Romeo-HRTF: A Multimicrophone Head Related Transfer Functions Database

Mounira Maazaoui and Yves Grenier*

Institut TELECOM, TELECOM ParisTech, CNRS-LTCI 37/39, rue Dareau, 75014, Paris, France

Abstract

In this article, we propose a Head Related Transfer Functions (HRTF) database. We extend the usual concept of binaural HRTF to the context of the audition of a humanoid robot where the robot head is equipped with an array of microphones. A first version of the HRTF database is proposed, that is based upon a dummy that simulates the robot; we call it Theo. A second version of the database is based upon the head and torso of the prototype robot Romeo. We also propose the corresponding Head Related Impulse Responses (HRIR). The heads of Theo and Romeo are equipped with 16 microphones. The databases are recorded for 72 azimuth angles and 7 elevation angles. The typical use of this database is for the development of algorithms for robot audition.

Résumé

Dans cet article, nous proposons une base de données de fonctions de transfert de tête (HRTF: Head Related Transfer Functions). Nous étendons le concept habituel de HRTF binaurale au context de l'audition d'un robot humanoïde dont la tête est équipée par un réseau de microphones. Une première version de la base de données de HRTF enregistrée avec un mannequin qui simule le robot est proposée, nous l'appelons Theo. Une seconde version de la base de données est basée sur la tête et le torse d'un prototype du robot Romeo. Nous proposons aussi les réponses impulsionnelles de tête (HRIR: Head Related Impulse Responses) correspondantes. Les têtes de Theo et de Romeo sont équipées de 16 microphones. Les bases de données sont enregistrées pour 72 angles d'azimuth et 7 angles d'élévation. Un usage typique de cette base de données est le développement d'algorithmes pour l'audition des robots.

1 Introduction

The Head Related Transfer Function (HRTF) is a response that characterizes how a source signal emitted from a specific direction is received at an ear. The HRTF of each ear captures the localization information of the source signal and the alteration introduced by the head and the pinna on the sound manifold [3]. HRTF are important cues for perception of surrounding sounds and for localization of sources; they form the core of binaural spatialization techniques. In the framework of the Romeo project, our goal is to take advantage of these cues for robot audition, in order to help source localization and separation.

The objective of this project is to develop a humanoid robot that can act as a comprehensive assistant for persons suffering from loss of autonomy. Romeo has to come with a very "human-friendly" interface, voice and gestures being the principal means of communication with the robot. It will have to understand what is said to him, carry out simple talks and even feel the intentions and emotions of its interlocutor in order to deduce the actions it has to realize [1].

In this article, we generalize the HRTF to the case of a humanoid with more than 2 ears (more than 2 microphones fixed in his head) and we propose "Romeo-HRTF", a 504×16 HRTF database recorded by

*This work is funded by the Ile-de-France region, the General Directorate for Competitiveness, Industry and Services (DGCIS) and the City of Paris, as a part of the ROMEO project

16 microphones from 72 azimuth angles and 7 elevation angles. The microphones are fixed in the head of an infant size dummy (Theo) or the head of the humanoid robot (Romeo). The plan of this article is as follows: the second section presents an overview of the HRTF, the third section explains the method that we used to estimate those functions and the fourth section is a description of the proposed database and the measurement process.

2 Head Related Transfer Function

Localization of a sound source by humans is accomplished through two binaural cues: interaural time differences (ITD) and interaural intensity difference (IID) [3]. ITD is the difference in arrival times of a sound wavefront at the left and right ear. IID is the amplitude difference of a sound that reaches the right and left ear. It is known that ITD and IID are important for sound perception in the horizontal plan. However, if the sound is allowed to vary in elevation and distance, ITD and IID do not specify a unique spatial position, there are identical for sound sources placed in a cone named the cone of confusion (*cf.* figure 1). The location of a sound source placed on this cone is not distinguishable if we use the ITD and IID parameters.

