Audio Engineering Society

Convention Paper

Presented at the 134th Convention
2013 May 4–7 Rome, Italy

This Convention paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This convention paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions

Piotr Majdak[1], Yukio Iwaya[2], Thibaut Carpentier[3], Rozenn Nicol[4], Matthieu Parmentier[5], Agnieszka Roginska[6], Yôiti Suzuki[7], Kanji Watanabe[8], Hagen Wierstorf[9], Harald Ziegelwanger[1], and Markus Noisternig[3]

[1] Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria
piotr@majdak.com, h.ziegelwanger@me.com

[2] Department of Electrical Eng. and Information Technology, Tohoku Gakuin University, Tagajo, Japan
iwaya@ieee.org

[3] UMR STMS IRCAM-CNRS-UPMC, Paris, France
thibaut.carpentier@ircam.fr, markus.noisternig@ircam.fr

[4] Orange Labs, France Telecom, Lannion, France
rozenn.nicol@orange.com

[5] France Television, Paris, France
matthieu.parmentier@francetv.fr

[6] Music Technology, New York University, New York, NY, USA
roginska@nyu.edu

[7] Research Institute of Electrical Communication, Tohoku University, Sendai, Japan
yoh@riec.tohoku.ac.jp

[8] Faculty of Engineering, Akita Prefectural University, Yuri-Honjo, Japan
kwatanabe@akita-pu.ac.jp

[9] Telekom Innovation Laboratories, Technical University of Berlin, Berlin, Germany
hagen.wierstorf@telekom.de

ABSTRACT

Head-related transfer functions (HRTFs) describe the spatial filtering of the incoming sound. So far available HRTFs are stored in various formats, making an exchange of HRTFs difficult because of incompatibilities between the formats. We propose a format for storing HRTFs with a focus on interchangeability and extendability. The spatially oriented format for acoustics (SOFA) aims at representing HRTFs in a general way, thus, allowing to store data such as directional room impulse responses (DRIRs) measured with a microphone-array excited by a loudspeaker array. SOFA specifications consider data compression, network transfer, a link to complex room geometries, and aim at simplifying the development of programming interfaces for Matlab, Octave, and C++. SOFA conventions for a consistent description of measurement setups are provided for future HRTF and DRIR databases.

1. INTRODUCTION

Head-related transfer functions (HRTFs) describe the spatial filtering of the incoming sound due to the listener's anatomy. HRTFs are crucially important for the binaural reproduction of virtual acoustics. HRTFs have been measured by a number of laboratories and are typically stored in each lab's native file format. While the different formats are of advantage for each lab, an exchange of such data is difficult due to incompatibilities between formats.

In this work, we propose specifications for an HRTF-data exchange format with a special focus on interchangeability and extendability. The spatially oriented format for acoustics (SOFA) aims at representing spatial data in a general way, allowing to store not only HRTFs but also more complex data, e.g., directional room impulse responses (DRIRs) measured with a multichannel microphone array excited by a loudspeaker array. In order to simplify the adaption of SOFA for various applications, examples of implementation of the format specifications are provided together with a collection of exemplary data sets converted to SOFA.

The AES-X212 HRTF file format standardization project is based on the SOFA format and was recently approved by the AES subcommittee SC-02 and assigned to the working group SC-02-08 on audio file interchange.

1.1. Typical measurement setups

One of the first publicly available HRTFs were those of a dummy-head microphone measured in an anechoic room [1]. Two microphones placed at the ear simulators were used for the recordings and one loudspeaker was used for the signal excitation. The loudspeaker was moved to the desired elevation and the mannequin was rotated to the desired azimuth. Taken together, HRTFs for 710 spatial positions were measured at elevations from -40° to +90° in steps of 10° and 360° azimuthal range in steps of 5° and a constant distance of 1.4 m. The HRTFs are provided as impulse responses (IRs) with the length of 512 samples at a sampling rate of 44.1 kHz.

One of the first publicly available HRTFs measured in *human listeners* was the CIPIC database [2]. The measurements were performed at a constant distance of 1 m for 1250 spatial directions around the listener. The HRTFs are available for 43 listeners as IRs of 200 samples at a sampling rate of 44.1 kHz. Since then many other HRTF/DRIR databases have been made publicly available [3–8].

All those measurement setups have the following properties in common. In an anechoic chamber or in a room, excitation signals are generated and microphones are used to record the incoming signals (see Fig. 1). The measurement is repeated while varying the spatial posi-

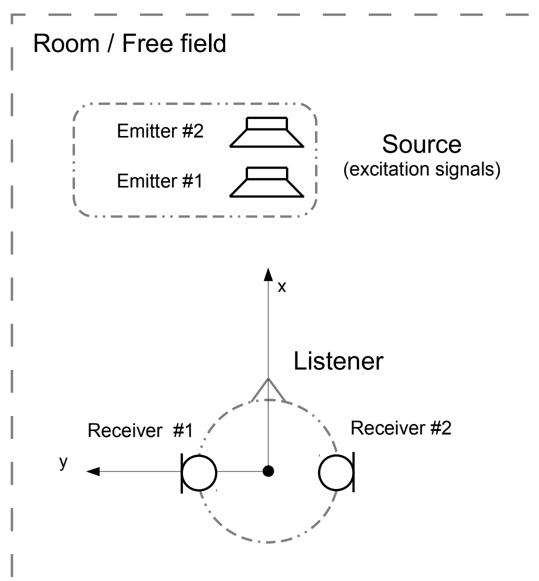

Figure 1: Typical HRTF/DRIR measurement setup.

tion of the excitation source relative to the listener, which is done by varying the position of the listener, the sound source, or both in different dimensions.

