Audio Engineering Society

Convention Paper 10385

Presented at the 148th Convention 2020 June 2-5, Online

An Innovative Method for Binaural Room Impulse Responses Interpolation

Valeria Bruschi[1], Stefano Nobili[1], Stefania Cecchi[1], and Francesco Piazza[1]

[1]*Department of Information Engineering, Università Politecnica delle Marche, 60131 Ancona (AN), Italy*

Correspondence should be addressed to Stefania Cecchi (s.cecchi@univpm.it)

ABSTRACT

The interpolation of room acoustic impulse responses is a widespread technique that allows to reduce measurement sets. In this paper, an innovative method for Binaural Room Impulse Responses (BRIRs) interpolation is presented and tested. In the proposed method, the BRIRs are decomposed in time and then divided in two frequency bands and an innovative peak detection and matching algorithm is applied for the early reflections, combined with a linear interpolation. A real room binaural impulse responses database has been employed in order to evaluate the algorithm comparing it with the state of the art. Experimental results have proved the effectiveness of the proposed approach.

1 Introduction

The increasing interest in the virtual reality has allowed the development of different spatial sound rendering techniques. The binaural reproduction over headphones is widely used in order to create an immersive audio scenario. This is achieved by considering the Head Related Transfer Functions (HRTFs), or their time-domain equivalent Head Related Impulse Responses (HRIRs), which contain the human ears characteristics and the localization cues [1]. The spatial audio virtualization is obtained by the real-time convolution with Binaural Room Impulse Responses (BRIRs), which are the combination of two parts: the Room Impulse Responses (RIRs) and the HRIRs, as described in [2]. In order to create a convincing 3D audio experience, the BRIRs need to change with the movement of the source or the listener, but it is not possible to measure the BRIRs in all the points of a room. The challenge is to reduce the number of measurements maintaining a correct perceptually spatialization and this is achieved by the BRIRs interpolation. In the literature, several approaches of HRIR and RIR interpolation have been proposed. In [3], a method for modeling a Room Transfer Function (RTF) is presented by using common acoustical poles and their residues. The residues correspond to the eigenfunctions of the room, so they are functions of the source and the receiver positions. In this way, the model can describe the RTF variations using simple residue functions. The effectiveness of this technique is more relevant for the low frequency component of the RTF. In [4], Kearney et al. have presented a novel method for the interpolation of the direct and early reflections components of the RIRs. They have employed

an algorithm known as Dynamic Time Warping (DTW) for the temporal alignment of the sparse reflections. In this case, the measured impulse responses are split in two parts: early reflections and diffuse decay. The first part is linearly interpolated after alignment and the tail is modeled according to the method of [5]. This technique prevents the distortions usually produced by the linear interpolation, but it is not suitable in the case of small rooms. In [6], a frequency domain approach is proposed and seems to reach better performances than the time domain techniques. A recent method for the interpolation of BRIRs has been presented by Garcia-Gomez in [7], in order to obtain a lower computational cost than the DTW algorithm of [5]. The BRIRs are decomposed in time and then divided in two frequency bands. The low band processing consists of a simple linear interpolation between the two BRIRs. The high frequencies part of the BRIR is elaborated through three steps: peak detection and matching, calculation of gravity points and warping vectors, and final reconstruction. The first elaboration consists of peaks research in adjacent impulse responses and their matching. Then, an algorithm for the gravity points calculation is applied. The gravity points represent the exact common sample index where the related peaks in both BRIRs must coincide. Once they are found, a non-linear elaboration is applied in order to align the peaks with warping vectors. At last, the final reconstruction of the interpolated BRIR is achieved by the linear interpolation of the two elaborated high frequencies BRIRs.

In this paper, the method presented in [7] is modified in order to improve its performance. In particular, the peak detection and matching algorithm is modified to avoid a mismatching of the peaks of the two impulse responses due to the HRIRs characteristics. This problem occurs when the impulse response has small values of sparsity, as in the case of the HRIR of the contralateral ear. The proposed algorithm is based on the evaluation of the difference between portions of the two impulse responses that must be interpolated. These segments are built around the peaks of the IRs in order to find the peak of the second impulse response that most matches with the peak of the first one.

This new approach has been tested using real room binaural impulse responses, taken from the MIT Media Lab database in [8]. Experimental results have proved the effectiveness of the proposed approach in comparison with the state-of-the-art algorithm. In addition, subjective tests have been carried out with the aim of getting further informations about the perceived audio quality.

