Audio Engineering Society

Convention Paper

Presented at the 120th Convention
2006 May 20–23 Paris, France

A System for Rapid Measurement and Direct Customization of Head Related Impulse Responses

Simone Fontana[1], Angelo Farina[2] and Yves Grenier[3]

[1] Ecole Nationale Supérieure des Télécommunications, Paris, France
fontana@tsi.enst.fr

[2] Università di Parma, Parma, Italia
farina@pcfarina.eng.unipr.it

[3] Ecole Nationale Supérieure des Télécommunications, Paris, France
grenier@tsi.enst.fr

ABSTRACT

Head-Related Impulse Responses (HRIRs) measurement systems are quite complex and present long acquisition times for an accurate sampling of the full 3D space. Therefore HRIRs customization has become an important research topic. In HRIRs customization some parameters (generally anthropometric measurements) are obtained from new listeners and ad-hoc HRIRs can be retrieved from them. Another way to get new listeners parameters is to measure a subset of the full 3D space HRIRs and extrapolate them in order to obtain a full 3D database. This partial acquisition system, of course, should be rapid and accurate. In this paper we present a system which allows for rapid acquisition and equalization of HRIRs for a subset of the 3D grid. Then a technique to carry out HRIR customization based on the measured HRIRs will be described.

1. INTRODUCTION

Binaural hearing systems aim to reproduce at one listener's ears the same sound field that he would perceive in a target listening space. This is done by direct binaural recording in the target listening space, or in a synthetic way, by the convolution of a dry signal with the target Room Impulse Response and the listener Head Related Impulse Responses corresponding to

specific locations; in the following we will focus on this last approach.

Head Related Impulse Responses (HRIRs) are defined as:

$$HRIR(f,\phi,\theta) = \frac{P_{\phi,\theta}(f)}{P_{REF}(f)},$$

where P(t) is the pressure at the eardrum of a subject due to a source coming from a direction defined by θ and φ and $P_{REF}(t)$ is some reference pressure, usually taken at the position of the listener head centre. This division also performs equalization, if the two pressures are obtained with the same measurement system. The Fourier Transform of an HRIR is called Head Related Transfer Function (HRTF).

As pointed out in [1], 'achievement of spatial consistency requires rendering static, dynamic, and environmental cues'. In fact it is known that the introduction of environmental cues, such as the room impulse response or some kind of reverberation, provides sound externalization in synthesized binaural recordings. Static cues provide localization and rely essentially on HRTFs. HRTFs are strictly individual, and can considerably vary from subject to subject; localization degradation due to non-individualized HRTFs has been observed, namely in term of front/back resolution. This effect can be reasonably be reduced by letting the listener moving its head to obtain front-back distinction. Dynamic cues are provided through HRIR real time interpolation and head tracking.

Individualized HRIRs are obtained from measurements of impulse response of sources placed on a 3D point grid. The most complete (public) databases for real subjects HRIRs are the CIPIC ([2]) and the IRCAM Listen ([3]) databases, which present HRTFs sets for 45 and 51 subjects. The databases include 3D HRIRs measurements with different spatial resolution and anthropometric measurements for the listeners. The underlying measurement systems are quite complex and present long acquisition times.

HRIR customization tries to overcome the full measurement process. The perfect approach is to solve the wave equation with the real subject head, for example by BEM ([4]). This technique still presents an important acquisition time. Several approached strategies are under investigation: they can be classed in *selection methods, structural methods, decomposition methods and interpolation methods.*

In selection methods the customized HRIRs set is chosen between some HRIRs sets contained in pre-recorded databases, on the basis of interactive choice of the new listener, or morphological proximity between the new listener and the database subject ([1], [5]). In structural methods ([6]) the HRTFs are obtained as cascade filters representing physical scattering on the different body parts that can be tuned according to the morphology of the new listener. Decomposition methods ([7]) associate Principal Components Representation of HRTFs to a reduced set of morphological parameters, in order to synthesize HRTFs from anthropometric measurements. In interpolation methods ([8]) a pre-clustering phase reduces the number of HRIRs to be measured on the new listener, and interpolation is then performed to obtain the full HRIR set.

The technique presented in this paper can be considered a selection method based on HRTFs proximity. A way to perform customization in this case is to measure a subset of the full 3D HRIR space and extrapolate it to a full 3D space. A way to do this is to compare the measured HRIR subset to the corresponding HRIR subset present in existing database, select the most fitting one and use the corresponding full 3D set as customized set. The database selection is not made on the basis of perceptual parameters or anthropometric measurements, but directly on the proximity of the new listener's HRTFs to the database HRTFs. This is why we call this customization approach *Direct Customization.* If the HRIRs subset is composed by a consequent number of measurements points, it could be interesting to *integrate* these measurements with the selected set more than simply use it as a reference. This means that, instead of using the measured subset as a tool to select a whole database set, it could preferable to use the measured HRTFs (the new listener own HRTF) and *complete* these with the HRTF of the most fitting database set. In this case the two subsets have to be rendered homogeneous, correcting the effects of the different measurements systems. In this case we will talk of *Integrative Direct Customization.*

To be applied in HRIR customization, the HRIR subset measurement system should be rapid and accurate.

In this paper we present a rapid measurement system that employs a circular loudspeaker array in an anechoic

room and a fast-and-easy-to-apply binaural microphone in *open meatus* configuration. The overall measurement process takes less than 5 minutes; the subject can leave with a full 3D customized HRTF set on his USB key 10 minutes after his arrival.

Accuracy is obtained through electronic chain equalization and a frequency-dependent time windowing post-processing step to window out parasite reflections. Low frequencies HRTF resolution is guaranteed by employing a proper time window.

After a preliminary section to show the link between low frequency resolution and space-time windowing, we will present the acoustical properties of the measurement system in section 3. The measurement process is described in section 4, while the post processing is treated in section 5, where we present the equalization and the frequency-dependent time windowing. The direct customization process is detailed in section 6.

2. PRELIMINARY REMARKS

2.1. Time-frequency remarks

Audio signals are wideband, in the range 20-20000 Hz. They can be thought as composed by sinusoidal components, according to the Fourier Theorem. Sampling a wideband audio signal means sampling each one of its sinusoidal components.

To correctly sample an audio signal, it is necessary to use a sampling frequency bigger than the Nyquist frequency, in order to avoid aliasing. Using the correct sampling frequency guarantees the correct representation of high-frequencies components. What is sometimes forgotten is that the sampling time windowing length is an important parameter for the resolution of low-frequencies. This is of scarce importance in long audio files, but for shorter files (some ms long, as in HRIRs), this parameter becomes important.

Let us suppose that we sample an audio file, and investigate how each frequency components is reproduced after sampling. In figure 1 we show a 50 Hz component sampled at 44100 Hz on 200, 1000, 10000 samples. It is quite clear from the figure that with 200 samples the 50 Hz component is not resolved, while resolution gradually improves augmenting the number of samples.

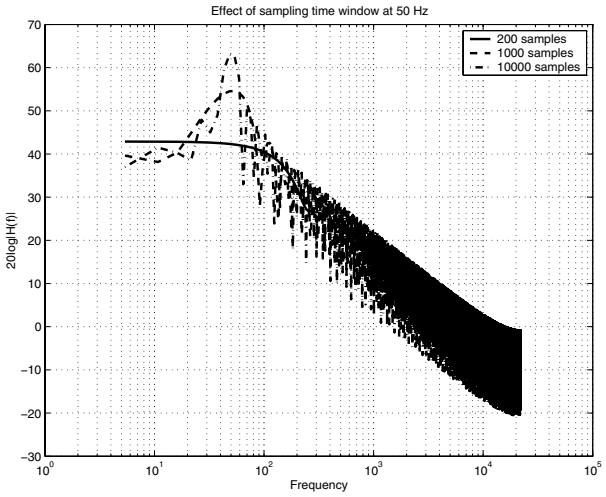

Figure 1. Sampling window effect at 50 Hz

In figure 2 it is easy to see that a sampling window of 50 points resolves the 1000 Hz frequency. This is due to the ratio between the lobe width and the observed frequency, and not to a reduction of the lobe width.