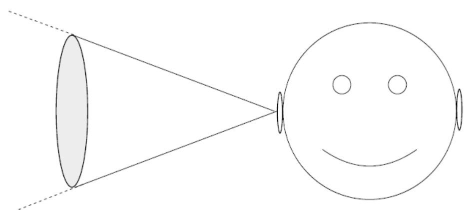

Figure 1: Cone of confusion

However, the human hearing system is still able to perceive the direction of this source. This is due to the spectral filtering of the sound source by the head and the pinna, and thus a transfer function between the source and each ear is defined: the Head Related Transfer Function (HRTF). The HRTF takes into account the ITD, the IID and the shape of the head and the pinna. The impulse response that corresponds to the HRTF is called Head Related Impulse Response (HRIR). Figure 2 represent an example of HRIR of Theo at elevation 0° and azimuth 0°.

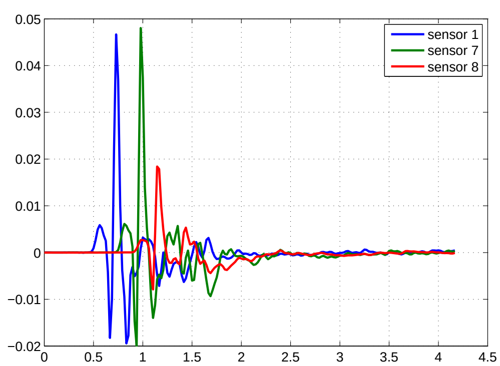

Figure 2: Examples of Theo HRIR at elevation 0° and azimuth 0° at sensors 1, 7 and 8

HRTF defines how a sound emitted from a specific location and altered by the head and the pinna is received at an ear. We consider a mono sound signal registered in anechoïc condition, non-localized. This sound can be artificially spatialized by taking its convolution with the specific left ear and right ear HRTF. The sound is then perceived as if it was coming from a specific location, the location from where the HRTF are measured: this technique is called binaural spatialization.

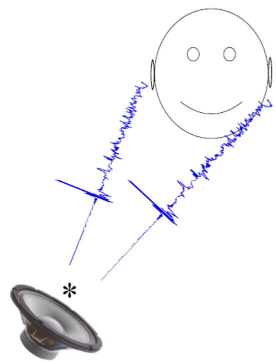

Figure 3: Binaural spatialization

The notions of HRIR and HRTF remain the same if we replace the human head by a dummy head and the ears by 2 microphones. We still keep the same notion if we increase the number of microphones to more than 2. Having more than 2 microphones fixed in the head allows to capture more accurately the effect of the head on the sound field and localize sounds more precisely. For example, with more than 2 microphones, the front/back indeterminacy that we have in the case of 2 microphone is avoided. We measured HRTF from 504 azimuth and elevation points. In this article, we measure the HRTF from different azimuth and elevation when a 16 microphone array is installed in the head of the dummy, Theo, and the prototype of the robot, Romeo. We see in figure 6 the localization of some of the microphone loosely fixed on the head of Theo. In the case of Romeo (figure 7), we see that 14 microphones are located at the surface of the skull, whereas two of them are located at a position that mimics the human pinna and auditory channel.

As far as we know, this is the first instance of a multimicrophone HRTF database with different azimuth and elevation angles; up to now, the HRTF databases that are available to download are binaural [8, 2, 4, 5]. The KEMAR is a binaural HRTF database measured with a dummy for 70 different positions [8] and the CIPIC HRTF database was measured with 45 subjects (humans) for 25 azimuths and 50 elevations [2].

3 Measurement of HRIR using Golay complementary sequences

We consider a discrete-time ($t \in \mathbb{Z}$) system characterized by an impulse response $h(t)$, an input signal $s(t)$ and an output signal $x(t)$. To identify the system, we need to estimate $h(t)$ for a known input signal $s(t)$ and output signal $x(t)$. In our case, the system represent the acoustic path and the influence of the head and we need to find the HRIR $h(t)$:

$$x(t) = s(t) * h(t) \tag{1}$$

where $*$ is the convolution operator.

