



Audio Engineering Society

Convention Paper 9764

Presented at the 142nd Convention
2017 May 20–23 Berlin, Germany

This Convention paper was selected based on a submitted abstract and 750-word précis 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 The complete manuscript was not peer reviewed.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. This paper is available in the AES E-Library, <http://www.aes.org/e-lib>. 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.

MySofa: Design Your Personal HRTF

Christian Hoene¹, Isabel C. Patiño Mejía², and Alexandru Cacerovschi³

¹*Symonics GmbH, Tübingen, Germany*

²*Universität Tübingen, Tübingen, Germany*

³*DAS Solutions SRL, Chişinău, Moldova*

Correspondence should be addressed to Christian Hoene (christian.hoene@symonics.com)

ABSTRACT

Binaural auralizations are present in increasing numbers of applications and devices. Albeit most of the time they only use generic Head-Related Transfer Functions (HRTFs), the recent standardization of the HRTF format SOFA has paved the way to support individualized HRTFs broadly. We have developed and implemented MySofa: a web service that helps users to design a personal HRTF. In MySofa, based on anthropometric and user inputs, algorithms calculate and tune HRTFs. The result is displayed in the web browser and the user can listen to test renderings to verify, whether the personalized HRTF matches his expectations. In order to foster the use of individualized HRTF, we also implemented a light weight C-library called libmysofa, which helps programmers to read SOFA files and lookup FIR filters.

1 Introduction

Already in 2000, Begault has quantified the benefits of individualized Head-Related Transfer Functions (HRTFs) and compared them to generic ones. With individualized HRTFs, the front-end confusion is reduced and the elevation can be judged more precisely [1].

Roughly estimated, only a few dozen laboratories in the world are able to measure individual HRTFs. In addition, albeit recent improvements [2] measuring an individual HRTF is cumbersome – even if you do not take in account the travelling that is required to get to the lab. Thus, in the recent years, numerous approaches have been developed to get a personalized HRTF more easily. For example, Nicol et al. suggest that a good HRTF can be achieved by selecting one HRTF out of seven [3]. Zotkin has described how to

personalize HRTF by tuning anthropometric properties [4], and Hwang personalizes HRTFs by changing 12 principal components (PCs), which he found in the CIPIC data set [5]. Last but not least, a HRTF can be calculated numerically by means of a Bounded Element Method (BEM) [6].

All these approaches show different trade-offs in respect to accuracy and effort. For example, if the user's head and torso can be scanned with an MRI at high resolution, a BEM simulation is a good choice. As another option, the head shape and dimensions can be measured visually by means of photometry or 3D depth scans. The resolution of such scans is typically lower and shapes such as the ear channel, which is not viewable from the outside, cannot be gathered. Then, anthropometric heuristics could lead to results of similar quality as compared to a BEM simulation that is

made from incomplete and noisy input data. In addition, heuristics approaches based on anthropometric features have the advantage that relevant distances can be simply measured with a calliper. Finally, if no input data is available, the user still can select his HRTF by means of listening tests or tune them – e.g. – by modifying parameterized HRTFs.

We believe that not only one algorithm shall be used to personalize HRTFs, but it shall be possible to choose one out from several algorithms – depending on the user’s requirements. Some use cases – such as guidance systems for soldiers are mission critical and an imprecise HRTF might even have lethal consequences. Thus, it is important to have a perfect HRTF. An audiophile might spend considerable time to tune his HRTF whereas an occasional user of binaural rendering solutions is happy to get his HRTF fast, even if it is only moderately well fitted.

Instead of providing one algorithm that fits all needs or instead of aiming for the perfect personalized HRTF, the web service MySofa lets the user decide which quality he needs. The HRTF can be optimized until the user is satisfied.

MySofa implements the following work flow: First, the user selects one of a few HRTF personalization algorithms. Then, he uploads or enters his anthropometric data. Afterward, the HRTF is calculated and the result is displayed. The user can listen to the results to verify whether it meets his expectations. Finally, the HRTF can be downloaded and – as it is stored in the standardized SOFA format [7] – it can be uploaded into a device or software for its use.

In the remaining paper, we describe how we have set up the MySofa web service, which components have been selected and combined, and which further enhancements are envisioned.

2 The AES SOFA format

Prior to the standardization of the AES SOFA format, researchers used various formats to store and exchange HRTFs. Thus, Madjak et al. defined a standard [8], which generalized the existing conventions on how to store a FIR filter and defined a common format on how to describe those HRTF data sets. As such, the AES SOFA format is well suited for researcher to exchange their lab data. Conveniently, it uses a well known HDF5 format, which is also used in MATLAB.

Whereas the SOFA format is well suited for researchers, using its reference implementation for spatial audio rendering software is not as straight forward because of several limitations.