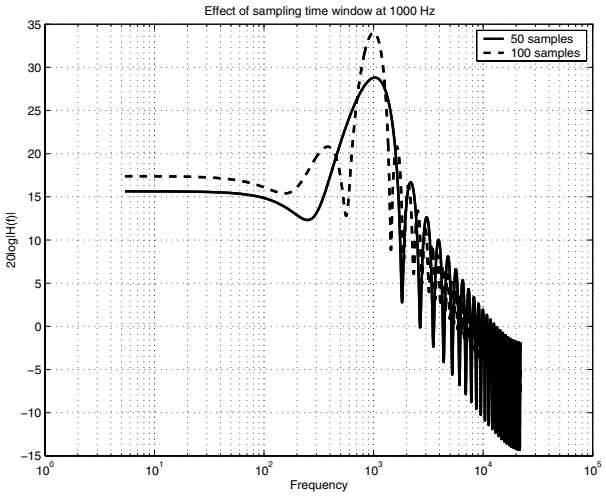

Figure 2. Sampling window effect at 1000 Hz

The main lobe due to the convolution of the ideal delta function with the sinc function related to the rectangular analysis window presents a width inversely proportional to the sampling window NT_s, where N is the number of samples and Ts the sampling time. In figure 3 we verify the theoretical result keeping fixed the frequency of interest (say f_0) at 50 Hz and plotting the frequency resolution

$$R = \frac{B_{-3dB}}{f_0}$$

as a function of the samples number. B_{-3dB} is the width of the principal lobe at -3dB.

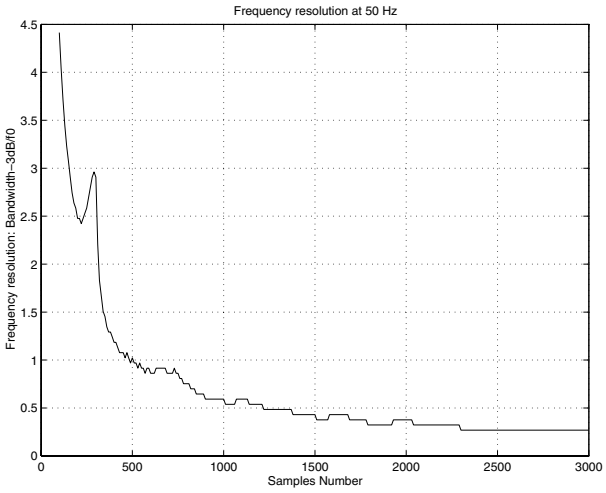

Figure 3. Frequency resolution at 50 Hz.

We observe the expected 1/N behaviour of the function and conclude that a 0.5 frequency resolution at 50 Hz is guaranteed with a sampling window of at least 1024 samples.

2.2. Space-time remarks

HRIRs measurement systems are usually located in enclosed spaces (rooms, anechoic rooms, etc.) and are composed by more or less extended measurement equipment (microphones, loudspeakers, loudspeakers supports, etc.). All these elements can potentially provoke interference in the measurements. Keeping long impulse responses on one side improves low-frequency resolution, but on the other side can dangerously make Room-and-Head Related Impulse Responses of supposed only-Head Related Impulse Responses.

In Table 1 we give some values of correspondence between samples number, impulse response length and space radius spanned by the response. For an N-samples impulse response to be 'pure' the corresponding radius should be free of diffracting or reflecting objects.

METRES	MS	~SAMPLES @44100 Hz	~SAMPLES @48000 Hz
0.5000	1.5	65	71
1	2.9	129	143
1.50	4.4	194	212
2	5.9	259	285
2.50	7.4	324	353
3	8.8	390	424
8	23.5	1037	1130
12	35.3	1557	1700
17	50	2205	2400

Table 1. Space-time correspondence

The goal of the measurement system is obviously to keep the largest number of samples avoiding (as long as it is reasonably possible) parasite reflections. This can be done by simply windowing out the impulse response which is thought to be 'spurious' or, in an optimized way, by frequency dependent time windowing, as we'll see in section 5.

3. SYSTEM CHARACTERIZATION

The aim of this section is to give an acoustical description of the system, in order to prove its possibility to provide pure HRIR and eventually define the better strategy to employ for equalization.

The electroacoustic chain is composed by a TASCAM MWE 24x24 direct-to-disk recording system, linked by Ethernet to a portable computer for audio downlink and uplink. The TASCAM output is input a 6 channels Yamaha 6150 power amplifier, which feeds 6 Tannoy system 600 loudspeakers. The signal is recorded by a Sennheiser MKE 2002 binaural microphone, linked to a Behringer preamp. The two-channel output is input to the TASCAM.

The measurement system is composed by electronic equipment in an acoustic environment (an anechoic room) that is supposed to be transparent to the measurement. The sweep is electronically filtered by the playback equipment and recording transfer function, and by the acoustic transfer function that represents the acoustical path from the loudspeaker to the microphone, and that includes reflections and scattering by the loudspeakers and the loudspeakers supports (and not by the walls, considered as perfectly absorbing).

We measured the impulse response of the electroacoustic chain, with the sine sweep method ([9]), in order to obtain the impulse response and transfer function from the TASCAM output to the TASCAM input.

3.1. Electronic chain

In order to characterize the only electronic chain we put one loudspeaker in an anechoic room (the ENST anechoic room, 4.20x4.50x4.60m) and a Schoeps MK2 microphone in the loudspeaker axis at the distance of 1 meter. The absorption by the air on the direct path is considered as belonging to the electronic chain. The obtained transfer function is plotted in figure 4, where it is compared to the one measured in similar conditions at IRCAM.

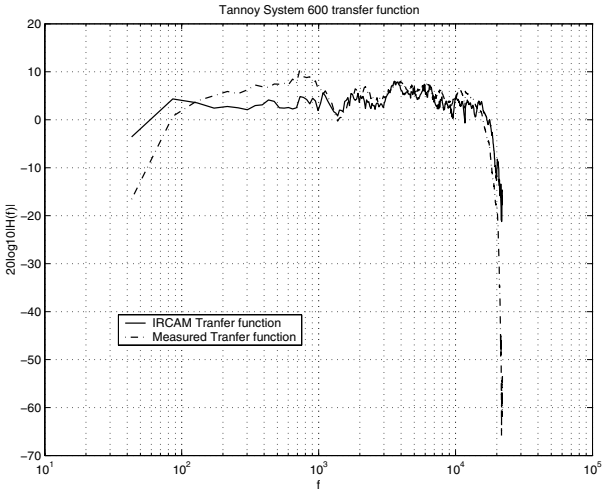

Figure 4 System Transfer Function, measured with the MK2 microphone (dB/Hz).

The measured transfer function represents the contributions of the cascade of the electronic chain components. We measured separately the contribution of the TASCAM, directly connecting its output and input. The response reveals to be flat in the interest frequency range. The response of the loudspeaker, preamp and amp has been obtained with a Schoeps MK2 microphone (assuming its frequency response flat).

The response we found is quite similar to the one measured at IRCAM. The major differences are sharpest decays at the range boundary, more energy in the low frequencies and a more marked notch around 1500 Hz, as it is possible to see in figure 4. We attribute these differences to the different playback and measurement equipment, and the different measurement conditions. It is important to remark that Tannoy claims a flat frequency response (-3dB values) in the range 52Hz-20KHz.

The response of the MKE2002 is qualitatively deduced by the previous measurements, comparing the transfer function measured with the MKE2002 and the previous transfer function. The two functions are shown in figure 5. We can observe that MKE2002 seems to loose sensitivity in the low frequency range: it presents -9dB at 50 Hz compared to the MK2-measured transfer function (1024 samples).

From these preliminary observations, we can observe that the system transfer function contains significant energy (at -3dB) in the range 100-19000 Hz; 512 points HRIRs guarantee 0.5 resolution at 100 Hz. In the following we present a technique that allows for low frequencies resolution down to 50 Hz (which is the claimed lower frequency bound for Tannoy loudspeakers), that corresponds to 1024 samples HRIRs.

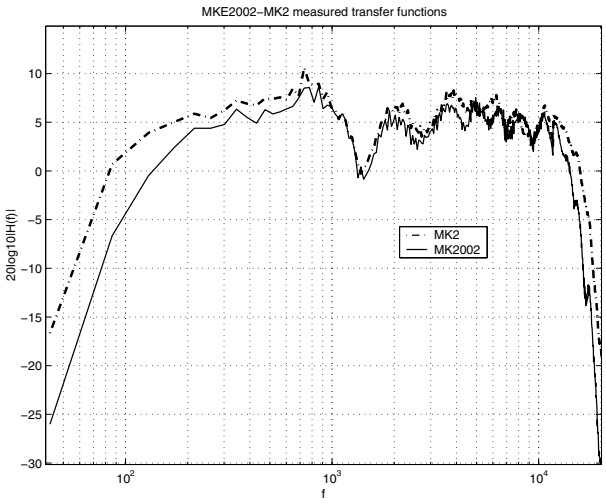

Figure 5. System Transfer Function, measured with the MKE2002 microphone.

