DOI: 10.61782/fa.2023.0729

forum acusticum 2023

RECENT ADVANCES IN THE SPATIALLY ORIENTED FORMAT FOR ACOUSTICS (SOFA, AES69)

Michael Mihocic[1*] **Piotr Majdak**[1] **Franz Zotter**[2]
Fabian Brinkmann[3] **Julien De Muynke**[4,5] **Markus Noisternig**[6]
[1] Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, A-1040 Vienna, Austria
[2] Institut für Elektronische Musik und Akustik, Inffeldgasse 10/III, 8010 Graz, Austria
[3] Audio Communication Group, Technische Universität Berlin, Einsteinufer 17c, 10587 Berlin, Germany
[4] Sorbonne Université, CNRS, ∂'Alembert, 4 place Jussieu, 75005 Paris, France
[5] Eurecat, Centre Tecnològic de Catalunya, Tecnologies Multimèdia, Barcelona, 08005, Spain
[6] STMS, IRCAM, Sorbonne Université, CNRS, Ministère de la Culture, 1 place Igor-Stravinsky, 75004 Paris, France

ABSTRACT

The spatially oriented format for acoustics (SOFA), also known as the AES69 standard, is a container file format for spatial acoustic data. SOFA can be used to describe head-related transfer functions (HRTFs), binaural, spatial, and directional room impulse responses (BRIRs, SRIRs, DRIRs), and directivities, among other data. SOFA was introduced in 2015 [1] to ease the exchange of spatial acoustic data. SOFA specifications consider structured data description, data compression, network transfer, and links to complex room geometries or other data in a hierarchical way. Since its introduction, SOFA has been embraced by many institutions and researchers, who have developed SOFA libraries for various programming environments. SOFA's recent revisions AES69-2020 [2] and AES69-2022 (as known as SOFA 2.1) [3] include a continuous representation of *Source* and *Listener* directivity composed of spherical harmonic *Emitter*s and *Receiver*s, new conventions describing the directivity of microphones, musical instruments, and loudspeakers, and new conventions describing multiple-input and multiple-output measurements of room impulse responses, enabling complex interaction between *Source*s and *Listener*s such as multiperspective representations.

Keywords: *head-related transfer function, spatialization, room impulse response, standard, file format*

1. INTRODUCTION

Until 2015 several databases for different types of acoustical data (such as head-related transfer functions, binaural and spatial room impulse responses) were collected by several labs worldwide. Each of them using an independent file format for storing data made it hard to compare and exchange data between researchers and users. The AES69 standard was established in 2015 by the AES to introduce and standardize the Spatially Oriented Format for Acoustics (SOFA) as a solution to the problem.

2. BASICS OF SOFA

SOFA stores the information in a single file by serializing the data into a binary stream. For the serialization, SOFA uses the numeric container developed by Unidata and called Network Common Data Form (NetCDF) [4]. Following the NetCDF terminology, SOFA stores all data

**Corresponding author*: *michael.mihocic@oeaw.ac.at.*

in variables and attributes. Variables are numeric or character matrices. Fig. 1 shows the elements of a SOFA file, structured in NetCDF dimensions, NetCDF variables, and NetCDF attributes.

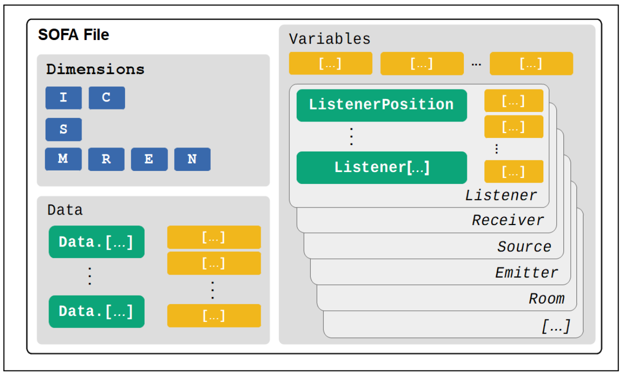

Figure 1. Elements of a Spatially Oriented Format for Acoustics (SOFA) file structured using the Network Common Data Form (NetCDF) and labeled with SOFA terminology, including SOFA objects [5].

SOFA supports storing data in spatially discrete Cartesian and spherical coordinate systems, and as spatially continuous spherical harmonics. Spatially discrete representations directly describe points in space by using one of these coordinate systems. Continuous direction representation uses the coefficients of a finite series of spherical harmonics converted from one of these coordinate systems. The geometry information can be transformed to whatever fits best to the user's internal data representation format [3].

