INTERCHANGE FORMATS AND THE ART OF SPATIAL AUDIO

Gary S. Kendall

Sonic Arts Research Centre
School of Music & Sonic Arts
Queen's University Belfast

ABSTRACT

The requirements of a spatial interchange format for electroacoustic music are examined with a focus on the interrelationship between the authoring of spatial content and its interpretation in audio reproduction.

1. AUTHORING AND INTERPRETATION

In the context of electroacoustic music, the role of a spatial audio interchange format is to be a bridge between the processes of authoring and interpreting spatial content (Fig. 1). It must enable the author to represent spatial content in a way that anticipates the means of reproduction in practical way. It must also represent the author's spatial content in a way that enables the interpreter to adapt it to a wide variety of reproduction settings. In scene-based multimedia, this interpretation is intended to be automated and to respond to user actions, but in electroacoustic music, the complexity of the interpretation is so great that one can hardly imagine how this could be accomplished except as an act of artistic judgment as happens in live diffusion. Because authoring and interpretation are so closely linked, the interchange format itself should include a description of the interpretation schemes that adapt content to the means of reproduction, even supporting alternative strategies. This is important to the author as well as the interpreter because they essentially participate in a process of artistic co-creation.

2. WHAT THE FORMAT REQUIRES

2.1. Interchange Format Primitives

In order to take into account the broad range of spatial audio practices observed in electroacoustic music, the interchange format must support a layer of more primitive descriptions than is typical in today's scene-based frameworks. In electroacoustic music, authors and interpreters especially need to capture specific spatial processing techniques because spatial content and DSP algorithms are intrinsically linked together, as in the case of granular and spectral-spatial processing [5,6]. Then too, content is also often linked to assumptions about the reproduction process, such as the arrangement of output channels. A layer of such primitives should not present any barrier to scene-based conventions and, in fact, should also support the creative manipulation of those conventions.

2.2. Conceptual Sources and Spatial Images

The complexity of the relationship between the conceptual source described in the interchange description and the perceived spatial image produced in reproduction is often underestimated. For example, in reproduction with 5.1 systems there is a considerable difference between the virtual image created by equal amplitude in the front-left and front-right loudspeakers and the image created by the center channel loudspeaker. Even though these sources can be described as originating from the same virtual spatial location, the perceived spatial images are different. The interchange format needs to differentiate virtual images from in-loudspeaker images (as well as from spatial images created with technologies like wave field synthesis).

One approach to ameliorating such differences in spatial imagery might be to attach spatial descriptors to the virtual source that set goals for the attributes of the spatial image created in reproduction. We can look to the literature of spatial quality evaluation for guidance on terminology for these perceptual attributes [1,4,7]. Following Rumsey [4], it is undoubtedly an important idea to separate the spatial attributes from the attributes of spaces, thereby separating source spatial properties from environmental properties. A practical set of spatial attributes can be given as direction, distance, width and depth [3,4]. These are attributes that clearly can be heard by listeners and that can be manipulated through recording and spatial processing techniques. We should be skeptical though of the notion that any general-purpose controls for these perceptual attributes can be applied to pre-existent sources independent of specific processing techniques. Such attributes are typically manipulated, not through perceptual controls, but in the settings for DSP processes that affect these attributes. These processes are sometimes quite idiosyncratic and often interact with other attributes of the source. For example, there are various ways of increasing source width that depend on temporal or spectral partitioning of the source [3]. This supports the requirement that the interchange format should incorporate the coding of specific spatial algorithms.

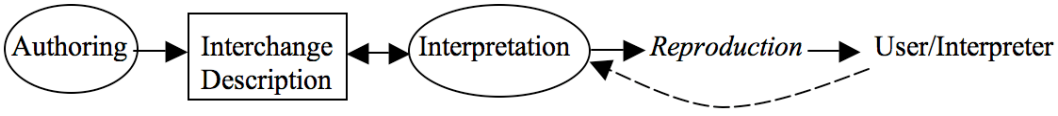

Figure 1. The relationship of the interchange format to authoring and interpretation.