Binaural HRTF measurement setups use only two microphones to record the left and right ear signals. However, HRTF/DRIRs measurements may also consider multiple microphones, e.g., three microphones per head side in hearing-assist devices [7], tens of microphones arranged in an array structure at different directions and distances from the center [9], a multichannel microphone array arranged around the listeners in a reciprocal HRTF measurement system [10], [11], multichannel microphone arrays for measuring DRIRs [12] or various microphone positions in a room, e.g., for concert-hall acoustics measurements [13]. As a generalization, microphones and an object comprising those microphones can be identified. Thus, in this article, a microphone as the single receiver of the sound field is called the *receiver*, and the comprising all the receivers is called the *listener,* see Fig. 1.

The sound source used for the excitation signal is not necessarily a single point source. Loudspeaker arrays were used, either to control the sound field surrounding the listener, e.g., wave-field synthesis [11], [14], [15], or higher-order Ambisonics [16], [17] or to control the radiation characteristics of the sound sources [18]. Similarly to the concept of listener and receivers, in this article, the particular sources creating the excitation signal are called *emitters* and the object comprising the emitters is called *source*. Note that a measurement setup with a source with multiple emitters and a listener with multiple receivers has already been considered [19].

In typical HRTF measurements, only the direction of the incoming signal is varied. In more recent setups also different sound-ear distances have been considered [4], [11], [20]. However, sometimes the variation of other parameters is of interest. For example, HRTFs were measured as a function of the head orientation relative to the torso [21], or the room IRs were measured as a function of the room temperature [22]. An HRTF file format should thus consider even such parameters.

1.2. Existing data formats

Until now, HRTFs are stored using different formats, all of them having advantages and disadvantages. The CIPIC database [2] provides a file per listener in either a plain text or Matlab (Mathworks, Inc.) file format. The directions are hard coded, i.e., the index of an HRTF corresponds to a predefined direction used in the measurements. While the representation of HRTFs from other directions is not allowed, anthropometric data have been stored within that format. The openDAFF package[1], while similarly storing HRTFs only in a regular angular distance, uses a key-value system for the description of the metadata which seems to be very promising. Other databases such as LISTEN [3] and ARI [6], consist of an HRTF matrix and additional matrices describing the direction of the corresponding HRTF, thus, allowing to represent HRTFs from any direction. In that formats, HRTFs from each listener are stored in a separate file. In the database storing the HRTFs as a function of distance [4], the data are stored in a separate file for each distance. Combined with the necessity to store a separate file for each listener, those three latter formats would result in many files. The MARL-NYU database [23] harmonized the format of CIPIC, LISTEN, MIT, and others databases, and stores all those data in a single file. This concept seems to be promising when combined with a network interface and partial file access in the future. Most of those HRTFs are stored in Matlab formats, i.e., they use a Matlab file convention to store predefined matrices. In contrast, SDIF [24], a general format for storing audio-related data, has been adapted to HRTFs, allowing to store HRTFs of a single listener in a mixed text-based and binary representation. The concert-hall data [25], stored as compressed ".wav" files, are another example for a mixed-binary format, which further requires a description (separate text files) in order to being able to interpret the data. The HRTFs measured in rooms (e.g., [8]) are also Matlab files and the relationship between the data and the geometry of the measurement setup is provided in separate publications.

From this summary, the requirements on a file format storing HRTFs and spatially oriented descriptions of acoustic systems are derived:

- Description of a measurement setup with arbitrary geometry, i.e., not limited to special cases like a regular grid, or a constant distance;
- Self-describing data with a consistent definition, i.e., all the required information about the measurement setup must be provided as metadata in the file;
- Flexibility to describe data of multiple conditions (listeners, distances, etc) in a single file;
- Partial file and network support;

1 see http://www.opendaff.org/

- Available as binary file with data compression for efficient storage and transfer;
- Predefined description conventions for the most common measurement setups.

SOFA aims at fulfilling all those requirements. SOFA specifications are described in the following sections. A HRTF/DRIR measurement setup is described by various objects (Sec. 2.1) and their relations (Sec. 2.2). The information is stored in a numeric container (Sec. 2.3) and structured by the *measurement*. Measurement is a discrete sampled observation done at a specific time and under a specific condition. A measurement consists of data, e.g., an IR (Sec. 2.4), and is described by its corresponding dimensions (Sec. 2.5) and metadata (Sec. 2.6). All measurements are stored in a single data structure, e.g., a matrix of IRs. Conventions for a consistent description of measurement setups are provided for two setups (Sec. 3).