All objects are described by their metadata. *Receiver* objects represent any acoustic sensors. The number of *Receivers* is a positive integer number greater than zero. *Receivers* can be ears of a human, or microphones in a microphone array. The *Listener* object represents a unit of all *Receivers*. There is exactly one *Listener* object per SOFA file. *Listener* objects can be: a human head, or a microphone array. *Emitter* objects represent any acoustic sources. The number of *Emitters* is a positive integer number greater than zero. *Emitters* can be vocalists in an orchestra, or loudspeaker drivers. The *Source* object represents a unit of all *Emitters*. There is exactly one *Source* object per SOFA file. *Source* objects can be: an orchestra, or a loudspeaker array. The *Room* object represents the volume enclosing the entire measurement setup; it can also be free-field. There is exactly one *Room* object per SOFA file.

The absolute position of objects is defined according to the Cartesian or spherical coordinate system within the *Room* object. *Receivers* and *Emitters* are defined in a local coordinate system, relative to the *Listener* and *Source* object, respectively.

The acoustic information, i.e., data, is represented by a group of NetCDF variables with the prefix `Data.`. The specific variable names depend on the data type used. The data type "FIR" describes the data in the time domain, which is usually a set of finite impulse responses (FIRs). The data type "TF" describes the data in the frequency domain as a set of complex-valued spectral transfer function. The data type "SOS" describes the data as a chain of second-order section (SOS) filters coefficients [5].

General metadata describe the general content of a SOFA file and are stored as NetCDF global attributes. Some metadata are mandatory (e.g. `SOFAConventions`, `SOFAConventionsVersion`, `DataType`) and must be defined according to the used conventions, or optional (e.g. `History`, `Comment`) to provide optional information in a standardized way. Metadata can be general (e.g. `SOFAConventionsVersion`), or object-related (e.g. `ListenerPosition` gives the position of the *Listener* object).

All NetCDF variables are organized as matrices of predefined dimensions. The dimensions **I** and **C** represent constant values of **I** = 1, and **C** = 3. Further mandatory dimensions represent the size of the largest string (**S**), the number of measurements (**M**), *Receivers* (**R**), *Emitters* (**E**), and the number of data samples in a single measurement (**N**).

3. CONVENTIONS

Conventions specify sets of data and metadata, commonly used in an application, for known measurement setups. Anyone can create new sets of conventions, which should be accessible to others who access the corresponding data. Proposing new conventions is possible by submitting an idea; once data is available, the convention's status is set to stable, and it may then be standardized.

SOFA 2.1 covers 15 standardized conventions for general purpose free-field databases (e.g., head- and pinna-related transfer functions), directivities (of acoustic sources and receivers), room impulse responses (DRIRs, SRIRs, BRIRs), and headphone impulse responses. Fig. 2 demonstrates a simple setup of the **Sim-**

pleFreeFieldHRIR convention, the most common convention, used e.g. for HRIR measurements. A *Source* is positioned around the *Listener*, it is defined by a discrete position. The data contain impulse responses, measured with an omnidirectional *Source* for a single *Listener*.

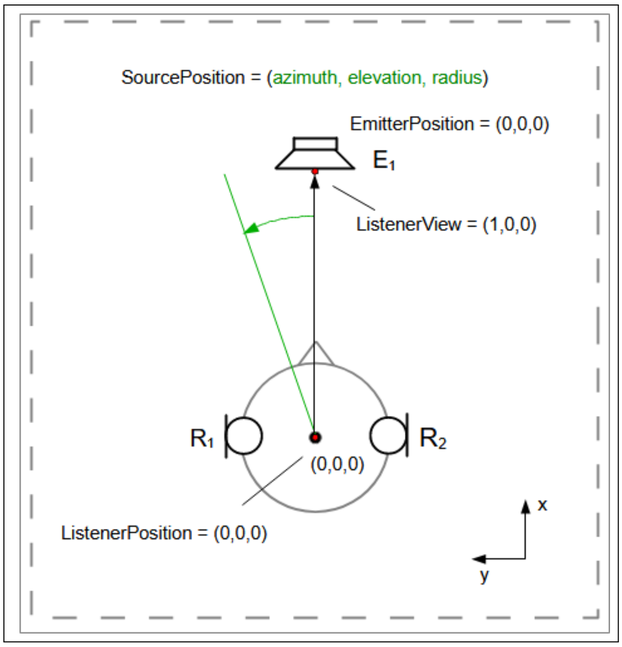

Figure 2. Simple setup of **SimpleFreeFieldHRIR** convention: A *Listener* with two *Receiver*s, and a *Source* with one *Emitter*, placed in a *Room*.

4. NEW FEATURES IN SOFA 2.X

SOFA's recent revisions include several new features, regarding representation of data and conventions [2,3]. Furthermore, the file type ".sofa" has been registered as a media type [6].

The first example of a new feature is the continuous-direction representation of data, which can be employed to represent Head-Related Transfer Functions (HRTFs) or other data without the spatial discretization of the sound-source direction. To this end, conventions such as **FreeFieldHRTF** and **FreeFieldHRIR** are provided to store HRTFs in both discrete and continuous ways. The continuous-direction representation can be used to easily interpolate between HRTF directions [5].