First of all, the storage format requires the use of third-party libraries such as libhdf5 and libnetcdf, which increase the code size and which are not well suited for cross platform compilation.

Second, despite SOFA being a common standard, it still allows for various notations on how to describe HRTFs. For example, coordinates may be described in Cartesian or spherical conventions and unit names may be written in British or American English.

Third, the SOFA format is not optimized for a fast lookup of HRTFs filter. Instead, HRTFs are stored in an unsorted manner.

Last not least, the loudness of HRTF is not standardized. Thus, if you switch from one SOFA file to the other, loudness changes may occur.

To overcome these issues, we have implemented a library that can read SOFA files. It is called libmysofa and is available under Apache 2 license at github [9]. While focusing just on the reading functionality, we have been able to reduce the code size. For example, the program size of a SOFA to JSON converter shrank from 798387 bytes to 90625 on a 64bit Linux. We based the implementation on the specifications of the HDF, NETCDF and SOFA file formats since SOFA is based on NETCDF and NETCDF is based on HDF. Where the HDF5 specification was incomplete [10], we reverse engineered SOFA files to figure out implementation and specification details.

A FIR filter, stored in the SOFA format, cannot always be used for real time auralizations. Thus, further processing steps are useful.

First, the FIR filters are resampled to match a requested sampling rate of the auralization. In libmysofa, we take advantage of the speex resampler [11].

Next, the loudness of the HRTF data set is normalized. We look for the FIR filter pair which is closest to the looking direction of the listener. Then, we determine the mean loudness of both left and right channel and normalized all filters so that the mean loudness of the left and right frontal looking FIR filters is one.

In a next step, we lower - the size of the FIR filters splitting them into min-phase filters and delays lines.

To achieve this, we reduce the length of each FIR filters until it carries 99% of the energy of the original filter. In the set of length reduced FIR filters, we take the longest and define its length as the length of all remaining filters. After this step, all FIR filters have the same length.

Because in a SOFA file the FIR filters are not sorted, looking them up is not straight forward. Thus, we use a K-D tree [12] to find a filter pair for a given orientation. On average, K-D trees have a complexity of $O(\log n)$ and a worst case complexity $O(n)$. We use Cartesian coordinates to store the FIR filters in the K-D tree because spherical coordinate would require a special wrap-round treatment.

Finally, to achieve an interpolation for missing values, we first lookup for the neighbors of each filter coordinate. The search for neighbors is done in spherical coordinates. These neighbors are stored in a table for each FIR filter pair. If a FIR filter pair for a given coordinate has to be interpolated, first we use the K-D tree to find the closest FIR filter pair. In a next step, we lookup suitable and nearby neighbors. Then, the interpolation of the filters is done linear, weighted by the reciprocal of the distance between the requested coordinate and the given coordinates of the filter pairs.

Following the advice of [13] and to make programming more easy and efficient, we provide access to SOFA files with a reduced set of function calls. The programmers can open a SOFA file, request filter pairs for a given coordinate, and finally, they need to free libmysofa memory structures again.

With the help of libmysofa, to add support for SOFA files and thus individualized HRTFs into auralization engines is straight forward. At least we had not any difficulties adding libmysofa to our spatial audio web conference service [14]. Time will show on how many other 3D audio solutions will support the SOFA format, too.

3 MySofa web service

MySofa is a web service, which can be used with modern web browser on devices such as smartphones, tables, and computers. Therefore, it is accessible to anybody in the world who has internet connectivity. MySofa helps the user to select, generate and tune his personal HRTF.

In the frontend, the part which runs in the browser, MySofa is based on the frameworks Bootstrap [15] and AngularJS [16] as those allow efficient programming. In the backend, we use an Apache server for delivering static content, a Tomcat server and the Spring Boot framework [17]. In addition, the various algorithms described in Section 4 to 6 are also running on the backend server.

4 Anthropometric input

MySofa supports different ways of providing anthropometric input. We assume that the user has access to 3D scanning solutions such as MRI, 2D to 3D photometry, or scanning with depth cameras.

For example, we tested MySofa with an RGBD sensor Asus Xtion Pro Live and an algorithm, which one of the authors has developed [18]. With this approach the methodology for the data acquisition is as follows: the sensor is located at the same height as the user's eyes. The user starts facing the sensor, then he turns until complete a 360° rotation with 10 seconds pause at 90° and 270°, this with the purpose of improving the ear data. In order to capture side view models, the user stands still in front of the camera showing his right and left side for approximately 10 seconds each. After finishing the scanning, the user uploads a 3D model of the head and torso. Meshlab [19] helps us to convert various input formats. For faster uploads via internet, we also allow those files being compressed. ZIP files containing DICOM MRI images are converted to a mesh automatically by a tool called DICOM2STL [20].

