The Viking HRTF Dataset v2

Simone Spagnol, Riccardo Miccini, Rúnar Unnthórsson

October 30, 2020

1 Introduction

The *Viking HRTF dataset v2* is a collection of head-related transfer functions (HRTFs) which includes full-sphere HRTFs measured on a dense spatial grid (1513 positions) with a KEMAR mannequin with 20 different pairs of artificial pinnae attached, one pair at a time. The artificial pinnae were obtained through a custom molding procedure from 20 different lifelike human heads.[1] The purpose of the present collection, which is available for free download, is to provide accurate input data for investigations on the relation between HRTFs and anthropometric data through machine learning techniques or other state-of-the-art methodologies [1, 2].

A first version of the dataset was released in May 2019 [3].[2] In this second version, the used artificial pinnae were re-casted from the existing inverse molds with 35 Shore-OO silicone for both the left and right channels of the KEMAR. Furthermore, the new HRTF measurements have been taken inside the anechoic chamber of the University of Iceland in Reykjavík and free-field compensated.

2 Measurement system

The HRTF measurements were taken automatically with the system pictured in Figure 1. The measurement system consisted of a KEMAR mannequin[3] mounted on a 360° rotating cylindrical stand and a Genelec 8020CPM-6 loudspeaker mounted on an L-shaped rotating arm. The configuration for the mannequin was the 45BB-4 with standard large anthropometric pinnae (35 Shore-OO hardness, GRAS KB5000/01) as reference and half-inch pressure microphones (GRAS 40AO) placed at the ear canal entrances. The two audio channels were fed to an RME Fireface 802 sound card connected to a host workstation running MATLAB.

The distance between the loudspeaker tweeter and center of the mannequin's head was a constant 1 m, independently of the orientation of the mannequin and arm. This distance value guarantees the collection of far-field spectral cues with reasonable accuracy [4]. Absolute references for azimuth and elevation were provided through a fixed marker on the bottom of the mannequin (0° azimuth) and a bubble level mounted on the side segment of the arm (0° elevation), respectively. Correct alignment in all three axes could be attained by targeting laser beams to the microphones and tip of the nose of the KEMAR. This experimental choice allowed HRTF measurements on a full-azimuth range with elevations from $-45°$ to $90°$, according to a vertical polar coordinate system.

[1] Courtesy of Ernst Backman, Saga Museum Reykjavík.

[2] https://itsadive.create.aau.dk/index.php/viking-hrtf/

[3] http://kemar.us

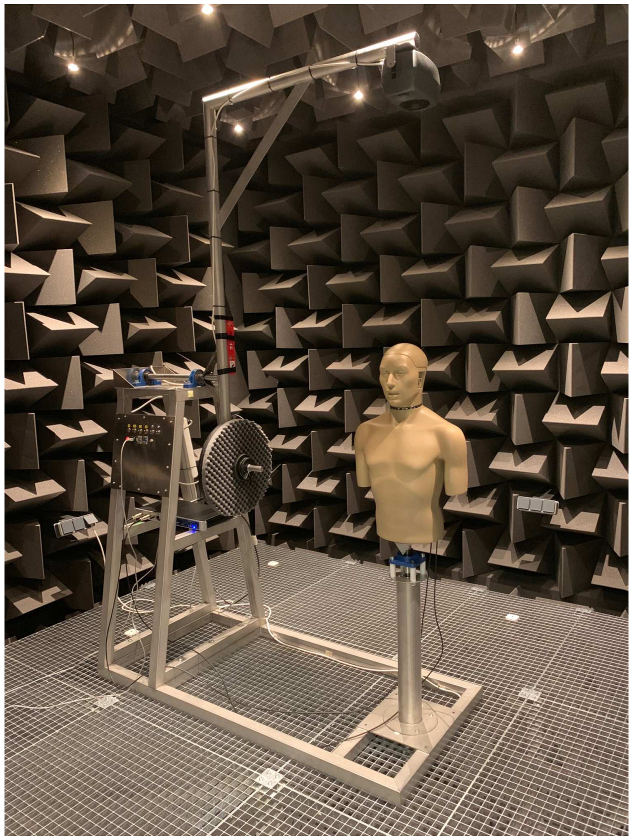

Figure 1: The HRTF measurement system: mechanical apparatus, loudspeaker, and KEMAR mannequin.