The second example is the source directivity which describes the directional distribution of the radiated sound energy. Although the directivity of sources such as loudspeakers can be described straightforwardly, describing the directivity of natural sources, such as singers or musical instruments, is more complicated. For example, identical musical notes can have different directivities—think of a guitar on which the same musical note can be played on different strings. The convention **FreeFieldDirectivityTF** (introduced in SOFA 2.0) aims at organizing the description of such directivities [5].

The third example is the new multiple-input multiple-output (MIMO) data representation. It represents SRIRs as FIR filters measured in a single *Room* with a single *Source* containing an arbitrary number of (by default omnidirectional) *Emitter*s (e.g., a loudspeaker array) and a single *Listener* containing an arbitrary number of (by default omnidirectional) *Receiver*s (e.g., a microphone array), with both the *SourcePosition* and *ListenerPosition* potentially varying [5].

Additional instances include the "general conventions" for representing any form of spatially oriented data. In SOFA 1.0, users had to create their own convention to represent new types of data. Initially, two conventions have been established (**GeneralFIR** and **GeneralTF**). In SOFA 2.0, these and three further conventions (**GeneralFIR-E**, **GeneralTF-E**, and their full generalization **General**) have been standardized to ease access to any type of spatial data. The most general convention is called **General**. It has minimum requirements on the information to be provided, namely, the mandatory global attributes, mandatory object-related metadata, and the data, with the data type being one of those standardized or defined by the user. This convention is thought to represent any existing and/or future acoustic setup, at the price of being less specific in its description. Conventions **GeneralFIR-E**, **GeneralTF-E**, **General-FIR**, and **GeneralTF** restrict the convention **General** to the corresponding data type. Spatially oriented data from one of these data types must be stored in one of the aforementioned conventions in case they do not fit in any of the more specific conventions [5]. The full list of supported conventions is available on the official website `https://sofaconventions.org` [7], together with their full list of specifications and requirements.

5. SUMMARY

The introduction of SOFA 2.x with the update of the AES69 standard opened up numerous exciting possibilities for representing spatial data. A significant addition is the inclusion of a method to represent continuous-direction data by means of spherical harmonics, commonly referred to as Ambisonics in the audio community. Although remaining backward compatible to the original AES Standard AES69-2015 (as known as SOFA 1.0), SOFA 2.0 introduced new conventions to describe a variety of spatial configurations, e.g., a convention describing the directivity of musical instruments and loudspeakers, with flexibility not covered by other AES standards; two conventions describing SRIR measurements enabling complex interaction between *Source*s and *Listener*s (such as MIMO data and multi-perspective representations); or conventions for a more comprehensive and flexible description of HRTFs, SRIRs, and equivalent data. Finally, general conventions allowing for a comprehensive representation of any spatial acoustic setup are provided as a basis for being further tailored down by the user. The details mentioned above have been further refined in the updated AES Standard AES69-2022, also referred to as SOFA 2.1, which is accessible to users on the SOFA website [5, 7]. A multitude of contributors from various parts of the world have supplied numerous datasets, which are obtainable through the SOFA repository [7]. SOFA is a registered media subtype [6], offering both application developers and users a convenient container for storage, transfer, and exchange of spatially oriented data.

6. ACKNOWLEDGMENT

This work was supported by the European Union (RIA action of Horizon 2020, project SONICOM, grant 101017743 to P.M.; Horizon 2020, project MOBIUS, grant 957185 to J.D.M; and Joint Programming Initiative on Cultural Heritage, project PHE, grant ANR-20-JPIC-0002 to J.D.M) and by the French National Research Agency (project RASPUTIN, grant ANR-18-CE38-0004 to M.N.).

7. REFERENCES

[1] AES Standards Comittee, *AES69-2015: AES standard for file exchange - Spatial acoustic data file format*. New York, NY, USA: Audio Engineering Society, Inc., 2015.

[2] AES Standards Comittee, *AES69-2020: AES standard for file exchange - Spatial acoustic data file format*. New York, NY, USA: Audio Engineering Society, Inc., 2020.

[3] AES Standards Comittee, *AES69-2022: AES standard for file exchange - Spatial acoustic data file format*. New York, NY, USA: Audio Engineering Society, Inc., 2022.

[4] R. Rew and G. Davis, "Netcdf: an interface for scientific data access," *IEEE Computer Graphics and Applications*, vol. 10, no. 4, pp. 76–82, 1990.

[5] P. Majdak, F. Zotter, F. Brinkmann, J. De Muynke, M. Mihocic, and M. Noisternig, "Spatially oriented format for acoustics 2.1: Introduction and recent advances," vol. 70, no. 7, pp. 565–584. Publisher: Audio Engineering Society.

[6] "Media type: Audio/sofa - iana.org."

[7] "SOFA (spatially oriented format for acoustics) - sofaconventions.org."