2. GENERAL SPECIFICATIONS

2.1. Objects

Receiver is any acoustic sensor like the ear or a microphone. The number of receivers in not limited in SOFA and defines the size of the data matrix.

Listener is the object incorporating all the receivers. For HRTFs, a listener can be a head or dummy-head microphone. For DRIRs, a listener represents the microphone-array structure such as a sphere or a frame. Incorporating the receivers in the listener as a single logical object is important because in measurements, usually the orientation and/or position of the listener vary without substantial changes in the head-microphone relation. For example, in measurements done for multiple positions in a room, the position of the head varies and the relation between the head and the microphones does not change. Note that only one listener is considered.

Emitter is any acoustic excitation used for the measurement. The number of emitters is not limited in SOFA. The contribution of the particular emitter is described by the metadata (see later).

Source is the object incorporating all emitters. In SOFA, source might be a multi-driver loudspeaker (with the particular drivers as emitters), or a speaker array (with the particular speakers as emitters), or a choir (with the particular human as emitter), etc. Note that only one source is considered but the source may incorporate an unlimited number of emitters.

Room is the volume enclosing the measurement setup. In the case of a free-field measurement, the room is not considered. An optional room description is considered for measurements performed in reverberant spaces, with a direct description of a simple shoebox, or with a link to a digital asset exchange file for a more complex description.

Optional *Objects* can be defined by including additional parameters of a measurement. For example, this might be the torso as in the measurements in which the angle between the torso and the head is varied as an independent variable.

2.2. Relation between the objects

We use two coordinate systems. Source and listener are defined in the coordinate system of the room, called global coordinate system. In free field, the global coordinate system is arbitrary.

Emitters and receivers have both their own coordinate system called local coordinate system. The local coordinate system of emitter and receiver are defined relatively to the coordinate system of the source and listener, respectively. With the source and listener in the origin and at default orientation, the local coordinate systems correspond to the global coordinate system.

Two vectors describe the basic orientation of the source/listener: the “view” vector defines the direction in which the source/listener looks; the “up” vector defines the top of the source/listener. In spherical coordinates, the view vector describes the azimuth and elevation angles of the source/listener. The up vector describes the roll, which is usually not considered in HRTF measurements and is optional. If given, we suggest the up vector to be orthogonal to the view-vector. The default basic orientation for the source/listener is the view vector on x-axis and the up vector on z-axis.

In order to be flexible in the future, the way the position and orientations are defined is specified separately for the listener, source, all emitters, and all receivers. The default coordinate type for the position, view, and up vectors is the Cartesian (x y z). When the spherical coordinate system is required, the format is (azimuth elevation distance).

The source/listener basic rotations can be further modified. Most HRTF measurements consider only rotations

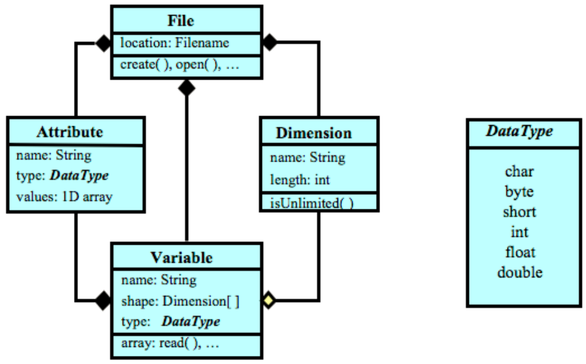

Figure 2: Classic netCDF data model. Figure from Unidata (http://www.unidata.ucar.edu)

described by the azimuth and elevation angles. These two angles provide the possibility to describe the rotation of the listener in an intuitive way. However, for arbitrary rotations in the 3-dimensional space the exact order of the rotations becomes important. Rotation descriptions like the "yaw-pitch-roll" system (which is known as DIN 9300 for aviation and more intuitive) or the unit quaternions (which avoid the gimbal lock and are computationally efficient) clearly define the order of rotation. Note that a complete agreement on the coordinates and coordinate systems has not been done yet.

2.3. Numeric container

SOFA stores the information in a single file by serializing the data into a binary stream. The serialization is usually done by a numerical container, which defines the format of the binary representation. SOFA files have the extension ".sofa".

In order to avoid custom development of a numerical container, SOFA relies on netCDF-4 (Unidata), which is a set of software libraries and data formats supporting the creation, access, and sharing of scientific data.[2] It is self-describing, network-transparent, and machine-independent; it supports huge files, partial access within a file, and allows for data compression. netCDF-4 is widely used in the field of climatology, meteorology, oceanography, and geographic information systems. It is based on the HDF5 (HDF5 Group)[3], a more basic numerical container, further supported by many institutions worldwide. For SOFA, netCDF offers a structured representation of multidimensional data and metadata. The open-access specifications are freely available and include a complete definition as well as examples of various implementations. Application-programming interfaces are available as pre-compiled libraries for programming languages like C++, Octave, and JAVA. Note that netCDF is natively supported in Matlab.

netCDF considers *conventions*, a set of recommendations in a community on the naming of attributes, variables, and dimensions within a netCDF file. Many conventions exist, mostly in the field of climate and geographical research.[4] SOFA proposes conventions related to the HRTF/DRIR measurement. In particular, SOFA conventions are proposed for typical HRTF/DRIR measurement setups. According to the netCDF terminology, SOFA defines *dimensions* and stores data in *variables* and *attributes*.