We observed the degree of anechoicity of the room on the impulse response of the system (figure 6). It is possible to see that the first significant reflection is 40 dB under the main peak and then supposed inaudible: the room can reasonably be thought as anechoic in this loudspeaker configuration.

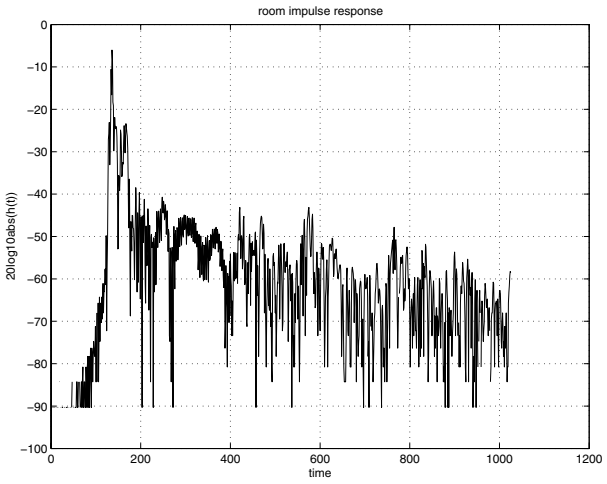

Figure 6. Anechoic Room impulse response (dB/ms).

The reflection takes place around samples 430 and corresponds to a reflection on a wall (3.2 meters path, for 1 meter direct path).

3.2. Electroacoustic chain

In this section we aim to characterize the full electroacoustic chain, including all the paths from a loudspeaker to the microphone, in presence of the complete measurement system, for two potential measurement systems.

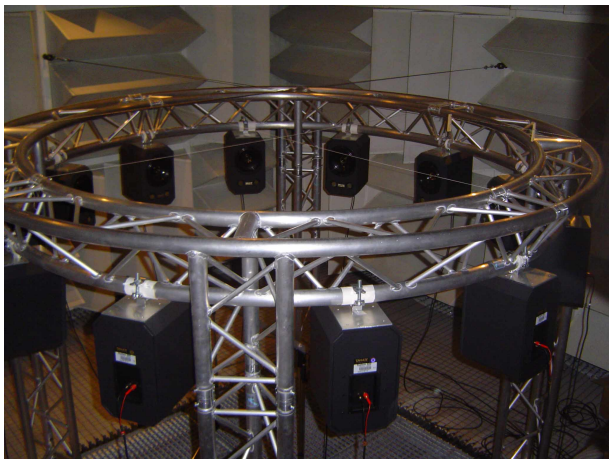

Figure 7. Ancient measurement system

We mounted a first measurement system in the anechoic chamber of the ENST, which measures 4.20 x 4.50 x 4.60 m. The structure was composed by an aluminum support and a set of 12 Tannoy system 600 passive loudspeakers (see figure 7).

Figure 8. New measurement system

Even if the structure was elegant and ergonomic, it only allowed for azimuthal HRIR measurements and, most of all, it provoked an acoustical interference that acoustical correction did not completely correct, and that affected in quite a dramatic way the purity of the measured HRIRs. We chose to redesign the system in order to allow for a more significant sampling of the 3D spatial grid, and to reduce the amount of interference due to the loudspeaker support.

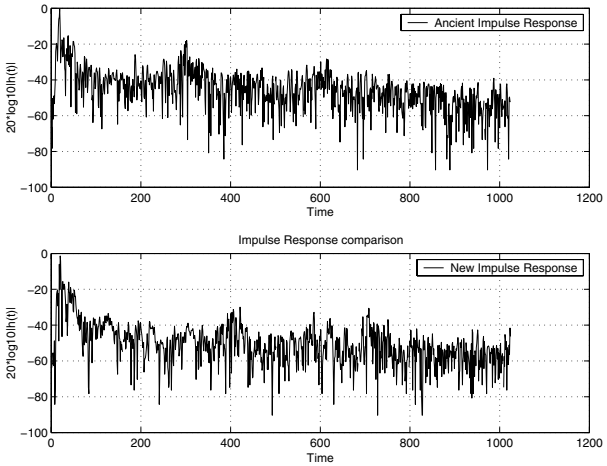

Figure 9. Old and New system impulse response comparison

The actual system is shown in figure 8. We reduced the number of measurement positions to 6, augmented the structure radius, used normal, thin, loudspeaker support, considered different elevations, and used glass wool to isolate loudspeakers supports and loudspeakers face. The new system presents a cleaner impulse response:

the first reflection is 11 dB less compared to the ancient impulse response (figure 9). First significant reflection is also translated in time from 270 to 430 samples after the direct front. In figure 10 we report the differences between the one-loudspeakers and the 6 loudspeakers transfer function. The effect of the structure is clearly visible around sample 400, where a reflection against the in-front loudspeaker takes place. The direct front response tail (until 390 samples) stays below 35 dB and we will consider it as immune from parasite reflections.

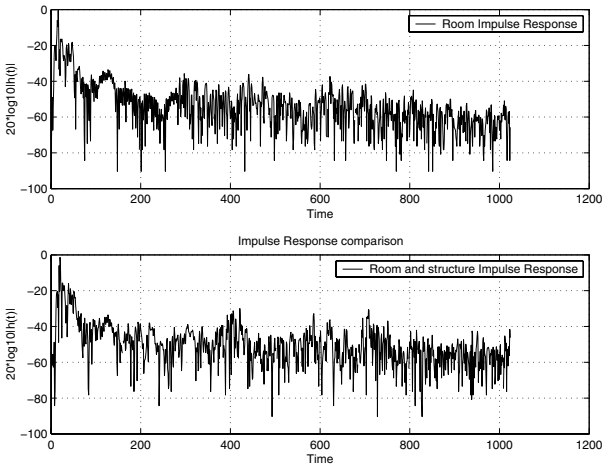

Figure 10. One loudspeaker and full structure impulse response comparison

3.3. System calibration

A calibration of the loudspeaker distance from the center reference point has been carried out, measuring the flight time from each loudspeaker to the microphone positioned at the center of the structure. The maximum distance difference is 4 samples, which correspond to a misplacement of 3 cm. A calibration for loudspeaker amplitude has equally been made, for a reference taken at the center of the structure.

The transfer functions at the center of the structure can vary from one loudspeaker to another, as it is shown in figure 11, considering only the direct front. This is due to the differences in loudspeaker manufacture and positions but mostly to the loudspeaker different elevation from the microphone, which determines out-of-axis coloration. Loudspeakers set at the same elevation present more similar functions (figure 12), which let us assume that loudspeakers transfer functions are quite constant for the 6 loudspeakers. Analogue results have been obtained for the same loudspeaker at two recording positions (for example at the two ears position). The difference in this case can be due to capsule differences too.

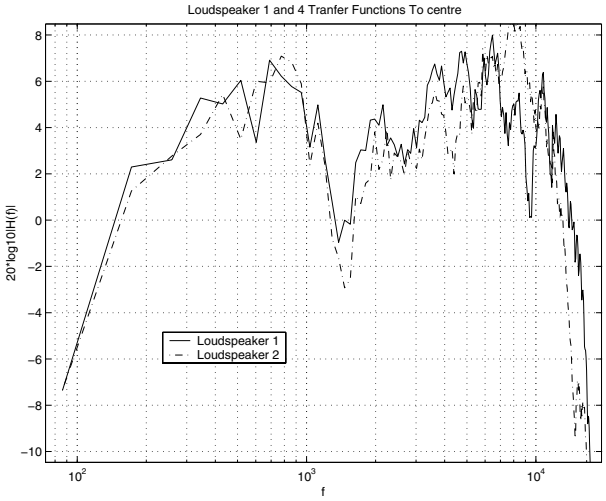

Figure 11. Loudspeakers response comparison

3.3.1. System sensitivity to little variations of measurement points

We put the MKE2002 at the virtual position of the ears, which are marked with the reference support that is visible in figure 8. Then we place it again, at the virtual position of a bigger head.