Rotation of the mannequin and arm was managed by two independent high-torque step motors (JVL MST001A, 1.2 Nm), controlled through two digital step drives (Geckodrive G213V) in full-step mode (200 steps per revolution). In order to increase the torque and angular resolution of the system, two 100 : 1 gearboxes were installed between the first motor and the arm and between the second motor and the mannequin. As a result, the minimum rotation angle of both the arm and mannequin was 0.018°. Furthermore, in order to reduce the needed torque on the arm, a 22 kg counterweight was applied to the shorter appendix so to balance the weight of the loudspeaker and the longer appendix of the arm. Communication between the step drives and host workstation was also managed in MATLAB.

3 Pinna samples

In order to provide a reasonable sample of pinna shapes for the HRTF measurements, we set up and applied a custom procedure to cast silicone replicas of pinnae from dummy heads to fit the KEMAR as follows.

Step 1: First negative mold. We applied two layers of silicone mix to the ear, first a thin one with a paintbrush and then a thick one (1/2 cup) with a spatula, letting each layer dry for 1 day before the next step. In order to reduce leakage onto vertical surfaces, we added a few drops of thixotropic additive to the silicone for the thick layer. Then, we applied wet plaster of Paris strips over the negative silicone mold (Figure 2a) in order to efficiently peel it from the dummy head and

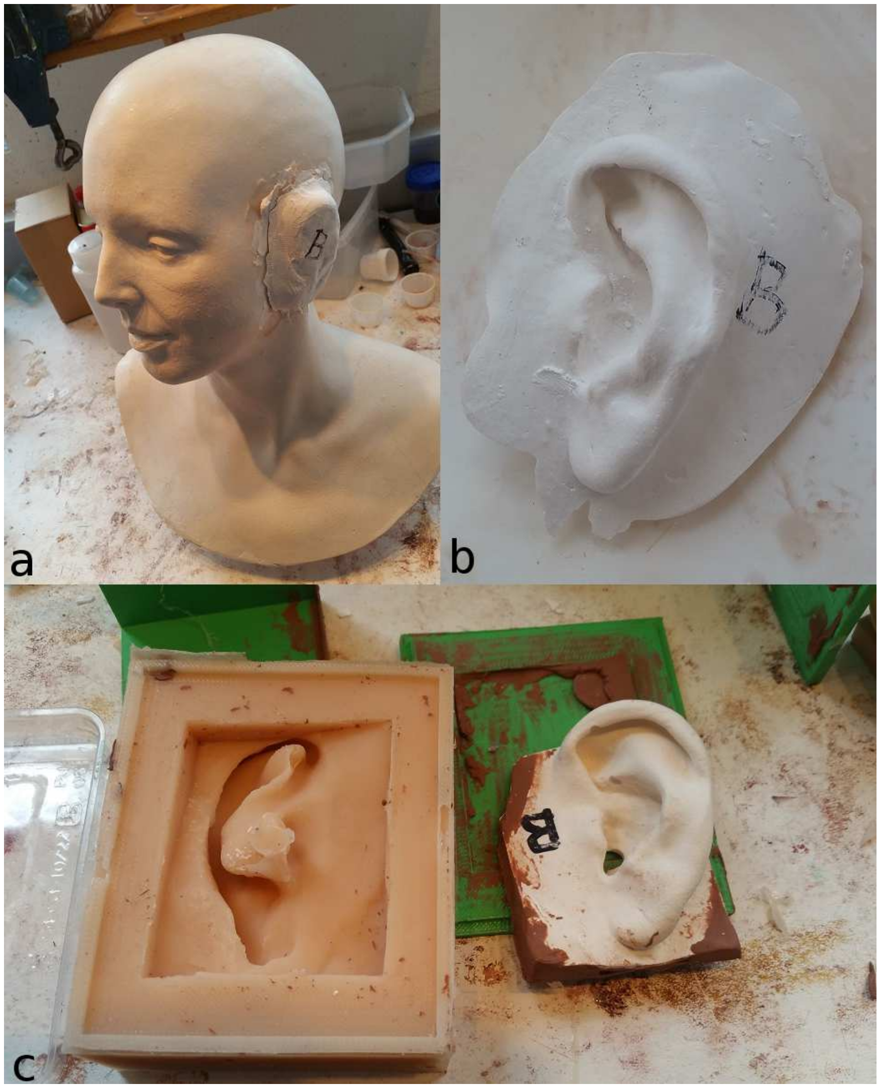

Figure 2: The making of a negative ear mold: (a) 1st negative mold; (b) Jesmonite®replica; (c) 2nd negative mold.

to give it support so it does not deform in Step 2.