SOFA uses the so-called enhanced data model from netCDF-4, which is based on the classic netCDF data model shown in Fig. 2. Since the enhanced data model is more complex and not well spread in various computer systems yet, we mostly use the classic data model parts from the enhanced model. This way allows a simple data representation but still full flexibility in the future. More deep knowledge of netCDF format details is not required to read or write netCDF datasets. More interested readers are referred to the User's Manual.[5]

Note that in SOFA, we sometimes refer to the data type "string", which is defined in the enhanced data model but is not provided in the classic model. Currently, some computer programs like Matlab and Octave have difficulties handling netCDF strings in a proper way, thus, at the moment, we encode strings as character arrays according to the classic model. Native support of strings and string arrays is planned after clarification of the technical requirements.

2.4. Data

Data represent the numeric description of the acoustic systems and consist of a multidimensional matrix of an arbitrary size. Data stored in this format have the flexibility to be in the domain that best accommodates the measurement and measurement system. Data can be time domain finite IRs (data type FIR) or infinite IR filter coefficients (IIRBiquad), with or without separately stored broadband delays. The broadband delay (i.e.,

2 see http://www.unidata.ucar.edu/software/netcdf/
3 see http://www.hdfgroup.org/HDF5
4 see http://www.unidata.ucar.edu/software/netcdf-java/formats/UnidataObsConvention.html and http://cf-pcmdi.llnl.gov/
5 http://www.unidata.ucar.edu/software/netcdf/docs/user_guide.html

Data type	Field	Size	Description
FIR	IR	[M R N]	IRs as a function of the sampling interval N.
FIRplusTOA	IR TOA	[M R N] [M R]	IRs as a function of sampling interval N. The IRs are supposed to be without any broadband delay (TOA), which is separately provided for each receiver in TOA.
IIRBiquadplusTOA	G (linear gain) B1, B2 A1, A2 TOA	[M R] [M R N] [M R N] [M R]	Biquad filters representation by the nominator (B1, B2), denominator (A1, A2), the gain (G), and broadband delay (TOA). N represents the number of biquads.
FIRnoTOA	IR TOAModel	[M R N] [M 5]	IRs as a function of sampling interval N without the TOA, which is modeled and represented by parameters TOAModel.

Table 1: Data types considered in SOFA. Multiple entries for the size separated by "," indicate that multiple sizes can be supported. For all data types, the parameter SamplingRate (in Hz) is mandatory.

time-of-arrival, TOA) can be stored as discrete delays in a matrix or as parameters of continuous-directional TOA model [26]. Data contain fields (e.g., Data.IR, Data.G) which are functions of the dimension N. The interpretation of N depends on the data type, e.g., for IRs, N represents the sampling interval (i.e., inverse of the sampling rate) or the number of FIR-filter taps. The interpretation is denoted in the attributes of the dimension variable N. The different data types and corresponsing fields are shown in Tab. 1.

Theoretically, the HRTFs/DRIRs (as a function of discrete spatial position) can be transformed to functions of continuous spatial frequency and represented in the spherical-harmonic (SH) domain. Advantages like the directional continuity or better compactness are the main reasons for such a representation. Even though not provided at the moment, SOFA aims at considering SH data in future conventions (see Sec. 3).

2.5. Dimensions

Each netCDF variable has fixed dimensions and its dimensions must be defined before creating the variable. Thus, in SOFA, the following dimensions are pre-defined as netCDF dimensions:

Name	Description	Default
Conventions	Declaration of a SOFA convention, read only	SOFA
DataType	Specifies the data type	FIR
SOFAConvention	Specific SOFA convention used for the measurement setup	-
SOFAConventionVersion	Version of the SOFA convention	-
ApplicationName	Name of the application editing the file	-
ApplicationVersion	Version of the application editing the file	-
AuthorContact	Contact information (e.g., email) of the author	-
License	Legal license under which the data are provided.	CC BY-SA 3.0
Copyright	Legal name of the copyright holder	-
Organization	Legal name of the organization of the author	-
History	Audit trail for modifications to the original data	-
Comment	Information about the data or methods used to produce it	-
RoomType	Defines the room type ("free field", "shoebox", "dae")	free field
DatabaseName	Name of the database	-
DatabaseTimeCreated	Date and time of creation of the file, in ISO 8601	-
DatabaseTimeModified	Date and time of the last modification of the file, ISO 8601	-
SubjectID	ID of the listener used for the measurement	-

Table 2: General metadata considered in SOFA, stored as global attributes in the netCDF file.

