SpatDIF

Spatial Sound Description Interchange Format

SpatDIF: Principles, Specification, and Examples

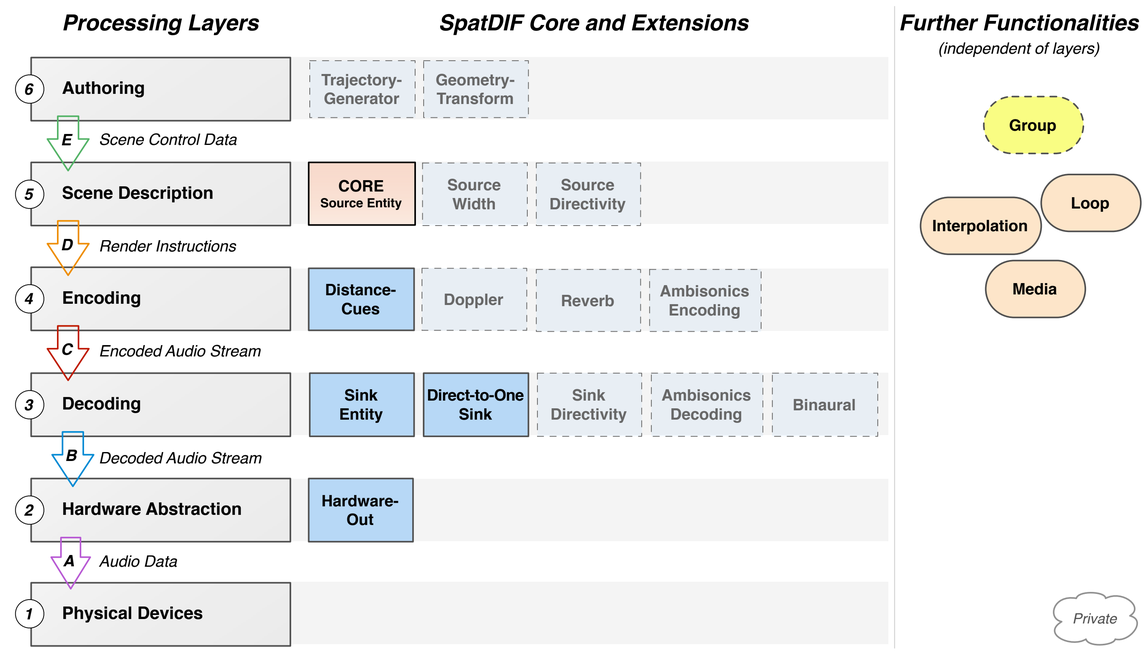
Nils Peters – ICSI, CNMAT, UC Berkeley
Trond Lossius – BEK, Bergen Center for Electronic Arts
Jan C. Schacher – ICST, Zurich University of the Arts

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.



The layer model of the SpatDIF namespace. Extensions with a dashed frame are work-in-progress.

Principles

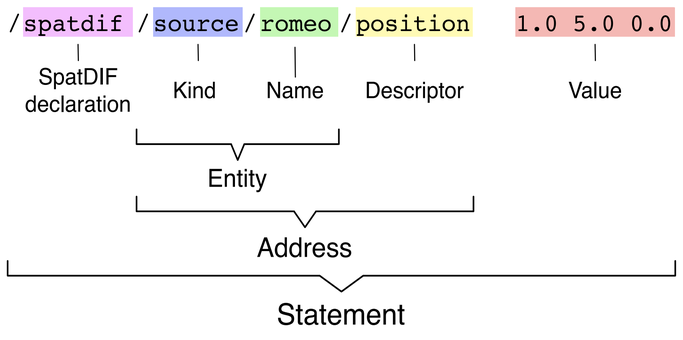
SpatDIF presents a hierarchical, unambiguous structure. The SpatDIF-syntax serves for structuring audio-scene related information.

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.



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**.

The values of descriptors may change over time. All entities and descriptors are defined within the SpatDIF namespace.

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.

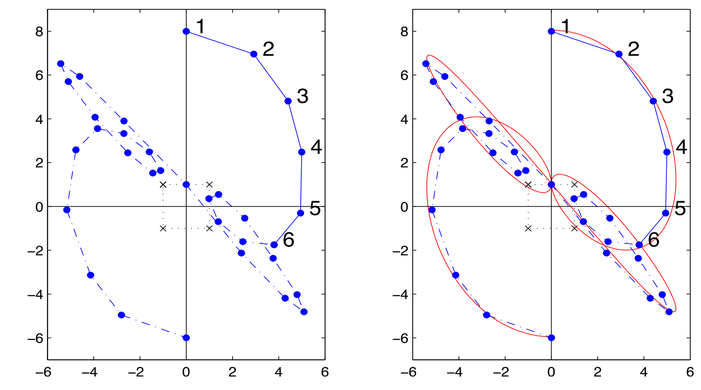
A SpatDIF representation that uses extensions must declare them in the meta-section.

Website and Downloads

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

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



The reconstructed (L) and interpolated (R) Lissajous trajectory.

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:

ID	LP1	LP2	LP3	LP4	Gain factor	Direct Ratio	Rev. Ratio	Pitch shift
1	0.7071	0.7071	0	0	0.2859	0.5347	0.4653	0.9773
2	0.7739	0.6333	0	0	0.2998	0.5475	0.4525	1.0314
3	0.8521	0.5234	0	0	0.3443	0.5867	0.4133	1.0884

Such channel-based instructions pose a challenge to the adaptation of the piece to other loudspeaker configurations. Knowning the loudspeaker positions and equal-power panning 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
    source:
        name: insect
            position: 0.0 0.0 7.99 aed
    time: 1.0
    source:
        name: insect
            position: 22.8 0.0 7.55 aed
    time: 2.0
    source:
        name: insect
            position: 42.5 0.0 6.52 aed

/spatdif/time 0.0
/spatdif/source/insect/position 0.00 7.99 0.0
/spatdif/time 1.0
/spatdif/source/insect/position 2.92 6.96 0.0
/spatdif/time 2.0
/spatdif/source/insect/position 4.41 4.81 0.0
```

The YAML– and OSC–style representations.