Step 2: Jesmonite® replica. We poured a quick mix of 1/2 cup of water and two teaspoons of Jesmonite®into the negative mold obtained in Step 1 placed above a vibration table (in order to avoid formation of air bubbles), and let it harden for $1 - 2$ days. Then, we removed the Jesmonite®ear (Figure 2b) and occasionally filled gaps or deformations in it with clay. Finally, we drilled a hole in the back of the concha and sanded the ear base with a belt sander to accurately fit the size of the KEMAR pinna slot.

Step 3: Second negative mold. We placed the Jesmonite®ear replica into a custom-made plastic box with open top and secured it to the bottom with clay. Then, we poured 2 cups of silicone mix inside the plastic box, placed on the vibration table, and let it dry for 1 day. We finally removed the plastic box and Jesmonite®ear (Figure 2c).

Step 4: Final pinna replica. We first poured car wax into the inner parts of the negative mold obtained in Step 3 as a release agent, and let it dry for half a day. Then we poured 1/3 cup of a

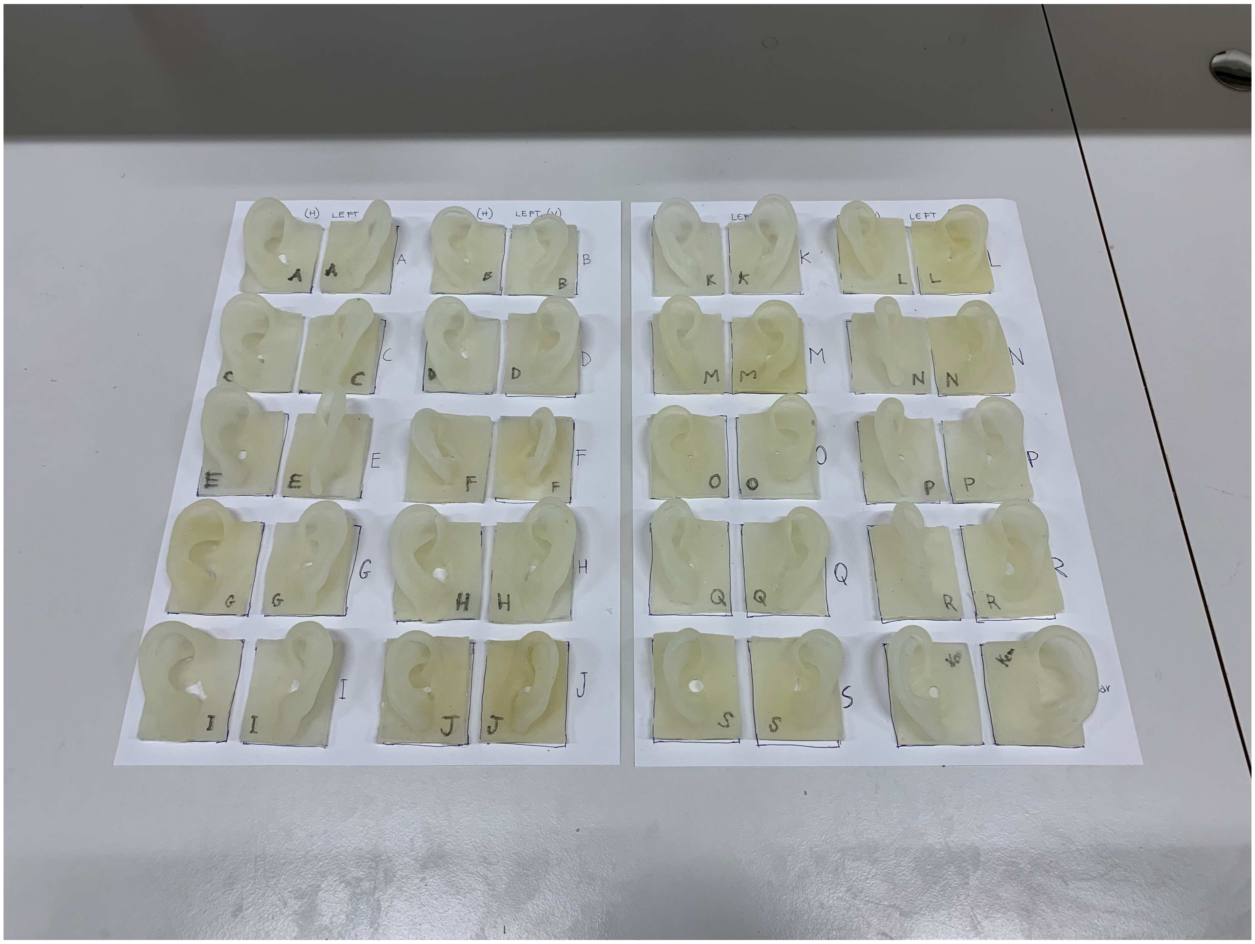

Figure 3: The 20 pairs of artificial pinnae obtained through the casting procedure.

different silicone mix from the one used in Steps 1 and 3 (35 Shore-OO hardness) into the negative mold, and let it dry for 1 day. We finally removed the silicone pinna replica and cut the excess parts on the base with a knife.

As pictured in Figure 3, we applied the procedure to both ears of 20 different artificial heads. These include the KEMAR with standard large anthropometric pinnae and 19 different lifelike dummy heads made out of plaster, borrowed from the Saga Museum in Reykjavík. The heads, manufactured between 2001 and 2003 for artistic purposes,[4] reproduce with high fidelity the anthropometric traits of 19 Icelandic humans (7 female), aged between 7 and 77 at the time of manufacturing.

4 Measurement procedure