To estimate this impulse response, we use Golay complementary sequences [7] as input signal $s(t)$. The Golay complementary sequences have the useful property that their autocorrelation functions have complementary sidelobes: the sum of the autocorrelation sequences is exactly zero everywhere except at the origin. Golay complementary sequences are not unique [10]. We use a pair of sequences $a(t)$ and $b(t)$ of length L is defined as follows:

$$
\begin{array}{l}
a\left(t\right) = \pm 1 \text{ for } 1 \leq t \leq L \\
b\left(t\right) = \pm 1 \text{ for } 1 \leq t \leq L
\end{array}
\tag{2}
$$

Those sequences are Golay complementary sequences if and only if:

$$
a\left(-t\right) * a\left(t\right) + b\left(-t\right) * b\left(t\right) = 2L\delta\left(t\right) \tag{3}
$$

The sequences that we used are defined recursively as:

$$
\begin{bmatrix} A_L \\ B_L \end{bmatrix} = \begin{bmatrix} A_{L/2} & B_{L/2} \\ A_{L/2} & -B_{L/2} \end{bmatrix} \text{ with } \begin{array}{l} A_L = [a(1)\cdots a(L)] \\ B_L = [b(1)\cdots b(L)] \end{array}
$$

$$
\text{and } \begin{bmatrix} A_2 \\ B_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}
$$

The responses of the system due to the inputs $a\left(t\right)$ and $b\left(t\right)$ are respectively:

$$
\begin{array}{l}
x_a\left(t\right) = a\left(t\right) * h\left(t\right) \\
x_b\left(t\right) = b\left(t\right) * h\left(t\right)
\end{array}
\tag{4}
$$

Using equations (3) and (4), the impulse response of the system is given by:

$$
h\left(t\right) = \frac{1}{2L}\left(a\left(-t\right) * x_a\left(t\right) + b\left(-t\right) * x_b\left(t\right)\right) \tag{5}
$$

4 Database description

4.1 Materials and software description

The Golay codes recording was done in the anechoïc room of Telecom ParisTech. The following equipment was used:

- Theo, an infant size dummy of $1m40$ height (*cf.* figure 5) for the acquisition of Theo_HRTF;
- head and torso of a Romeo, for the acquisition of Romeo_HRTF;
- 2 Echo Audiofire Pre8 audio recording devices[1] which support also playback (*cf.* figure 4);
- AKG C417 pp[2] 16 microphones (*cf.* figure 6);
- 7 Tannoy System 600 loudspeakers[3] ;
- Brüel & Kjær Type 9640 turntable[4].

Figure 4: Echo AudioFire Pre8 recorder

[1]http://www.echoaudio.com/Products/FireWire/AudioFirePre8/index.php
[2]http://www.akg.com/mediendatenbank2/psfile/datei/35/c4174055c447d8838.pdf
[3]http://www.tannoy.com/products/158/uman_System600.pdf
[4]http://www.bksv.com/products/telecomaudiosolutions/electroacousticsaccessories/turntablesystemtype9640.aspx

The Audiofire Pre8 are linked by FireWire to a PC with a real time core Linux exploitation system. The software used are:

- Jack audio connection kit (JACK-control) [9] which is an open-source application to control the JACK sound server daemon specific for the Linux Audio Desktop infrastructure.
- Ffado [6], an open-source driver for FireWire based pro-audio devices. We use ffado-mixer to control the synchronization between the recorders and access to the channels mixing table of the recorder.