Name	**Size**	**Description**	**Default**
ListenerPosition	[C], [M C]	Position	(0 0 0)
ListenerPosition:Type	A	Type of coordinate system used for the position	cartesian
ListenerPosition:Unit	A	Unit of the coordinates	meter
ListenerView	[C], [M C]	View vector for the basic orientation	(1 0 0)
ListenerUp	[C], [M C]	Up vector for the basic orientation	(0 0 1)
ListenerRotation	[C], [M C]	Rotation for the additional modification of the basic orientation	optional
ListenerRotation:Type	A	Type of the rotation	din9300
ListenerRotation:Unit	A	Unit of the coordinates used for the rotation	degrees
ListenerDescription	String	Informal description of the listener	-

Table 3: Listener metadata considered in SOFA. Similar applies to the source. Multiple entries for the size separated by “,” indicate that multiple sizes can be supported. A: Attribute.

Name	**Size**	**Description**	**Default**
ReceiverPosition	[R C], [R C M]	Position, in the local coordinate system	(0 0 0)
ReceiverPosition:Type	A	Type of coordinate system used for the position	cartesian
ReceiverPosition:Unit	A	Unit of the coordinates	meter
ReceiverView	[R C], [R C M]	View vector for the orientation	(1 0 0)
ReceiverUp	[R C], [R C M]	Up vector for the orientation	(0 0 1)
ReceiverDescription	String	Informal description of the receivers	-

Table 4: Receiver metadata considered in SOFA. By replacing R by E, similar applies to the emitter metadata.

Room Type	**Parameter**	**Size**	**Description**	**Default**
free field	none	-	-	-
shoebox	RoomCorner	[2 C], [2 C M]	Coordinate [A B] of the shoe box	-
	RoomCorner:Type	A	Type of coordinate system used for the room	cartesian
	RoomCorner:Units	A	Units of coordinates	meter
	RoomCorner:Description	A	Informal description of the room	-
dae	RoomFileName	String	Filename of the DAE file describing the room	*
	RoomFileName:Description	A	Informal description of the room	a room

Table 5: Room types. A: Attribute. “*”: empty string

- M: number of measurements;
- R: number of receivers;
- N: number of data samples describing one measurement. Data is a function of N;
- E: number of emitters;
- C: coordinate dimension, always three with the meaning depending on the coordinate type;
- Q: quaternions (optional).

Data and metadata are described by using these dimensions. In order to take care for other dimensions that might be required to define a variable, additional dimension variables are created by appending “DIM” to the name of the corresponding variable.

Throughout this document, the matrix sizes are denoted by [A_1 A_2 ... A_I] where A_i represents the length of the dimension i of the I-dimensional matrix. For example, assume a database consisting of one thousand measurements, i.e., M = 1000, obtained for 1000 different rotations of the listener, i.e., ListenerRotation is [M C], using two microphones, i.e., two IR per measurement, and sampling rate of 48 kHz. Then, in the netCDF file, M =

1000, R = 2, and C = 3. Further, the netCDF variables "Data.FIR", "ListenerRotation", and "Data.SamplingRate" have dimensions [M R N], [M C], and [1], respectively.

2.6. Metadata

Metadata consist of variables and their attributes. General metadata (Tab. 2) consider the most important properties of the measurement and are valid for the global measurement setup. General metadata are represented as global attributes in netCDF.

Other metadata can be a matrix of numeric (integer or float) variables or a string. Attributes can accompany a variable where appropriate. Object-specific metadata consider the description of objects listener, receivers, source, emitters (Tabs. 3 and 4). Room-specific metadata describe the room used in the measurements and depend on the attribute RoomType (Tab. 5). Measurement metadata describe other measurement-specific data like the time of a particular measurement (MeasurementTime) and have the prefix "Measurement".

In order to keep it simple, nested structures within the metadata are not allowed, but grouping by prefixes, e.g., ListenerPosition and ListenerOrientation is encouraged. Attributes for the geometry description (e.g., source position, listener orientation) extend a value by further coordinate triplet C. When saved as a variable, date and time uses integer number of seconds from 1974-02-22 00:00:00. When saved as attributes string in ISO 8601 format "yyyy-mm-dd HH:MM:SS" is used.

3. SOFA CONVENTIONS

In order to meet the different requirements coming from different application fields, SOFA conventions are proposed, i.e., definitions of data and metadata consistently describing particular HRTF/DRIR measurement setups. SOFA conventions are specified in the Common Data Form Language (CDL) as suggested for netCDF. CDL files allow for a platform-independent interpretation of a convention and can be compiled to a binary netCDF file using the ncgen tool from the netCDF package:

```
ncgen -b -o mySOFAfile.sofa -k3
mySOFAconvention.cdl
```

In the following, two SOFA conventions are described. Note that instead of aiming at foreseeing the future, conventions should be developed only for known measurement setups. The known features should be consistently described while not limiting the development of future conventions.