The HRTF measurements were carried out during the month of November 2019 inside the anechoic chamber recently installed at the University of Iceland. The chamber has a size of $5.2 \times 4.3 \times 3.9$ m (LWH), and the mannequin was placed roughly in the center of it. Based on ISO 3745 and ISO 26101, the anechoic free field in the chamber has been certified (as of September 2019) compliant for measurements within an area equivalent to a distance of approximately 0.2 m from wedge tips

[4]`https://sagamuseum.is/sagadesign/`

Table 1: HRTF measurement positions.

Elevations [deg]	[-45,45]	[50,70]	[75,85]	90
Step [deg]	5	15	45	360
No. of azimuths	72	24	8	1

on the walls, ceiling and floor in the frequency range from 200 Hz to 10 kHz. In order to remotely monitor the data acquisition process and to prevent possible mechanical or electrical failures, video footage was streamed to the host workstation in real time via a webcam mounted inside the anechoic chamber and pointing towards the measurement system.

We used the logarithmic sweep method to record the single acoustical responses. The input sweep signal, whose level was kept constant throughout the whole measurement schedule, spanned frequencies between 20 Hz and 20 kHz in 0.9 s, at a sampling rate $f_s = 48$ kHz. The average SPL level at 1 kHz for a frontal stimulus as collected through a Class 1 sound level meter at the center of the reference system was 90 dB.

Sound source location was specified through the azimuth angle θ and elevation angle ϕ in vertical-polar coordinates. Elevations were uniformly sampled in 5° steps from $-45°$ to 90°, while azimuths were sampled in different increments as shown in Table 1 in order to obtain roughly uniform density towards the upper pole of the sphere. The total number of spatial positions per measurement was 1513. At each measurement session, and for each elevation, the mannequin was consecutively rotated of the corresponding angular step and sweep responses were acquired at each azimuth angle. After completion of all azimuth angles for the current elevation, the arm moved down to the next elevation angle. In order to maintain the HRTF measurements as silent as possible, 0.5-ms pauses were introduced between all motor commands and the previous/next recording. The duration of one single measurement session was approximately 80 minutes.

