SpatDIF: Principles, Specification, and Examples

Abstract
SpatDIF, the Spatial Sound Description Interchange Format, is an ongoing collaborative effort offering a semantic and syntactic specification for storing and transmitting spatial audio scene descriptions.

The SpatDIF core is a lightweight minimal solution providing the most essential set of descriptors for spatial sound scenes. Additional descriptors are introduced as extensions, expanding the namespace and scope with respect to authoring, scene description, rendering and reproduction of spatial audio.

A general overview of the specification is provided, and two use cases are discussed, exemplifying SpatDIF's potential for file-based pieces as well as real-time streaming of spatial audio information.

Principles
SpatDIF presents a hierarchical, unambiguous structure. The SpatDIF-syntax serves for structuring audio-scene related information. Authoring and rendering of spatial audio may occur at completely separate times and places, and be executed with tools whose capabilities cannot be known in advance.

SpatDIF is a syntax rather than a programming interface or file-format. SpatDIF may be represented in any structured mark-up language or message system. SpatDIF describes only the aspects required for the storage and transmission of spatial information.

```
SpatDIF declaration
  Kind Name Descri  o
  Entity
  Address
  Statement
  The SpatDIF terminology.
```

Terminology
A SpatDIF scene is the combination of a space and the actions that are unfolding within it. A scene consists of a number of SpatDIF entities. Entities are all objects that are affecting or interacting with the sound of that scene.

Entities can be of different kinds, e.g., sources or sinks. Each entity instance is assigned a name, so that it may be uniquely identified within the scene. The properties of entities are described and transmitted via SpatDIF descriptors.

A complete SpatDIF statement consists of an address unambiguously identifying an entity, its descriptor and its associated value.

Structure
A SpatDIF scene can consist of two sections. The meta section serves to configure and initialize the system, while the time section describes the temporal unfolding of a scene.

Meta Section
This section contains meta descriptions and is located at the top of a SpatDIF representation. Its information is not executed at runtime; timed events are excluded from this section. Descriptions include extension setup information, general annotation and documentation, ordering of the time section and more. The meta section is mandatory for a SpatDIF representation.

Time Section
This section holds information about how entities and their properties unfold over time. Each statement is located at a specific point in time.

Several ordering principles exist. Ordering by time is equivalent to a score, while ordering by entities groups the statements into tracks. In real-time streaming ordering by time is necessary. In storage cases other ordering principles may be applied.

SpatDIF Core
The basic SpatDIF namespace is defined in the core and contains the most essential set of functionalities for describing spatial sound scenes. A SpatDIF compliant audio renderer must understand and interpret all core statements.

Extensions
SpatDIF extensions introduce additional descriptors in a modular way. The extensions expand the namespace and scope of SpatDIF in relation to authoring, description, rendering and reproduction of spatial sound. Extensions might introduce new kinds of entities, expand the set of descriptors for existing entities, augment descriptors defined by other extensions or address meta-descriptors or extend and introduce new time-methods.

A SpatDIF representation that uses extensions must declare them in the meta-section.

Use-Case: Turenas (J. Chowning)
Completed in 1972, Turenas is one of the first electronic compositions that created the impression of moving sound sources in a 360-degree space. It is composed for FM synthesizers, 4 loudspeaker channels and a reverberation unit. It is famous for its use of Lissajous figures as sound trajectories

YAML and OSC example representations
The original Turenas score is a channel-based rendering instructions for creating a decoded audio stream (stream B) on the Hardware abstraction layer:

```
<table>
<thead>
<tr>
<th>ID</th>
<th>LPH</th>
<th>LPH</th>
<th>LPH</th>
<th>LPH</th>
<th>Volume</th>
<th>Pitch</th>
<th>Pan</th>
<th>Gain</th>
<th>Volume</th>
<th>Gain</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.2515</td>
<td>1.2515</td>
<td>8</td>
<td>0</td>
<td>0.2098</td>
<td>0.1031</td>
<td>1.72</td>
<td>0.6173</td>
<td>0.6173</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.9615</td>
<td>0.6255</td>
<td>8</td>
<td>0</td>
<td>0.4387</td>
<td>0.0087</td>
<td>0.72</td>
<td>1.0986</td>
<td>1.0986</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Such channel-based instructions pose a challenge to the adaptation of the piece to other loudspeaker configurations. Knowing the loudspeaker positions and equal-power paring law applied to the gain values, we were able to “reverse-engineer” the trajectory. By using the position descriptor of the SpatDIF core, the sampling points can now be described in the time section in the following ways:

```
/spatdif/time 0.0
  name: insect
  position: 2.0 0.0 1.99 and
  time: 1.0
/spatdif/time 2.0
  name: insect
  position: 2.0 2.0 1.99 and
  time: 2.0
/spatdif/time 4.0
  name: insect
  position: 2.0 2.0 1.99 and
```

The YAML – and OSC-style representations.

Website and Downloads
The specifications and example implementations can be downloaded from http://www.spatdif.org.