert, “Instrumental Analysis and Synthesis of Auditory Scenes: Communication Acoustics,” in *Proc. AES 22nd Int. Conf.* (2002), pp. 387–395.

[8] F. E. Toole, “Subjective Measurements of Loudspeaker Sound Quality and Listener Performance,” *J. Audio Eng. Soc.*, vol. 33, pp. 2–32 (1985 Jan./Feb.).

[9] A. Gabrielsson and H. Sjögren, “Perceived Sound Quality of Sound Reproducing Systems,” *J. Acoust. Soc. Am.*, vol. 65, pp. 1019–1033 (1979).

Table 4. Proposed definitions of miscellaneous spatial attributes in reproduced sound.

Attribute	Definition
Scene left–right skew	Degree to which a spatial audio scene is skewed to the left or right from a stated reference position
Scene front–back skew	Degree to which a spatial audio scene is skewed to the front or back from a stated reference position
Source stability	Degree to which individual sources remain stable in space with respect to time (assuming nominally stationary sources)
Scene stability	Degree to which the entire scene remains stable in space with respect to time
Source focus	Degree to which individual sources can be precisely located in space (this may be closely related to ISW)
Scene width homogeneity	Evenness of distribution of scene elements compared with a reference scene

[10] N. Zacharov and K. Koivuniemi, "Unravelling the Perception of Spatial Sound Reproduction," in *Proc. AES 19th Int. Conf.* (2001), pp. 272–286.

[11] F. Rumsey, "Subjective Evaluation of the Spatial Attributes of Reproduced Sound," in *Proc. AES 15th Int. Conf.* (1999), pp. 122–135.

[12] T. Nakayama, T. Muira, O. Kosaka, M. Okamoto, and T. Shiga, "Subjective Assessment of Multichannel Reproduction," *J. Audio Eng. Soc.*, vol. 19, pp. 744–751 (1971 Oct.).

[13] M. R. Schroeder, "Normal Frequency and Excitation Statistics in Rooms: Model Experiments with Electric Waves," *J. Audio Eng. Soc.*, vol. 35, pp. 307–316 (1987 May).

[14] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating Interactive Virtual Acoustic Environment," *J. Audio Eng. Soc.*, vol. 47, pp. 675–705 (1999 Sept.).

[15] N. Zacharov and K. Kuovuniemi, "Unravelling the Perception of Spatial Sound Reproduction: Analysis and External Preference Mapping," presented at the 111th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 49, p. 1217 (2001 Dec.), preprint 5423.

[16] J. Berg and F. Rumsey, "Correlation between Emotive, Descriptive and Naturalness Attributes in Subjective Data Relating to Spatial Sound Reproduction," presented at the 109th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 48, p. 1106 (2000 Nov.), preprint 5206.

[17] A. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA, 1990).

[18] J. Berg, "Systematic Evaluation of Perceived Spatial Quality in Surround Sound Systems," Ph.D. thesis, Luleå University of Technology, School of Music at Piteå, Sweden (2002).

[19] J. Berg and F. Rumsey, "Cluster Analysis of Scaled Verbal Descriptors," presented at the 108th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 48, p. 360 (2000 Apr.), preprint 5139.

[20] S. Bech, "Methods for the Subjective Evaluation of the Spatial Characteristics of Sound," in *Proc. AES 16th Int. Conf.* (1999), pp. 487–504.

[21] J. Berg and F. Rumsey, "Spatial Attribute Identification and Scaling by Repertory Grid Technique and Other Methods," in *Proc. AES 16th Int. Conf.* (1991), pp. 51–66.

[22] T. Neher, "Stimulus Manipulation for 3D Audio Evaluation," Int. Rep., Institute of Sound Recording, University of Surrey, UK (2001).

[23] R. Mason and F. Rumsey, "An Assessment of the Spatial Performance of Virtual Home Theater Algorithms by Subjective and Objective Methods," presented at the 108th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 48, p. 359 (1999 Apr.), preprint 5137.

[24] J. M. Jot, "Efficient Models for Reverberation and Distance Rendering in Computer Music and Virtual Audio Reality," in *Proc. Int. Computer Music Conf.* (Thessaloniki, Greece, 1997), pp. 236–243.

[25] J. P. Jullien, E. Kahle, M. Marin, and O. Warusfel, "Spatializer: A Perceptual Approach," presented at the 94th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 41, p. 386 (1993 May), preprint 3465.

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

[27] F. Rumsey, "Controlled Subjective Assessments of Two-to-Five-Channel Surround Sound Processing Algorithms," *J. Audio Eng. Soc.*, vol. 47, pp. 563–582 (1999 July/Aug.).

[28] Draft IEC 60268, "Sound System Equipment—Part 13: Listening Tests on Loudspeakers," International Electrotechnical Commission, Geneva, Switzerland (1997).

[29] EBU Rec. 562-3, "Subjective Assessment of Sound Quality," European Broadcasting Union, Geneva, Switzerland (1990).

[30] F. Rumsey, Subjective Assessment of the Spatial Attributes of Reproduced Sound," in *Proc. AES 15th Int. Conf.* (1998), pp. 122–135.