3.1. Free-field HRTF measurement of a single listener (SimpleFreeFieldHRTF)

This convention essentially defines the setup used in the ARI, LISTEN, FIU, CIPIC, and other similar HRTF databases (Fig. 2). The measurements are done at a constant distance in free field with a single excitation source assuming an omnidirectional loudspeaker. Human listeners are considered and thus, the number of receivers is two. Azimuth and elevation angles are varied and the tilt of the head is not considered during the measurement. The measured HRTFs are represented as FIR filters, with a single HRTF set of a listener per file.

The SOFA convention SimpleFreeFieldHRTF defines such a setup. It requires:

- *General attributes*: SOFAConvention: SimpleFreeFieldHRTF, Datatype: FIR, RoomType: free field, the other general attributes are arbitrary.
- *Data*: The number of receivers is two, thus, the size of Data.FIR is [M 2 N]. Only a single sampling rate for the file is allowed, thus SamplingRate is a scalar.

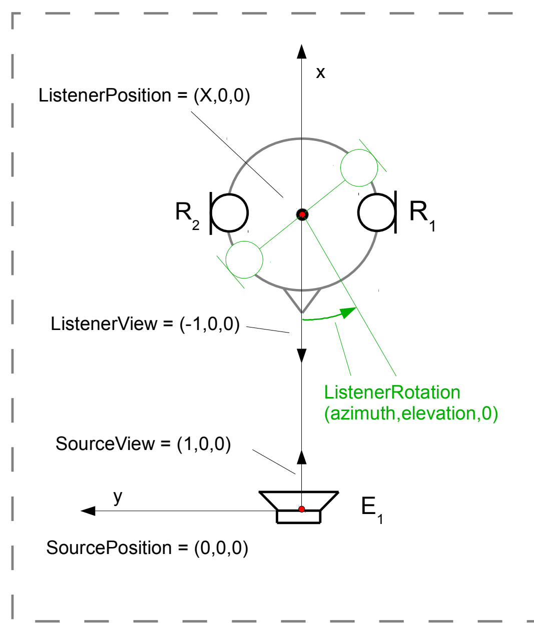

Figure 3: Free-field HRTF measurement setup considered in the SOFA convention "SimpleFreeFieldHRTF"

- *Source*: Source is in the origin of the setup and consists of a single emitter. SourcePosition: (0 0 0), SourceUp: (0 0 1), SourceView: (1 0 0), SourceRotation, and EmitterPosition: (0 0 0).
- *Listener*: The listener is in the measurement distance X (in meter) from the source, facing the source. ListenerPosition: (X 0 0), ListenerView: (0 0 0), ListenerUp: (X 0 1).
- *Receivers*: Two receivers (=ears) on a head with radius H (in meter): ReceiverPosition: (0 -H 0; 0 +H 0).
- The different azimuth and elevation angles of the measurement are described by the ListenerRotation as [M 3] matrix (in degrees). The coordinate type is “din9300”.

The corresponding CDL file is provided in the Convention directory of the SOFA package.

3.2. Microphone-array measurement in a room (SimpleDRIRMicArray)

This convention defines a setup used for measuring DRIRs in a room with a microphone array (Fig. 4). The setup considers measurements in a room with a single excitation source and a microphone array with an arbitrary number of omnidirectional microphones (i.e., receivers). Because of the special interest in spherical microphone arrays for DRIR measurement, the spherical coordinate system is considered for the description of the receivers. The positions of both the source and the listener are considered as variant. The DRIRs are represented as FIR filters for a single room per file. SimpleDRIRMicArray consists of:

- *General metadata*: SOFAConvention: SimpleDRIRMicArray, Datatype: FIR, RoomType: dae. Other general metadata are optional.
- *Data*: The amount of the receivers varies, thus, the size of Data.FIR is [M R N]. Only one sampling rate is allowed, thus SamplingRate is [1] (in Hz).
- *Source*: The position and the orientation of the source may vary by providing SourcePosition, SourceUp, SourceView. The coordinate type is cartesian and unit is meter.
- *Emitters*: Source consists of a single omnidirectional emitter, which position is fixed, EmitterPosition: (0 0 0). EmitterUp and EmitterView are omitted.

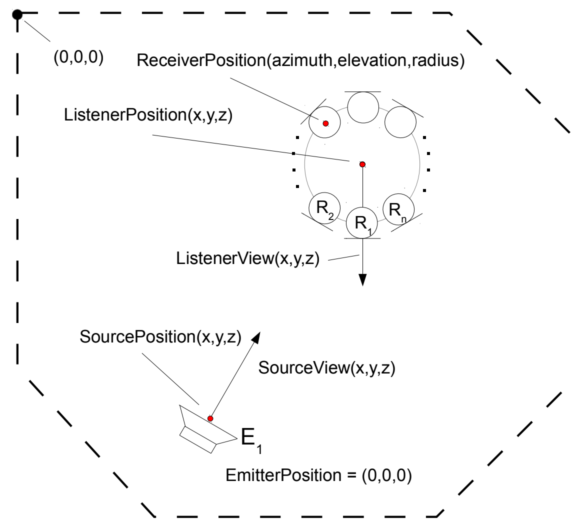

Figure 4: DRIR measurement setup in a room considered in the SOFA convention “SimpleDRIRMicArray"

- *Listener*: Position and the orientation of the listener vary and thus ListenerPosition, ListenerView, and ListenerUp are mandatory. The coordinate type is cartesian and unit is meter.
- *Receivers*: the position of all receivers via ReceiverPosition is provided. The coordinate type is spherical and unit is “degrees, degrees, meter”. The receivers are considered to be omnidirectional, thus, ReceiverUp and ReceiverView are omitted.
- An additional rotation of the listener is not considered, thus, ListenerRotation is not provided.
- For each measurement, an ID and the time stamp is considered, with MeasurementID as [M strlen] and MeasurementTimeCreated as [M].[6]

The corresponding CDL file is provided in the “Convention” directory of the SOFA package.