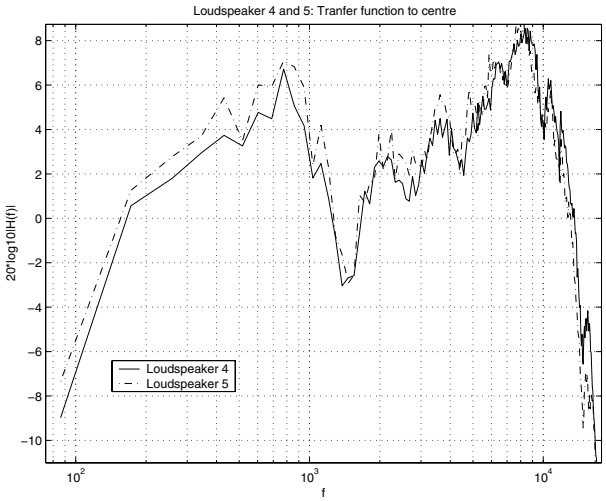

Figure 12. Equal elevation loudspeaker response comparison

The Interaural Time Difference (ITD) is coherent with the new head size, and a 0 value is guarantee in both cases for azimuth zero, as it is possible to see in figure 13. Only minor spectral differences (<1dB) are reported for the transfer functions corresponding to the same ear position, in the two measurement situations.

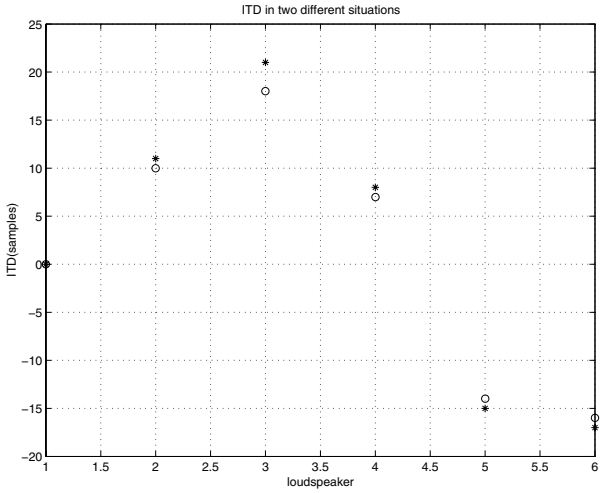

Figure 13. ITD for two heads of different dimensions.

3.4. System characterization: conclusions

Experimental data, obtained through the observation of the 6 impulse responses let us fix the direct front boundary to **350** samples after the direct front arrival. This length guarantees a 0.5 frequency resolution at 150 Hz. An equalization step is necessary to remove the coloration due to electroacoustic chain and guarantee 50 Hz frequency resolution through longer windowing.

The sensitivity of the system to different loudspeakers and measurement position has been considered comparing the corresponding transfer functions. Even if some differences have been observed, a compromise has to be found between the use of a different equalizer for each loudspeaker and each position, and the use of a reduced number of equalizers robust to variations and easily scalable with different measurement geometry choices.

4. MEASUREMENT

The measurement points are marked with a fixed support, so that they not vary from one measurement to the other. The person is simply asked to sit and tune the seat height to make the back of the ears touch the reference supports, and then the MKE 2002 (figure 14) is applied.

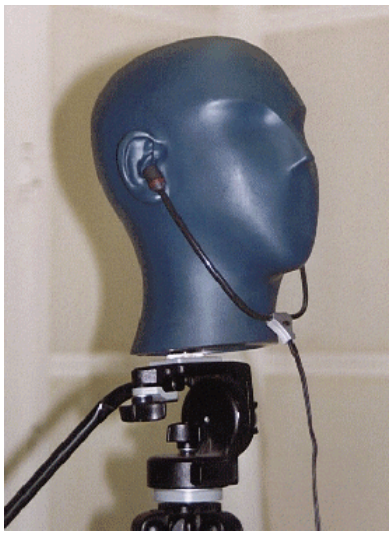

Figure 14. MKE2002 microphone

The 6 signals are played through the corresponding loudspeaker. The overall measurement process takes less than 5 minutes. The recorded wave file (44100 Hz) consists in the sequence of the 6 recorded chirps. The signal is downloaded from the TASCAM and directly provided to the processing module. In this preliminary phase we measured four sets of HRTFs.

MKE2002 performs *open meatus* HRIR measurements: occlusion of the ear channel by a plug would results in an unnatural high frequency reflection on the plug, as the capsule is not directly fixed on it. *Closed meatus* techniques are usually preferred because ear channel doesn't provide more spatial information and only adds individual information. For these reasons the blocked-entrance pressure is more likely to be representative of a population [10].

In order to fully characterize the individual hearing properties, open meatus measurements can be used. Open meatus measurements with the MKE2002 are taken 6mm outside the ear channel entrance and include the ear channel resonance. The first constraint does not deteriorate the spatial information contained in the HRTF ([9]) and the resonance can be removed at the headphones equalization stage.

5. PROCESSING

The processing module carries out the following tasks: impulse response extraction, equalization and frequency dependent time windowing, HRTF customization (figure 16). The processing is performed in differed time. The overall processing time is 20 seconds

The extraction module performs the sweep deconvolution, the synchronization with the recorded clock and the HRIR selection through peak finding and windowing, to isolate with a rectangular window the 6x2 1024 samples impulse responses. The equalization and customization modules are treated in detail in the next sections.

5.1. Equalization

The goal of a HRIRs measurement system is to obtain the free field impulse response from a source in a given position to the listener's ears. This is the condition that guarantees the good working of the convolution process, following the Green Theory. In practice we obtain the response to an impulse filtered by the electronic chain in en enclosed space. Both the electronic coloration and the room reflections are spurious elements to be eliminated from the measurements. The electroacoustic chain is depicted in figure 15.

The full electroacoustic chain can then be expressed as

$$H(f,r) = H_S(f,r)HRTF(f,r),$$

where $H_S = H_{TO}H_Y H_T H_R H_M H_B H_{TI}$, and r is the measurement position.

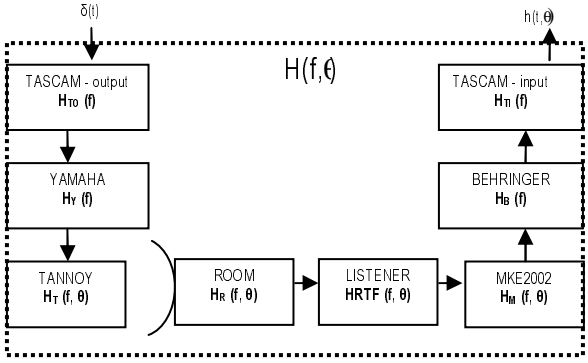

Figure 15. Electro-acoustic chain

If we want to separate the acoustical (two dimensional-space, time) and the electronic (mono dimensional, time) parts of the chain, we assume that we can write the transducers transfer functions as

$$H_T(f,r) = H_{T_{r_0}}(f,r_0)H_{Tr}(f,r), \text{ and}$$
$$H_M(f,r) = H_{M_{r_0}}(f,r_0)H_{Mr}(f,r),$$

where $H_{T_{r_0}}(f,r_0)$ and $H_{M_{r_0}}(f,r_0)$ are intended to be measured at a reference position r_0 in free field. In this way we can write the transfer function of the electronic chain $H_E(f,r_0)$ as

$$H_{TO}(f)H_Y(f)H_{T_{r_0}}(f,r_0)H_{M_{r_0}}(f,r_0)H_B(f)H_{TI}(f),$$

We remark that $H_E(f,r_0)$ does not depend on the measurement position. The transfer function of the acoustic chain without the listener $H_{AWL}(f,r)$ as

$$H_{Tr}(f,r)H_{Mr}(f,r)H_R(f,r),$$

so that

$$H(f,r) = H_E(f,r_0)H_{AWL}(f,r)HRTF(f,r)$$

In an ideal system H should be equal to HRTF, and then we need to equalize the measured response, which suffers from interference of the room and the electroacoustic chain. Of course, it is possible to invert directly the product of the two transfer functions.

To quantify the performances of equalization we define, similarly as in [11], the equalization efficiency in the frequency domain as the cepstral distance between the target function and the equalized transfer function. We suppose that the target function is the Fourier transform of a delta function that is a flat spectrum. We assume the value of the constant spectrum to be the mean value of the equalized response.

$$j(k) = 10 * \log_{10}\left|\tilde{H}(k)\right| - 10 * \log_{10}\left|\overline{\tilde{H}}\right|, \text{ where}$$

$\tilde{H}(k)$ is the equalized electroacoustic response. Other choices for the target function ([15]) are possible.