[31] G. Soulodre, T. Grusec, M. Lavoie, and L. Thibault, "Subjective Evaluation of State-of-the-Art Two-Channel Audio Codecs," *J. Audio Eng. Soc. (Engineering Reports)*, vol. 46, pp. 164–177 (1998 Mar.).

[32] M. Gardner, "Distance Estimation of 0 Degrees or Apparent 0-Degree-Oriented Speech Signals in Anechoic Space," *J. Acoust. Soc. Am.*, vol. 45, pp. 47–53 (1969).

[33] R. Mason, N. Ford, F. Rumsey, and B. de Bruyn, "Verbal and Nonverbal Elicitation Techniques in the Subjective Assessment of Spatial Sound Reproduction," *J. Audio Eng. Soc.*, vol. 49, pp. 366–384 (2001 May).

[34] M. Morimoto, "How May Auditory Spatial Impression be Controlled?" in *Proc. 2nd Int. Workshop on Spatial Media* (University of Aizu, Japan, 2001).

[35] D. Griesinger, "Spatial Impression and Envelopment in Small Rooms," presented at the 103rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 45, pp. 1013, 1014 (1997 Nov.), preprint 4638.

[36] J. Berg and F. Rumsey, "Verification and Correlation of Attributes Used for Describing the Spatial Quality of Reproduced Sound," in *Proc. AES 19th Int. Conf.* (2001), pp. 233–251.

[37] M. Morimoto and K. Iida, "A Practical Evaluation Method of Auditory Source Width in Concert Halls," *J. Acoust. Soc. Jpn. (E)*, vol. 16, no. 2, pp. 59–69 (1995).

[38] R. Mason and F. Rumsey, "A Comparison of Objective Measurements for Predicting Selected Subjective Spatial Attributes," presented at the 112th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 50, p. 521 (2002 June), preprint 5591.

[39] D. Griesinger, "The Psychoacoustics of Apparent Source Width, Spaciousness and Envelopment in Performance Spaces," *Acta Acustica*, vol. 83, pp. 721–731 (1997).

[40] J. Blauert, *Spatial Hearing. The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, 1997).

[41] R. Mason and F. Rumsey, "An Investigation of Interaural Time Difference Fluctuations, Part 4: The Subjective Effect of Fluctuations in Decaying Stimuli Delivered over Loudspeakers," presented at the 111th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 49, pp. 1225, 1226 (2001 Dec.), preprint 5458.

[42] D. Griesinger, "Objective Measures of Spaciousness and Envelopment," in *Proc. AES 16th Int. Conf.* (1999), pp. 27–41.

[43] D. Begault, *3D Sound for Virtual Reality and Multimedia* (Academic Press, New York, 1994).

[44] W. Martens, "The Impact of Decorrelated Low Frequency Reproduction on Auditory Spatial Imagery: Are Two Subwoofers Better than One?" in *Proc. AES 16th Int. Conf.* (1999), pp. 67–77.

[45] J. Berg and F. Rumsey, "Identification of Perceived Spatial Attributes of Recordings by Repertory Grid Technique and Other Methods," presented at the 106th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 47, p. 525 (1999 June), preprint 4924.

[46] M. Barron and H. Marshall, "Spatial Impression Due to Early Lateral Reflections in Concert Halls: The Derivation of a Physical Measure," *J. Sound Vibr.*, vol. 77, pp. 211–232 (1981).

[47] J. Bradley and G. Soulodre, "Objective Measures of Listener Envelopment," *J. Acoust. Soc. Am.*, vol. 98, pp. 2590–2597 (1995).

[48] M. Gerzon, "Signal Processing for Simulating Realistic Stereo Images," presented at the 93rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 40, p. 1054 (1992 Dec.), preprint 3423.

[49] J. Berg, Personal communication (2001).

THE AUTHOR

Francis Rumsey graduated in 1983 with first class honours (BMus Tonmeister) in Music with Applied Physics and received a Ph.D. degree in 1991 from the University of Surrey (UniS).

He was appointed a lecturer at UniS in 1986 and is currently a reader at its Institute of Sound Recording. He is also a visiting professor at the School of Music in Piteå, Sweden.

Dr. Rumsey was the winner of the 1985 BKSTS Dennis Wratten Journal Award, the 1986 Royal Television Society Lecture Award, and the 1993 University Teaching and Learning Prize. He is the author of over 100 books, book chapters, papers, and articles on audio, and in 1995 was made a fellow of the AES for his significant contributions to audio education. His book *Spatial Audio* was recently published by Focal Press.

Dr. Rumsey has served the AES as member of the Board of Governors, chair of the British Section (1992–1993), vice president, Northern Region, Europe (1995–1997), and vice chair of the 19th International Conference in 2001. He is currently chair of the AES Technical Committee on Multichannel and Binaural Audio Technologies and chair of the AES Membership Committee. He was a partner in EUREKA project 1653 (MEDUSA), studying the optimization of consumer multichannel surround sound. His current research includes a number of studies involving spatial sound quality evaluation, and he is leading a project funded by the Engineering and Physical Sciences Research Council concerned with subjective quality tradeoffs in consumer multichannel audio and video delivery systems, in collaboration with the Psychology Department at UniS, Bang & Olufsen, and BBC Research & Development.