4. TECHNICAL ASPECTS

4.1. Application Programming Interface

SOFA specifications also consider an application-programming interface (API) with similar calls for various programming languages (Matlab, Octave, C++) and computer platforms.

6 strlen is the maximum length of all strings, 64 characters per default

For Matlab and Octave, the API provides functionality to create, read, and write SOFA files. The data and metadata are handled in structures considering consistency checks of all information. A variable search is provided to find data entries for specific attributes, which allows for loading only the relevant subset of the file. Behind the user functions there are two different sets of low-level functions built on top of the netCDF-library support in Matlab and the netCDF Toolbox for Octave.

The SOFA C++ API is quite similar to the Matlab API; it is developed as a layer on top of the C-based netCDF library and provides read/write support for SOFA files. Numerical data and metadata can be efficiently accessed in whole or in part. It further allows converting SOFA files to human readable plain text and vice versa. Any C-compliant (C/C++/Objective-C) application can link against the SOFA library. The API interfaces use portable primitives to ensure cross-platform compatibility. The library tackles issues such as endianness and platform-dependent file I/O so that application programmers should not be concerned by the low-level file structure. Memory for numerical arrays can either be allocated by the library or passed in by the programmers. The API calls are intended to be rather similar across the different languages (Matlab, Octave, C++).

4.2. Networking

In principle, netCDF files can be transferred via networks by using the Open Data Access Protocol (OpenDAP), which is a protocol for providing local data to remote locations regardless of local storage format.[7] SOFA, being technically speaking a netCDF convention, should be able to use OpenDAP.

4.3. Directionality

The directionality of the emitters and/or receivers is sometimes of interest. The measurement and reproduction of sound radiation patterns is topic of ongoing research (see e.g., [27]) and the wave-field expansion in spherical coordinates is a potential candidate for a description format that is independent of the sound-field rendering technique and scalable in terms of reproduction detail. SOFA, as a basis for the description of directionality, allows for arbitrary spectral and spatial resolutions and addresses directional characteristics of sound sources when applying surrounding microphone arrays. Corresponding future SOFA conventions are planned, considering compatibility to industry standards such as the common loudspeaker description format (CLF).[8]

7 see http://opendap.org

5. CONCLUSIONS AND OUTLOOK

SOFA and its conventions aim at describing data resulting from various HRTF/DRIR measurement setups. This includes but is not limited to free-field HRTF measurements, BRIR measurements at various listener positions, microphone-array measurements in a room with a line-array of speakers, measurements of the directivity patterns of a spherical microphone array, BRIR measurements with variable torso, or room impulse responses as a function of the room temperature.

Currently, SOFA is in the development phase. Two conventions are available, together with a draft of the API for Matlab. The SOFA package with its current development status is accessible at SourceForge.[9] For the development and evaluation of the conventions, the binaries of the nc-tools are available at the Unidata.[10] For debugging and numeric representation of the binary SOFA files, HDF5Viewer is available at the HDF5-Group.[11]

Two SOFA conventions have been defined and both are under review of the contributing organizations. Further, the AES standardization committee has considered the standardization of an HRTF file format as highly relevant to facilitate the personalization of binaural rendering. Focusing on a patent-free and scalable solution, the standardization project AES-X212 builds upon SOFA as a future HRTF format. Contributions to the draft from the audio industry are welcome to ensure a wide dissemination of personalized binaural audio solutions.

6. ACKNOWLEDGMENTS

We thank Wolfgang Hrauda for valuable contributions during his internship. This study is supported by the French project Binaural Listening (BiLi, FUI-AAP14) and the Austrian Science Fund (FWF, P 24124-N13).

7. REFERENCES

[1] W. G. Gardner and K. D. Martin, “HRTF measurements of a KEMAR,” *J Acoust Soc Am*, vol. 97, pp. 3907–3908, 1995.

[2] V. R. Algazi, R. O. Duda, D. M. Thompson, and C.

8 see http://www.clfgroup.org/

9 see http://sf.net/projects/sofacoustics

10 http://www.unidata.ucar.edu/software/netcdf/docs_rc/winbin.html

11 see http://www.hdfgroup.org

Avendano, "The CIPIC HRTF database," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001, pp. 99–102.

[3] O. Warusfel, "LISTEN HRTF database," 2003. http://recherche.ircam.fr/equipes/salles/listen/.