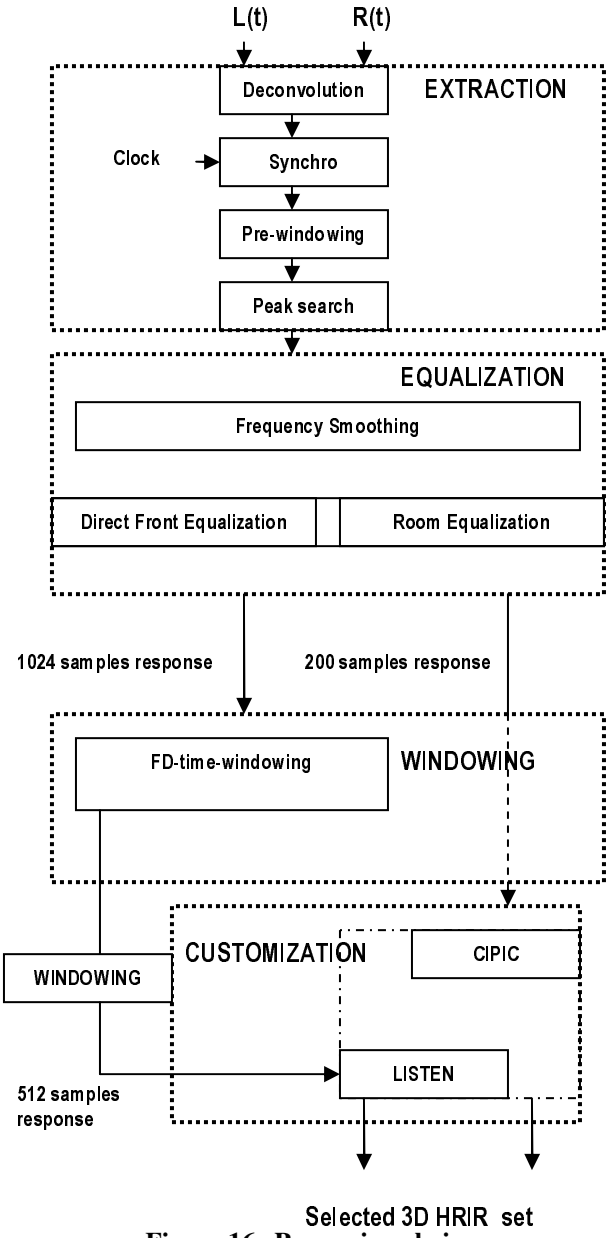

Figure 16. Processing chain

We also define the standard deviation, J, as:

$$J = \sqrt{\frac{1}{K}\sum_{k=1}^{K} j(k)^2} \ ,$$

where K is the number of points on which we perform the FFT. Another relevant parameter is the maximum deviation

$$J_{\max} = \max(|j(k)|)$$

These values should be better qualified with perceptual tests or with psychoacoustic weights, but we can use it as reference.

5.1.1. Electro-acoustic chain equalization

One could think that the best equalization strategy would be to store the twelve impulse responses from each loudspeaker to each capsule, and build 12 different equalizers to apply in post-processing to each measurement. This is a possible technique, but not optimal, nor scalable. In fact little errors on reference position and measurement points can lead, due to a complex room interference pattern, to severe deteriorations in the equalization process, most of all for high frequencies. Moreover changing in the loudspeaker setup would need another impulse response measurement session.

We will use two equalizers, obtained from measurements on one loudspeakers (say loudspeaker 1), for the two channels left and right, and use them for the equalization of all the loudspeakers. We consider exact and smoothed equalization. In exact equalization we directly use the equalizers issued from inverse electroacoustic channel computation, following the algorithm presented in [12].

The basic idea of smoothing is that it's not worth to try to perfectly correct the transfer function of each loudspeaker at one point: the matching of the inverse filter for one position is not necessarily optimal for lightly different positions (due, for example, to different loudspeaker positions). Because of this, it could be interesting to 'smooth' the reference impulse response, so as this becomes more representative of the impulse response in proximity of the measurement point or of the responses of loudspeakers at different positions. This is, of course, a cause of deterioration of the performances for the reference point, but it should reveal better for robust equalization. Smoothing is performed in the frequency domain, following [11], that is the one implemented in the Aurora Kirkeby Inverse Filtering plug in ([14]). The smoothing is performed on half octave windows.

5.1.2. Separated equalization

When we apply equalization on the electroacoustic transfer function with smoothing, we do not control the percentage of smoothing on the direct front and on the room. We could think that applying perfect equalization for the direct front and a smoothed equalization on the structure response could lead to an optimal equalization.

We can obtain $H_E(r_0)$, the transfer function of the electronic chain, by measuring the response of a single loudspeaker with a microphone positioned at a distance of 1 meter, oriented to the center of the speaker and elevation 0, in order to avoid not-in-axis coloration, in an anechoic room that is supposed equivalent to free field.

The electronic transfer function represent the direct wave front contribution, and can be equalized in quite a drastic way, because it is supposed not to depend on the error on the measurement position (r_0 is the real angle at which the measurement takes place). On the other side, H_E lightly depends on errors on the measurement, due, for example to different head dimensions, or imperfect replacement or placement of the binaural microphone, or to the loudspeaker position.

Once the direct front has been equalized for all the loudspeakers, we can obtain the room equalizers. To do this, we just consider the direct-front equalized impulse response for a loudspeaker, obtain the Kirkeby inverse filter from this partially equalized response, and then use it as we did for direct front equalization, for the equalization of all the direct-front-equalized transfer functions.

5.1.3. Results

The obtained results are shown in table 2.

Electronic Equalization of the direct front presents the best results. It is performed on a 350 samples windowed version of the HRIR and results are reported only for the direct front equalization. As in this case the response is not sensitive to measurement positions, perfect equalization can be used in a robust way. This explains the superior quality of equalization. In Electronic Equalization smoothing the transfer function is a suboptimal operation that does not improve robustness, but only determines performances deterioration. Direct front equalization only allows for 350 samples-long HRIRs. In order to obtain longer HRIRs, an equalization of the response tail is necessary, as explained. Electroacoustic equalization performs jointly the two equalizations. We can compare Electroacoustic Equalization and Acoustic Equalization of the previous direct-front equalized impulse response.

Electronic	**J (dB)**	**J_{max}(dB)**
Perfect	0.75	2.8
Smoothing	0.85	4.5
Electroacoustic		
Perfect	0.93	5.85
Smoothing	0.94	4.84
Acoustic		
Perfect	0.82	7.5
Smoothing	1.12	15

Table 2. Equalization Strategies Comparison

It is possible to observe that smoothing plays an important role (as forecast) in **Electroacoustic Equalization**, providing robustness to equalization (see J_{max}). Perfect equalization work perfectly for the first loudspeaker (the one taken for the reference transfer function): this explains the inferior J value compared to smoothed equalization. For the other loudspeakers smoothing provide better performances, as it is possible to see in figure 17.

For **Acoustic Equalization** of the full direct-front-equalized response we can observe that perfect equalization leads to better results only for the mean value of J, and a smoothed equalization provides in general worst performances than the electroacoustic equalization. These results can be explained not by the worst global performances of the separated equalization, but by its more marked sensitivity to particular bad cases, namely in very high frequencies (above 10 KHz).

Performances are in general lightly better than those of electroacoustic equalization, but we observed that in some particular cases (namely loudspeaker 5), the equalization process leads to strong artifacts,

concentrated in very high frequencies, where the room equalization is particularly critic. Performing complete electroacoustic equalization appears to smooth these ill-conditioned cases.

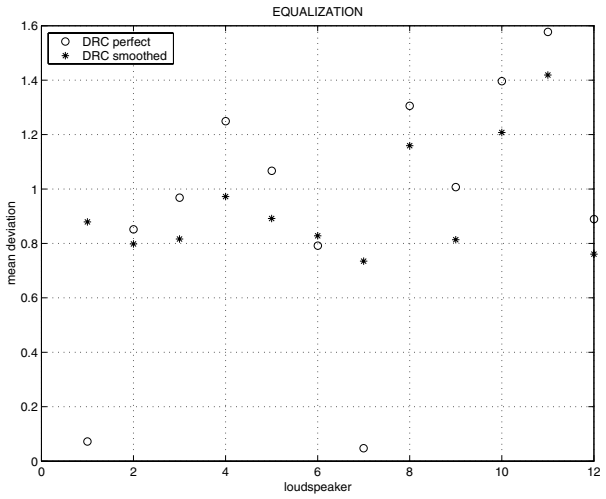

Figure 17. J for electroacoustic equalization

5.1.4. Conclusions

Taking a look to one equalized impulse response (figure 18), we see that Electroacoustic Equalization doesn't eliminate completely the reflections around sample 400. The residual weak reflections present in the obtained impulse responses determine a weak frequency comb filtering effect that can affect the local behavior of the impulse response though preserving the global properties of the measured HRTFs.