To verify the presence of a 3D face, we run an algorithm by Spreeuwers [21][22] to detect the nose and orientation of the face.

Detection of acoustically relevant measurements as head's width, height, and depth, as well as the closest distance from the front and the back of the head to the ears is performed by the algorithm described in [18]. The measurements are estimated based on the location of some crucial feature points (see Fig. 1): points located 3cm in front of the ear pits are used to calculate the width of the head, the top of the head and the chin estimate the height, the depth is obtained using the back of the head and the forehead, the closest point from the front and from the back of the head to the ears are represented by the nose bridge and the nape respectively. Since the user's rotation direction

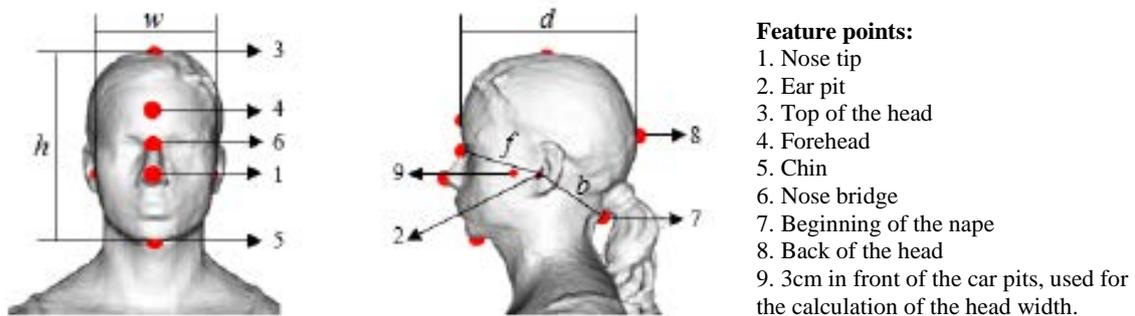


Figure 1: Feature points and measurements. w is head's width, h represents head's height, d symbolizes the depth, f and b the closest distance to the ears from front and back of the head respectively [18].

and undesired displacements affect the quality of the 360° model, the ear pits are located first on the side models through a concavity analysis. Then, these pit's positions are used to estimate the ear pits on the 360° model. The forehead, the top and the back of the head are determined based on the geometry of the human head. The chin is identified as the point with the largest concave value on the chin region. The nose bridge and the nape are determined using the Pythagorean theorem.

If the user does not have the capabilities to scan his head in 3D, the anthropometric measurements can be manually measured and added or the default values can be used.

5 HRTF personalization

One key idea of MySofa is to provide the user different means of getting his personal HRTF. Depending on the effort the user wants to spend, he can choose from a variety of algorithms.

Another aspect is to defer the responsibility of getting a good HRTF to the user. In the end, only he can judge whether the HRTF is good enough to meet his expectations. MySofa provides just the tools to make the personalization process easier.

One of the tools is a selection of one of few HRTFs. This approach is based on the observation described in [23] that only seven perceptual different HRTFs may be sufficient to satisfy most of the listeners. For this algorithm, anthropometric data is not required, just a good headset and a browser.

In the current MySofa prototype, we have included a heuristic algorithm, which is based on the MRI scan and HRTF measurements of 47 persons. Bomhardt

and Fels [24] analyzed the data, found its principle components and made a regression analysis to map anthropometric features to PCs. Based on this algorithm, individualized HRTFs are calculated.

Another algorithm, which we are currently adding, is implemented by Hölzl [25]. It allows the user to tune principle components in order to get his HRTF. Also, the BEM simulator of Ziegelwanger [26] is scheduled for addition.

6 HRTF quality control

After calculating an HRTF, the user verifies whether the results are good enough. Due to faulty input data, usage mistakes, and algorithms that operate beyond their tested scope, a HRTF may be degenerated. Thus, the user listens to binaural test renderings with the aim of checking, whether the sound is heard as expected.

Various test sounds, both stationary and in motion, can be played to get an impression about the quality of the externalization, to verify front vs. back, and to locate sounds moving along the cone of confusions.

In order to provide acoustic samples, we have taken a media server called Freeswitch [27] and extended our 3D audio rendering engine libspace, which is used in [27], too. With this renderer, different test samples are generated and transmitted to the browser via WebRTC. Finally, as a gimmick, the HRTF is displayed in 3D via the library three.js [28].

7 Summary and Outlook

Libmysofa allows every programmer of auralizations to support individualized HRTF easily. We hope that with the help of libmysofa the individualized HRTF

using the SOFA format will become widely available, preferable on any device that supports spatial audio. For MySofa web service, numerous algorithms and software frameworks have been compiled to get a web service that can be used by everybody. To the best of our knowledge, MySofa is the first web service of its kind.