2.3. Diffuse Soundfields

Interestingly, the spatial attributes of diffuse soundfields are more resilient when interpreted in diverse reproduction settings than are point sources [2]. There are a variety of techniques for creating diffuse soundfields, both in authoring and in interpretation, and these are once again dependent on specific techniques. What these techniques largely have in common is the need for multiple channels of reproduction. The interchange format could support the authoring of diffuse fields by encoding: 1) sources, techniques and goals, or 2) multichannel audio and goals. In the conventions of live diffusion, interpretation can involve the distribution of source material among a variety of types of loudspeakers, generally ranging from tweeters to subwoofers. This too creates diffuse soundfields, but at the stage of interpretation. The interchange format can encode such techniques by relating goals to categories of loudspeakers deployed in the reproduction setting.

2.4. Compatibility and Capture

Our conventional audio formats from stereo to 5.1 to 8 and more channels represent legacy formats that are authored and interpreted according to recognized conventions. A particular benefit of absorbing these formats into the interchange format is that the author can describe specific goals and techniques for the interpretation of the spatial content. This can be particularly useful in high-resolution reproduction settings where the number of channels exceeds the format of the original material. Then too, live diffusion in such situations illustrates how artistic co-creation between the author and interpreter takes place in actual practice: the interpreter adds content through dynamic manipulation of spatial imagery. The interchange format must accommodate the storage of dynamic control data in order to capture dynamics of live diffusion as well as specific diffusion performances.

2.5. Spatial Conventions and Idioms

The vitality of a listener's experience depends not only on the clarity of the spatial imagery, but also on the depth of the listener's engagement with the content. Hierarchical scene graphs provide a framework for both organizing and understanding audio content in terms of conventions based in everyday experience (for example, being enclosed in a room with a sound source). Scene-based formats illustrate that such conventions can be supported by relatively high-level controls. Of course, the authors of audio content will use every means available, including the violation of these conventions, for expressive purposes and we should acknowledge this from the beginning. From this perspective, we can see that the common conventions of audio scenes are simply a canvas on which the author paints content. Then too, the convention of modeling sources in physical rooms is too low a standard for common audio production. A competing model can be found in the conventions of film sound that illustrate how audio practice will gravitate toward expressive and practical conventions. For example, consider the following conflicts between audio conventions and audio realism:

1) *dynamic range* (must keep softest content audible; the intensity/distance rule needs to be flexible)
2) *good acoustics* (not all rooms sound good; idealized reverberation must be employed)
3) *source/reverberation interaction* (reverberation must be matched to source characteristics; reverberation characteristics must be tunable)
4) *EQ* (the overall mix must sound clear; environmental and source filtering has to be adjusted)
5) *Doppler shift* (tonal sources must stay in tune; Doppler shift must be tunable down to zero)

3. CONCLUSION

Just a few decades ago, standards and practices for stereo were simple issues. Our own technological inventiveness has seemingly raced ahead of our ability to maintain scalability and portability and this problem is strongly felt in the electroacoustic community. We must be mindful of artistic content and practice if an interchange format is to be viable across communities of authors and interpreters. Observations of common practice suggest a flexible combination of high-level controls and in-depth graphs of signal-flow.

4. REFERENCES

[1] Berg, J., Rumsey, F. "Systematic Evaluation of Perceived Spatial Quality." *Proceedings of the AES 24th International Conference*, 2003.

[2] Kendall, G. "The Decorrelation of Audio Signals and Its Impact on Spatial Imagery." *Computer Music Journal,* 19(4), 1995.

[3] Kendall, G. "The Artistic Play of Spatial Organization: Spatial Attributes, Scene Analysis and Auditory Spatial Schemata." *Proceedings of the 2007 International Computer Music Conference*, Denmark, 2007.

[4] Rumsey, F. "Spatial Quality Evaluation for reproduced Sound: Terminology, Meaning, and a Scene-based Paradigm." *Journal of the AES*, 50(9), 2002.

[5] Torchia, R. and C. Lippe. "Techniques for Multi-Channel Real-Time Spatial Distribution Using Frequency-Domain Processing." *Proceedings of the 2004 Conference on New Interfaces for Musical Expression*, Hamamatusu, Japan, 2004.

[6] Truax, B. "Composition and diffusion: space in sound in space." *Organised Sound,* 3(2), 1999.

[7] Zacharov, N. and K. Koivuniemi. "Unravelling the Perception of Spatial Sound Reproduction." *Proceedings of the AES 19th International Conference*, 2001.