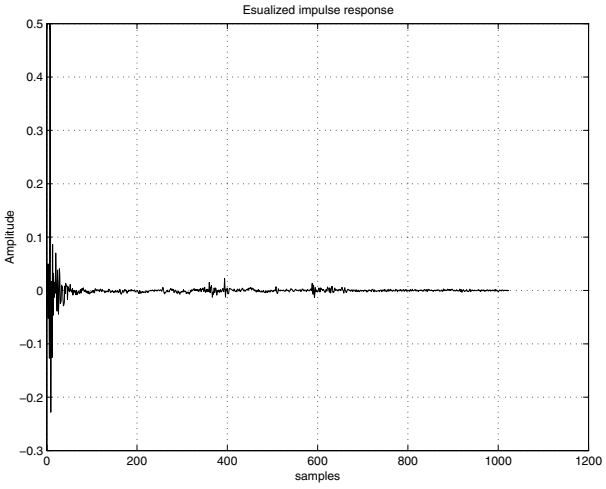

Figure 18. Direct front equalized impulse response.

To obtain pure HRIRs, these reflections should be windowed out. On the other side, windowing the impulse response at sample 350 would not allow for a frequency resolution below 150 Hz.

Another consideration can then be made. The reflections on the loudspeakers and the measurement structure only affect components above a certain frequency. It is known that scattering on an object takes place for wavelengths comparables with the object dimensions. In this case we could fix this frequency to 400 Hz, for which half the wavelength (around 40 cm) is comparable with the loudspeaker dimension. We could think to filter the response tail above sample 350 with a low-pass filter with cutoff frequency of 400 Hz, and keep this part of the impulse response, that could be mixed with the all-pass direct front response. We just keep the high frequency response until sample 350, which includes all the useful reflections. The mix with the low-passed impulse response from sample 350 to sample 1024 guarantee the low frequencies resolution for sinusoidal components below 400 Hz, that are not interested by structure scattering

5.2. FREQUENCY DEPENDENT WINDOWING

5.2.1. Hard Frequency Dependent Time Windowing

Looking at the spectrogram (figure 19) of one of the impulse responses, we can clearly see reflections around 9.5, 9.7, 11, 16 and 17 ms, even if the first significant reflection is -36 dB under the main front.

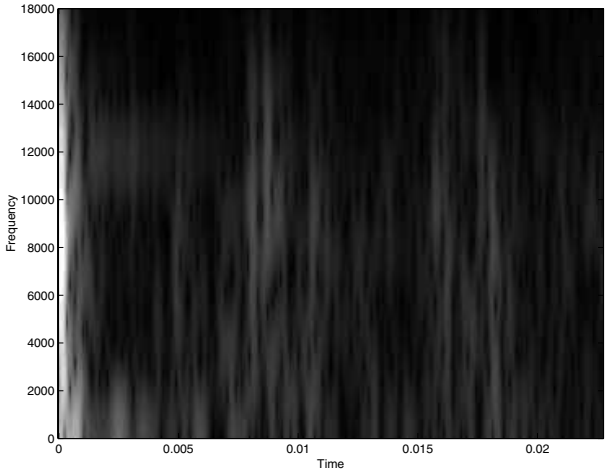

Figure 19. Direct front equalized impulse response spectrogram

If we high-pass filter the impulse response, with a cutoff frequency of 6000 Hz, and we look at the spectrogram (figure 20) , we can see that reflections are still present and easy to identify. On the other hand, filtering the impulse response with a low pass filter with cutoff frequency of 4000 Hz, we can see (figure 21) how the reflections have disappeared, and there is not strong discontinuity in the slow-varying low frequencies impulse response around 10 ms.

We consider as useful information the impulse response from sample 1 to 350 for high frequencies (direct front), then we window the direct front with a Tukey window, which allows avoiding sharp transitions at the cutting points, preserving the useful information. In a similar way, we filter the impulse response tail with a low pass filter and window it from sample 350 to sample 1024, with another Tukey window. We mix the two impulse responses with 20 samples overlapping.

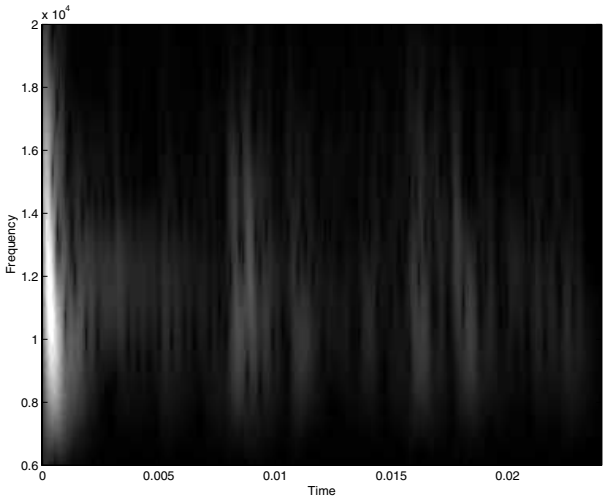

Figure 20.Direct front equalized high-passed impulse response spectrogram

This processing is usually called Frequency Dependent Time Windowing (FDTW) and it is usually used for room equalization, where the amount of correction is directly proportional to the length of the impulse response and inversely proportional to the considered frequency band. We call this kind of processing, where only one point of transition is applied 'Hard' FDTW.

One drawback of this process is not the temporal sharp transition, which is avoided by Tukey Windowing and overlapping, but the time-frequency sharp transition between an all-pass impulse response and a low passed impulse response. This phenomenon is clearly visible in figure 22, where we report the spectrogram of the H-FDTW processed impulse response. One way to solve this problem is to use what we call 'Soft'-FDTW that allows for smoother time-frequencies transitions.

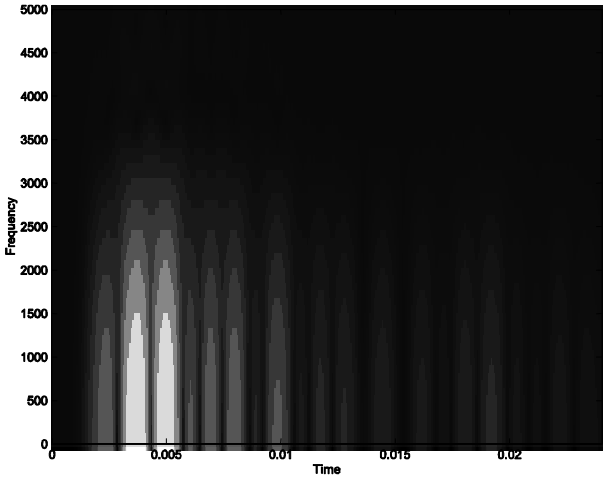

Figure 21.Direct front equalized low-passed impulse response spectrogram

5.2.2. Frequency-dependent-soft windowing

S-FDTW has been carried out, with the Denis Sbragion toolbox for Digital Room Equalization ([15]) that also provides C-written functions for homomorphic deconvolution, frequency dependent ringing truncation and peak limiting.

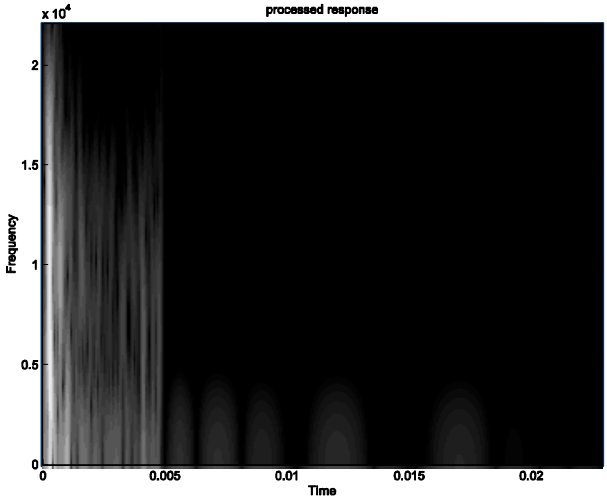

Figure 22. H-FDTW impulse response spectrogram

In the DRC toolbox, frequency dependent windowing is implemented following a band windowing approach or a sliding low pass filtering approach. The two techniques present similar results, even if the second techniques is said to be more stable and flexible. In both techniques two windows are defined at the frequency range boundary: a longer window for low-frequencies versions of the impulse response and shorter windows for higher frequency versions of the impulse response.