Twenty-one measurement sessions were scheduled in total. The first 20 sessions were devoted to the 20 pinna pairs, including the 19 casted from the plaster heads (alphabetically labeled as sets A to S) and the KEMAR pinna replicas (set T). In the final session (set Z) we measured the response of a pair of flat 35 Shore-OO silicone baffles filling the two pinna slots flush with the head (so as to simulate a "pinna-less" condition). Between each two measurement sessions, the HRTF measurement system was manually calibrated to the correct starting position by tuning each motor to its corresponding absolute reference.

On top of the above measurements, free-field reference measurements for the system were collected, allowing for removal of the influence of any element other than the mannequin. This was done by removing the KEMAR, replacing it with a thin wooden pole, and mounting a KEMAR microphone onto it, so that its position would be roughly at the center of the interaural axis with the head absent. These measurements were taken on the median plane only, at 5° elevation steps.

A post-processing script was written in MATLAB to recover head-related impulse responses (HRIRs) from sweep responses in accordance with the sweep method and the free-field compensation method. This means that each recorded signal was filtered with (1) the inverse reference spectrum of the input signal, low-passed and high-passed with second-order digital Butterworth filters to compensate for the original zero sound pressure level below 20 Hz and above 20 kHz, and (2) the inverse complex free-field response for the current elevation, where the phase spectrum of the free-field response was the minimum phase corresponding to the amplitude spectrum. Before step 2, a 128-sample half-Hann window was applied to each HRIR starting from its onset with the aim of removing the possible impact of reflections occurring far away from the mannequin.

5 Dataset contents and use

The HRTF dataset is provided in the standard SOFA format[5] and is available to download at [5]. The 21 SOFA files contain the post-processed HRIRs for each different pair of articial pinnae (files **subj_A.sofa** to **subj_T.sofa**), plus the "pinna-less" condition (**subj_Z.sofa**).

3D scans of the 20 left pinna replicas are also included in the dataset as STL files (files **subj_A.stl** to **subj_T.stl**). The scans were captured at 1 mm resolution with a Creaform Go!SCAN 20 white light scanner.[6]

The dataset is provided under the CC-BY 4.0 license[7] that grants unlimited access for everyone. If you use this data please reference

- Simone Spagnol, Kristján Bjarki Purkhús, Sverrir Karl Björnsson, and Rúnar Unnthórsson (2019). ***The Viking HRTF dataset***. In: *Proceedings of the 16th Sound & Music Computing Conference (SMC 2019)*, pages 55-60, Málaga, Spain.

- Simone Spagnol, Riccardo Miccini and Runar Unnthórsson (2020). ***The Viking HRTF dataset v2***. DOI: 10.5281/zenodo.4160401

Acknowledgment

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 797850, and from NordForsk's Nordic University Hubs programme under grant agreement No. 86892.

References

[1] R. Miccini and S. Spagnol, "HRTF individualization using deep learning," in *Proc. 2020 IEEE Conf. Virtual Reality and 3D User Interfaces Work. (VRW 2020)*, Athens, GA, USA, Mar. 2020, pp. 390–395.

[2] M. G. Onofrei, R. Miccini, R. Unnthórsson, S. Serafin, and S. Spagnol, "3D ear shape as an estimator of HRTF notch frequency," in *Proc. 17th Int. Conf. Sound and Music Computing (SMC 2020)*, Torino, Italy, Jun. 2020, pp. 131–137.

[3] S. Spagnol, K. B. Purkhús, S. K. Björnsson, and R. Unnthórsson, "The Viking HRTF dataset," in *Proc. 16th Int. Conf. Sound and Music Computing (SMC 2019)*, Malaga, Spain, May 2019, pp. 55–60.

[4] S. Spagnol, "On distance dependence of pinna spectral patterns in head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 137, no. 1, pp. EL58–EL64, January 2015.

[5] S. Spagnol, R. Miccini, and R. Unnthórsson, "The Viking HRTF dataset v2," Oct. 2020, DOI: 10.5281/zenodo.4160401.

[5] `https://www.sofaconventions.org/`

[6] Courtesy of AAU Create Prototyping Lab.

[7] `https://creativecommons.org/licenses/by/4.0/`