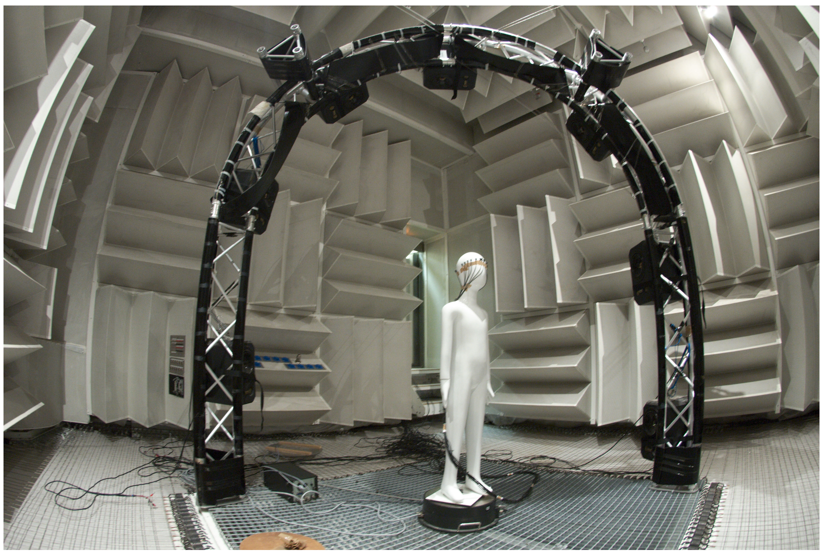

Figure 5: The dummy used to measure the HRTF (in the anechoïc room)

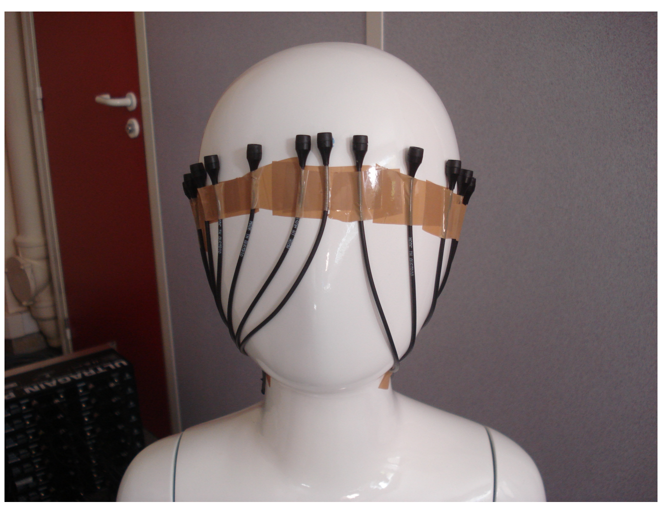

Figure 6: The 16 microphone array of Theo

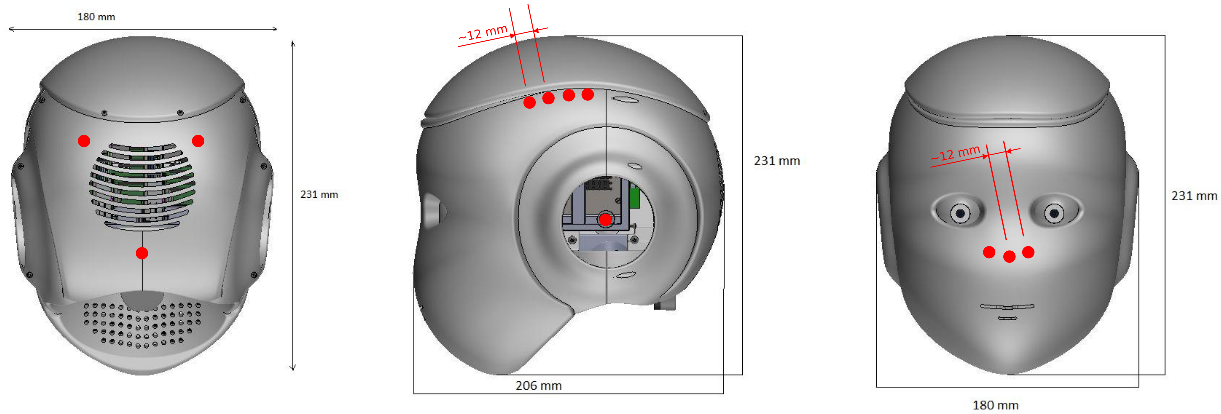

Figure 7: Romeo head and the localization of the microphones (in red)