The first procedure simply applies a filter bank to the impulse response and uses different time windows for each subband signal. These signals are then summed back together to obtain the S-FDTW version of the impulse response. The second procedure filters the impulse response with a low pass filters with a cutoff frequency that decreases with the time window length.

In the first procedure the window length for each band is computed as

$$W = \frac{1}{A * (F + Q)^{WE}}$$

Where W is the window length, F the normalized frequency, A and Q are chosen in order to fit the boundary window length.

The parameter named WE is the window exponent and determines the decreasing slope for window length reduction. Smaller WE values determines biggest fraction of the time frequency plan to be suppressed, as it is shown in figure 23.

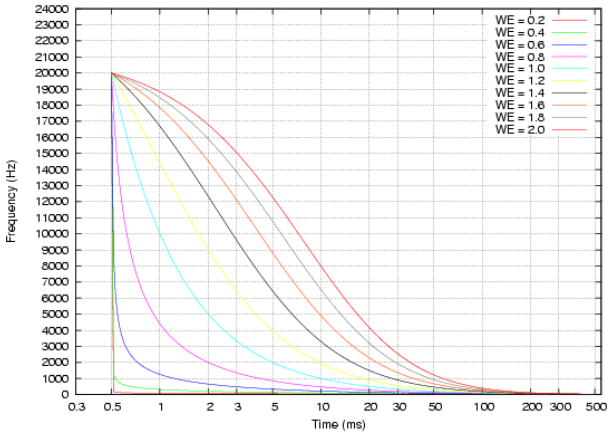

Figure 23. Window exponent for S-FDTW

In the second procedure

$$F_c = \frac{1}{A * (W + Q)^{WE}},$$

where F_c is the cut-off frequency of the sliding filter.

The S-FDTW impulse response spectrogram is reported in figure 24. It clearly presents a smoothed transition to the impulse response tail in the time-frequency plan.

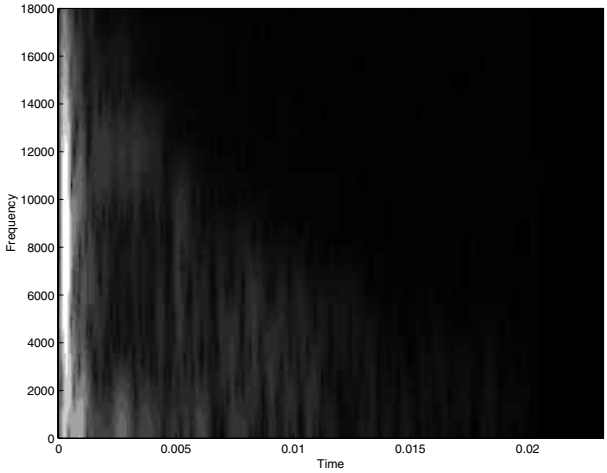

Figure 24. S-FDTW impulse response spectrogram

6. HRIR DIRECT CUSTOMIZATION

The two retained public individual HRIRs databases are the CIPIC and the IRCAM LISTEN ones. The two databases are briefly presented in the following. The direct customization and the integrative direct customization module are then described.

6.1. CIPIC Database

The CIPIC database contains HRIRs measured for 47 subjects. The measurements have been made at the CIPIC Interface Laboratory at the University of California. For each subjects a set of 27 relevant anthropometric measurements has been provided and two sets, one for the right ear and one for the left ear of HRIRs sampled on a grid of 750 points.

The grid does not represent a uniform 3D sampling. The measurements are given for 50 different elevations and 25 azimuths, following the convention showed in figure 25. θis called azimuth and its range is [-80, 80], more

densely sampled in the midsagittal [-45, 45] plane; φis the elevation and its range is [-45,230.625], uniformly sampled with a step of 5.625 degrees. These values don't match all the measurement positions of our system. In table 3 we show the values of ENST system points, and the ones of CIPIC that more match these positions.

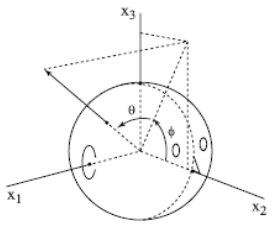

Figure 25. Interaural-polar coordinates system

In order to obtain the maximum precision, we should interpolate CIPIC HRIRs, or change the measurements points in ENST system. In this phase we just eliminate the [90,15] point, that presents an unacceptable difference in azimuth, and use the remaining 5 positions as index. We tolerate an error of localization of 5 degrees for the azimuth and 1.8750 for elevation.

The measurements have been made with the Snapshot System by Aureal Semiconductors, and Etymotic Research ER-7C probe microphones (in closed meatus configuration).

θ, φENST	θ, φCIPIC
0,0	0.0
30,0	30,0
(90,15	80,16.8750)
60,120	65, 151.8750
-60,120	-65,151,8750
-60,15	-65,16.8750

Table 3. Measurement position comparison

The ear channel resonance is not included in the measurements. 6.4-cm diameter Bose Acousticmass loudspeakers have been used. The measured HRIRs have been windowed with a 200 window (4.5 ms) in order to eliminate parasite reflections (even if some floor or knees reflections are present for some positions). The responses have been compensated with a perfect equalizer obtained from free field measurement at the position of the center of the head. Some low-frequency compensation has been introduced (see the documentation of the CIPIC database for more details).

We can assume that the obtained HRIR can be written in the form

$$HRTF_{\varphi\theta}^{CIPIC}(f) = H_{\varphi\theta}(f)\tilde{H}_{\varphi\theta}(f)HRTF_{\varphi\theta}(f)$$

Where $HRTF_{\varphi\theta}^{CIPIC}(f)$ is the Fourier Transform of the CIPIC HRIR, and $H_{\varphi\theta}(f)\tilde{H}_{\varphi\theta}(f) = H_{\varphi\theta}^{CIPIC}(f)$ is the global CIPIC transfer function that takes into account the CIPIC measurement system transfer function and the CIPIC equalization for the real HRIR.

6.2. IRCAM LISTEN Database

The IRCAM LISTEN database contains HRIRs measured for 51 subjects. The measurements have been made in the IRCAM anechoic room in Paris. For each subjects a set of 27 relevant anthropometric measurements have been provided and two sets, one for the right ear and one for the left ear of HRIRs. The sampling grid consists of 10 elevation angles starting at -45° ending at +90° in 15° steps vertical resolution. The steps per rotation vary from 24 to only 1 (90° elevation). As a whole, there are 187 measurement points.

The provided points match ENST system measurement points. The measurement has been carried out with a Tannoy system 600 loudspeaker, using the sweep sine method. A pair of Knowles FG3329 probe microphones has been used in closed meatus configuration. The measurement technique and equipment are similar to the one used in the present paper.

Diffuse field equalization [13], and windowing with a 512 samples windows have been carried out (see the LISTEN web site for more information). As we did with CIPIC, we define the global IRCAM transfer function as

$$H_{\varphi\theta}(f)\tilde{H}_{\varphi\theta}(f) = H_{\varphi\theta}^{IRCAM}(f).$$

6.3. DIRECT Customization

As we did for the two databases, we define as $H_{\varphi\theta}(f)\tilde{H}_{\varphi\theta}(f) = H_{\varphi\theta}^{ENST}(f)$ the ENST global transfer function (that also contains the ear channel resonance due to open meatus measurement). In Direct customization we assume that

$$H_{\varphi\theta}^{ENST}(f) = H_{\varphi\theta}^{CIPIC}(f) = H_{\varphi\theta}^{IRCAM}(f) = 1(f),$$

which is coherent with the fact that each one of these transfer function is intended to be an equalized version of the electroacoustic chain transfer function.

Let us suppose to have N measured reference K point HRTFs and M numbers of HRIRs sets in the database. No difference between left and right ears. Direct customization follows the scheme presented in figure 26. A proximity criterion W between the reference and each one of the databases HRTF sets is obtained from a metric characterizing the distance between two HRTFs. This metric can be defined simply as the cepstral distance between the two functions,

$$D_{ij} = \frac{1}{K}\sqrt{\sum_{k=1}^{K}(\log(HRTF_i(k)) - \log(HRTF_j(k)))^2}$$

The proximity criterion between the measured reference set and the candidate full database set can be defined, for example, as the mean of the metrics for each reference position. The minimum value of the proximity criterion over all the databases ears provides the customized set index. Note that with this technique we made no distinction between right ears and left ears, and that the two customized sets for the new listener can come from ears belonging to different subjects in the database.