[4] H. Wierstorf, M. Geier, A. Raake, and S. Spors, "A free database of head-related impulse response measurements in the horizontal plane with multiple distances," in *130th Convention of the Audio Engineering Society (aes)*, 2011, ebrief 6.

[5] T. Nishino, S. Kajita, K. Takeda, and F. Itakura, "Interpolation of head related transfer functions of azimuth and elevation," *J Acoust Soc Jpn*, vol. 57, pp. 685–692, 2001.

[6] P. Majdak, M. J. Goupell, and B. Laback, "3-D localization of virtual sound sources: effects of visual environment, pointing method, and training.," *Atten Percept Psychophys*, vol. 72, pp. 454–69, 2010.

[7] H. Kayser, S.D.Ewert, J.Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP J Advances Sig Proc*, Article ID 298605, 10 pages, 2009.

[8] M. Jeub, M. Schäfer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *16th International Conference on Digital Signal Proc*, 2009, pp. 1–5.

[9] I. Balmages and B. Rafaely, "Open-sphere designs for spherical microphone arrays," *IEEE Trans Audio Speech Lang Proc*, vol. 15, pp. 727–732, 2007.

[10] D. N. Zotkin, R. Duraiswami, E. Grassi, and N. A. Gumerov, "Fast head-related transfer function measurement via reciprocity," *J Acoust Soc Am*, vol. 120, pp. 2202–2215, 2006.

[11] M. Pollow, M., Nguyen, K.-V., Warusfel, O., Carpentier, T., Müller-Trapet, M., Vorländer, M., and Noisternig, "Calculation of head-related transfer functions for arbitrary field points using spherical harmonics decomposition," *Acta Acust United Ac*, vol. 89, pp. 72–82, 2012.

[12] B. Khaykin, D., and Rafaely, "Acoustic analysis by spherical microphone array processing of room impulse responses," *J Acoust Soc Am*, vol. 132, pp. 261–270, 2012.

[13] T. Pätynen, J., Tervo, S., and Lokki, "Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses," *J Acoust Soc Am*, vol. 133, pp. 842–857, 2013.

[14] A. J. Berkhout, "Holographic approach to acoustic sound control," *J Audio Eng Soc*, vol. 36, pp. 977–995, 1988.

[15] S. Ahrens, J., and Spors, "Wave field synthesis of a sound field described by spherical harmonics expansion coefficients," *J Acoust Soc Am*, vol. 131, pp. 2190–2199, 2012.

[16] M. A. Gerzon, "Ambisonics. Part two: Studio techniques," *Studio Sound*, vol. 17, pp. 24–26, 1975.

[17] M. Zotter, F., Pomberger, H., and Noisternig, "Energy-preserving ambisonic decoding," *Acta Acust United Ac*, vol. 98, pp. 37–47, 2012.

[18] B. Rafaely, "Spherical loudspeaker array for local active control of sound," *J Acoust Soc Am*, vol. 125, pp. 3006–3017, 2009.

[19] S. Clapp, A. Guthrie, J. Braasch, and N. Xiang, "The use of multi-channel microphone and loudspeaker arrays to evaluate room acoustics," in *Proceedings of the Acoustics*, 2012, vol. 131, p. 3208.

[20] S. Hosoe, K. I. Takanori Nishino, and K. Takeda, "Development of micro-dodecahedral loudspeaker for measuring head-related transfer functions in the proximal region," in *Proceedings of the IEEE Conference on Audio, Speech and Signal Processing (ICASSP)*, 2006, pp. 329–332.

[21] M. Guldenschuh, A. Sontacchi, and F. Zotter, "HRTF modelling in due consideration variable torso reflections," in *Proceedings of the Acoustics'08*, 2008, pp. 99–104.

[22] G. W. Elko, E. Diethorn, and T. Gänsler, "Room impulse response variation due to thermal fluctuation and its impact on acoustic echo cancellation," in *International Workshop on Acoustic Echo and Noise Control (IWAENC2003)*, 2003.

[23] A. Andreopoulou and A. Roginska, "Towards the creation of a standardized HRTF repository," in *131th Convention of the Audio Engineering Society (AES)*, 2011, Convention Paper 8571.

[24] D. Schwarz and M. Wright, "Extensions and applications of the SDIF sound description interchange format," in *Proceedings of the International Computer Music Conference*, 2000.

[25] J. Merimaa, T. Peltonen, and T. Lokki, "Concert hall impulse responses - Pori, Finland," 2005. http://www.acoustics.hut.fi/projects/poririrs/.

[26] H. Ziegelwanger and P. Majdak, "Continuous-direction model of the time-of-arrival in the head-related transfer functions," *J Acoust Soc Am*, submitted.

[27] M. Noisternig, F. Zotter, and B. F. Katz, "Reconstructing sound source directivity in virtual acoustic environments," in *Principles and Applications of Spatial Hearing*, Y. Suzuki, D. S. Brungart, and H. Kato, Eds. Singapore: World Scientific Publishing, 2011, pp. 357–373.