At the moment, MySofa only contains a basic set of features, but time and usage experience will allow us to enhance MySofa with more algorithms and testing procedures. For this optimization process, user feedback is essential. In addition, further enhancements are envisioned by using crowdsourcing approaches [29] on existing algorithms. For example, a contest between multiple HRTF individualization approaches could be done in order to see which is preferred.

Acknowledgments

This work has been funded by German Federal Ministry of Education and Research, funding code 01IS14027A. We would like to thank Ramona Bomhardt, Piotr Majdak, Chris Pike, and the entire HDF group for their help and support.

References

- [1] D. R. Begault, E. M. Wenzel, and R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source" *Journal of the Audio Engineering Society* 49.10 (2001).
- [2] J.-G. Richter, G. Behler, J. Fels, "Evaluation of a Fast HRTF Measurement System," *Audio Engineering Society Convention 140*, 2016A.
- [3] R. Nicol, M. Emerit, E. Roncière, and H. Déjardin, "How to make immersive audio available for massmarket listening," *EBU Technical Review*, 14 July 2016.
- [4] D. Y. N. Zotkin, et al. "HRTF personalization using anthropometric measurements." *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2003.
- [5] S. Hwang, Y. Park and Y. Park, "Modeling and customization of head-related transfer functions using principal component analysis" *2008 International Conference on Control, Automation and Systems*, Seoul, 2008, pp. 227-231.
- [6] H. Ziegelwanger, P. Majdak, W. Kreuzer, „Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization" *The Journal of the Acoustical Society of America*, 2015, 138 , 208-222
- [7] Majdak, Piotr, et al. "Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions." *Audio Engineering Society Convention 134*, 2013.
- [8] Audio Engineering Society, "AES Standard for file exchange – Spatial acoustic data file format," AES69-2015, March 2015.
- [9] <https://github.com/hoene/libmysofa>, 2017.
- [10] The HDF Group, "HDF5 File Format Specification Version 3.0", <https://support.hdfgroup.org/HDF5/doc/H5.format.html>, April 2016
- [11] Jean-Marc Valin, "The Speex Codec Manual Version 1.2 Beta 3" December 8, 2007
- [12] J. L. Bentley, "Multidimensional binary search trees used for associative searching" *Communications of the ACM*. 18 (9): 509.
- [13] J. Bloch "How to design a good API and why it matters" *Companion to the 21st ACM SIGPLAN symposium on Object-oriented programming systems, languages, and applications*. ACM, 2006.
- [14] Symonics Meeting web service <https://meeting.symonics.com>
- [15] Bootstrap. <https://getbootstrap.com/>
- [16] AngularJS. <https://angularjs.org/>
- [17] Spring Framework. <https://spring.io/>
- [18] I. C. Patiño Mejía and A. Zell, "Head Measurements from 3D Point Clouds" *The Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA2016)*, Oulu, Finland, Dec. 2016
- [19] Meshlab. <http://www.meshlab.net/>
- [20] David T. Chen, "DICOM2STL", <https://github.com/dave3d/dicom2stl>
- [21] L.J. Spreeuwens "Breaking the 99% barrier: optimisation of 3D face recognition" *IET Biometrics*, 4 (3). pp. 169-177. ISSN 2047-4938, 2015

- [22] L.J. Spreeuwens "Fast and Accurate 3D Face Recognition Using Registration to an Intrinsic Coordinate System and Fusion of Multiple Region classifiers" *International Journal of Computer Vision*, 93 (3). pp. 389-414. ISSN 0920-5691
- [23] B. Katz, P. Gaëtan "Perceptually based head-related transfer function database optimization." *The Journal of the Acoustical Society of America* 131.2 (2012): EL99-EL105.
- [24] R. Bomhardt and J. Fels, "Individualization of head-related transfer functions by the principle component analysis based on anthropometric measurements", *J. Acoust. Soc. Am.* 140, 3277 (2016).
- [25] J. Hölzl "A Global Model for HRTF Individualization by Adjustment of Principal Component Weights", Diploma Thesis, *Graz University of Technology*, March 2014.
- [26] H. Ziegelwanger, W. Kreuzer, and P. Majdak, P "Mesh2HRTF: Open-source software package for the numerical calculation of head-related transfer functions" in *Proceedings of the 22nd International Congress on Sound and Vibration*, Florence, IT, 2015
- [27] FreeSWITCH. <https://www.freeswitch.org/>
- [28] Three.js. <https://threejs.org/>.
- [29] Doan, Anhai, Raghu Ramakrishnan, and Alon Y. Halevy. "Crowdsourcing systems on the world-wide web." *Communications of the ACM* 54.4 (2011): 86-96.