6.4. INTEGRATIVE Direct Customization

Direct customization is a method to choose, on the basis of a reduced reference HRTF subset, a 3D set of HRTF. The chosen set is then used in its integrality, and the reference HRTFs subset is discarded. The reference HRTFs are of course the own listener HRTFs and one could think to keep them and just complete the reduced 3D subset, especially if the number of measured personal HRTFs can be consequent.

In direct customization the global transfer functions are considered equal to the flat spectrum. Even if this assumption is not completely true (see for example the results in section 5), the customization process only means that we are comparing HRTFs measured with two different systems, and then use only one of them: The listener would be using HRTFs that present an additional coloration due to the measurement system used for measurement.

On the other side, integrating two HRTFs subsets, issued from different measurement systems, would represent a problem due to different coloration determined by equalization residuals, which in general are not the same for the two systems. This would lead to different perception of sound coming from locations associated to different subsets, without a physical reason to be. The two systems have to be 'adapted'. To do this, once the customized set has been found, a system equalizer is built.

Let us suppose to have the HRTFs issued from two measurement systems, say A and B, where A is the subset reference measurement systems that provide personal HRTF and B a customization database. We can write:

$$HRTF_i^A(f) = H_{\varphi\theta}^A(f)HRTF_i(f) \text{ , and}$$
$$HRTF_i^B(f) = H_{\varphi\theta}^B(f)HRTF_i(f) \text{ i=1:N.}$$

The inverse filter obtained from B is

$$\varphi_i(f) = \frac{1}{H_{\varphi\theta}^B(f)HRTF_i(f)}.$$

Applying it to the measured reference HRTFs we obtain

$$\Phi_i(f) = HRTF_i^A(f)\varphi_i(f) = \frac{H_{\varphi\vartheta}^A}{H_{\varphi\vartheta}^B} \text{ , and}$$

$$HRTF_i^B(f)\Phi_i(f) = H_{\varphi\vartheta}^A HRTF_i(f).$$

We call Φ the database equalizer: applying Φ to the database customized (sub)set makes the two subsets homogeneous and then integration becomes possible. Using the database equalizer the ear channel resonance is corrected, too. We have to note that Φ depends on the measurement position, because measurement systems present loudspeaker-to-microphone transfer functions that are variable (see section 3). To obtain a significant mean database equalizer, we can use

$$\Phi(f) = \frac{1}{N}\sum_{1}^{N}\varphi_i(f).$$

or frequency smoothing techniques, as we did for electroacoustic chain equalization.

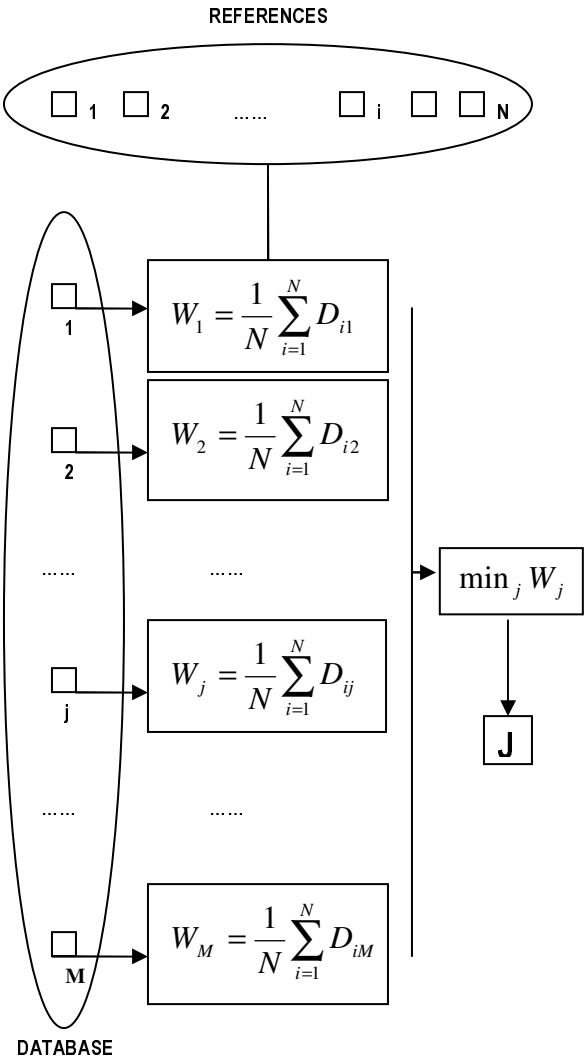

Figure 26. Direct customization principle

7. CONCLUSIONS AND FUTURE WORKS

This paper presents a rapid measurement system for index HRIR to use for direct customization. The acoustic characteristics of the system have been described with detail. The post processing module for impulse response extraction, equalization and frequency-dependent windowing has been tested in order to obtain some information about the performance of acoustical channel correction.

The reported results show that the smoothed equalized electroacoustic transfer functions present 0.94 mean deviation from the flat spectrum and max deviation of 4.84 dB; 0.75 and 2.8 dB are the corresponding values for direct front equalization and soft frequency dependent time windowing. These values let us think that measured HRIRs can be considered as representative of the only scattering by the head and the torso of the listener, and provide binaural cues for binaural synthesis.

The utilization of these measured HRIRs as index for direct and integrative direct customization has been described.

Localization and externalization quality of measured HRIRs have been proved with some informal tests on 4 subjects. In a future work we will carry out a measurement campaign with several subjects, and a perceptual tests campaign. The aim is to make clearer the following points:

1. The perceptual value of Equalization.
2. The perceptual value of low frequency resolution enhancement.
3. The perceptual value of direct customization, by direct in situ comparison of the real playback and the binaural playback with the measured HRIRs and the customized HRIRs.

8. REFERENCES

[1] D.N.Zotkin, R.Duraiswami, L.S.Davis, **"Customizable Auditory Displays"**, *Proc. 2002 International Conference on Auditory Displays,* Kyoto, Japan, July 2-5, 2002.

[2] V. R. Algazi, R. O. Duda, D. M. Thompson and C. Avendano, **"The CIPIC HRTF Database",** *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics,* pp. 99-102, Mohonk Mountain House, New Paltz, NY, Oct. 21-24, 2001.

[3] http://recherche.ircam.fr/equipes/salles/listen/

[4] B. Katz, **"Boundary element method calculation of individual head-related transfer functions. I Rigid model calculation",** JASA, 110(5) November 2001.

[5] J.C Middlebrooks, E.A. Macpherson. Z.Onsan, **"Psychophysical customization of directional , transfer functions for virtual sound localization",** JASA, 108(6), December 2000.

[6] C. Brown, O. Duda **"A structural model for Binaural Sound Synthesis",** IEEE Transactions on speech and audio processing, 6(5), September 1998

[7] C.Jin, P.Leong, J. Leung, A. Corderoy, S.Carlile, **Enabling individualized virtual auditory space using morphological measurements",** Proc. First IEEE Pacific-Rim Conf. on Multimedia, 2000

[8] C. Fahn, Y.C. Lo, **"On the clustering of Head-Related-Transfer Functions used for 3-D Sound Localization"**, Journal of Information Science and Engineering 19, 141-157 (2003)

[9] A. Farina, **"Simultaneous Measurement of Impulse Response and Distortion with a Swept Sine Technique",** AES 180th Convention, 2000 February 19-22, Paris, France

[10] D. Hammershoi, H. Moller, **"Sound transmission to and within the human ear canal"**, JASA, 100(1), July 1996.

[11] P. Hatziantoniou, J.N Mourjopoulos, **"Generalized Fractional-Octave Smoothing of Audio and acoustic Responses",** JAES, 48(4), 2000, April

[12] O.Kirkeby, P.Rubak, A. Farina, **"Design of Cross-Talk Cancellation Networks by using fast deconvolution"**, 106th AES Convention, Munich, 8-11 May 1999.

[13] J.M.Jot, V.Larcher, O.Warusfel, "**Digital signal processing issues in the context of binaural and transuaral stereophony",** 98th AES Convention, Paris, France, Feb. 25-28, 1995, preprint no. 3980.

[14] http://www.ramsete.com/aurora/homepage.html

[15] http://drc-fir.sourceforge.net/