4.2 Experimental process

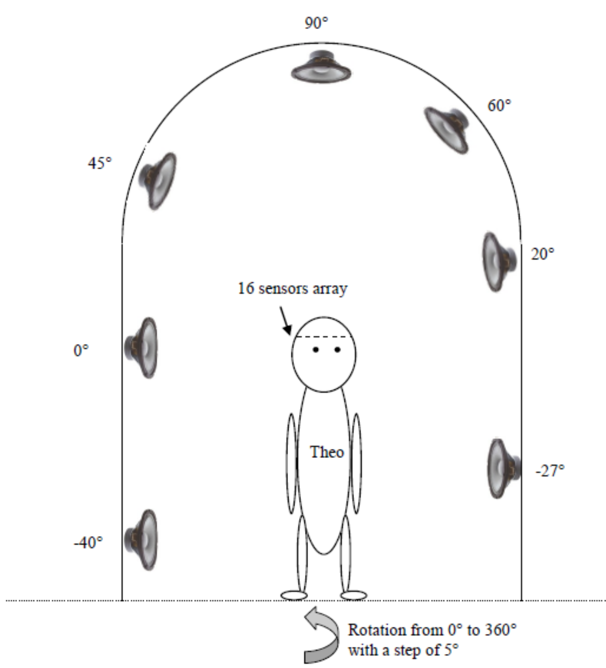

Figure 8: Theo and loudspeakers position for the Golay codes recording

In the anechoïc room, we measured the HRTF from 504 azimuth/elevation points distributed as follow (*cf.* figure 8):

- 73 azimuth angles from 0° to 355° with a 5° step

- 7 elevation angles: -40°, -27°, 0°, 20°, 45°, 60° and 90°

Theo (or Romeo) is fixed on the turntable in the center of the loudspeakers arc. The sequence of complementary Golay codes (*cf.* section 3) is emitted sequentially from each loudspeaker (elevation) and recorded with the 16 sensors array for each azimuth. The sampling frequency of the recorders is 48 kHz.

5 How to get Romeo-HRTF?

Romeo-HRTF database is under Open Database License[5] and it can be downloaded from:
`http://www.tsi.telecom-paristech.fr/aao/?p=347`

6 How to cite Romeo-HRTF?

The use of Romeo-HRTF should be reported by citing this article.

7 Conclusion

We propose a multi-microphone HRTF database Romeo-HRTF recorded with 16 microphone fixed on the head of an infant size dummy Theo and the head of the humanoid robot Romeo. We recorded 72 azimuth angles from 0° to 355° with a step of 5° and 7 elevation angles. We give a description of the experimental process including the measurement of the HRIR with Golay complementary sequences and the hardware and software used for the HRIR acquisition.

References

[1] Romeo project: www.projetromeo.com.

[2] V.R. Algazi, R.O. Duda, D.M. Thompson, and C. Avendano. The cipic hrtf database. In *2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pages 99 –102, 2001.

[3] Jens Blauert. *Spatial hearing, the psychophysics of human sound localization.* MIT Press, 1983.

[4] Ircam AKG HRTF database. http://recherche.ircam.fr/equipes/salles/listen/index.html, 2003.

[5] Shimada Laboratory HRTF Database. http://audio.nagaokaut.ac.jp/hrtf/, 2005.

[6] FFADO. http://www.ffado.org/?q=release/beta.

[7] S. Foster. Impulse response measurement using golay codes. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '86*, volume 11, pages 929 – 932, April 1986.

[8] Bill Gardner and Keith Martin. Hrtf measurements of a kemar dummy-head micophone. Technical report, MIT Media Lab Perceptual Computing, 1994.

[9] JACK Audio Connection Kit. http://qjackctl.sourceforge.net/.

[10] Matthew G Parker, Kenneth G Paterson, and Chintha Tellambura. Golay complementary sequences. In *Encyclopedia of telecommunications*, 2004.

[5] `http://www.opendatacommons.org/licenses/odbl/`