The paper is organized as follows: Section 2 presents the problem of the BRIR interpolation. Section 3 describes the reference algorithm that has been implemented. Section 4 focuses on the proposed peak detection and matching algorithm. Section 5 shows the results obtained through objective and subjective tests. Finally, Section 6 contains the conclusions and provides suggestions for possible future developments.

2 Basics and set-up of the interpolation problem

Focusing on the interpolation problem, it aims at calculating the missing BRIRs starting from a dataset of measured impulse responses. In this way, the problem shifts from measure the BRIRs to derive them from a database.

2.1 Linear Interpolation

The simpler way to calculate a new impulse response from a given dataset is the one dimension linear interpolation described by the following equation:

$$h_{\text{int}} = h_1 + (h_2 - h_1)\frac{x_{\text{int}} - x_1}{x_2 - x_1}, \tag{1}$$

where h_1 and h_2 correspond to the measured impulse responses at the points x_1 and x_2 respectively, and h_{int} represents the estimated BRIR at position x_{int}, with $x_1 < x_{\text{int}} < x_2$. It is worth nothing that the calculation of the position x_{int} depends on the measuring points of the employed BRIRs database as it will be described below.

2.2 Employed BRIRs database

In this work, the MIT Media Lab database [8] has been employed. This dataset was recorded with a sampling frequency of 44.1 kHz and the impulse responses were estimated using maximum length pseudo-random binary sequence (MLS). The BRIRs were recorded in 710 different positions and consist of the left and right ear impulse responses from a loudspeaker mounted at $r = 1.4$ m from a KEMAR dummy head microphone. The delay introduced by the playback/recorded have been discarded in order to reduce the size of the dataset and each impulse response has a length of $N = 512$

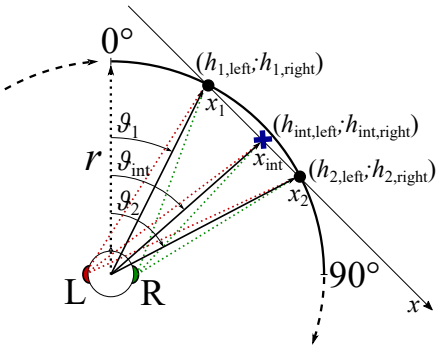

Fig. 1: Interpolation set-up.

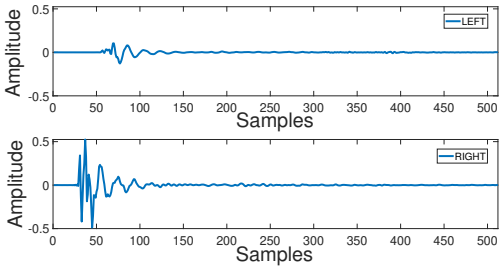

Fig. 2: Binaural Room Impulse Responses of MIT database [8], considering an angle of 95°.

samples. Hence, the measured BRIRs depend only on the azimut ϑ and the elevation φ. In this paper, an elevation of $\varphi = 0°$ is considered, so a 2D set-up can be taken into account, as shown in Figure 1 and a linear approximation must be applied following the Equation (1). Considering the positions x_{int}, x_1 and x_2, they are calculated as follows:

$$x_{\text{int}} = 0, \tag{2}$$

$$x_1 = -r \cdot \left|\sin(\vartheta_1 - \vartheta_{\text{int}})\right|, \tag{3}$$

$$x_2 = r \cdot \left|\sin(\vartheta_2 - \vartheta_{\text{int}})\right|, \tag{4}$$

where ϑ_1 and ϑ_2 are the angles related to the measured BRIRs $h_1$and h_2 respectively, and ϑ_{int} is the angle of the interpolated impulse response h_{int}. The simple application of Equation (1) is not sufficient to obtain a correct interpolated BRIR, so the alignment of impulse responses before the interpolation is required. In case of BRIRs, the alignment of the main peak of the response is a difficult task. This is due to the number of peaks introduced by the HRTF. Figure 2 shows the binaural room responses taken from the MIT database, related to the angle $\vartheta = 95°$. With reference to Figure 1, the source is aligned in front of the right ear of the listener and the impulse response of the right ear has an higher amplitude and the main peak is clearly visible, while in the case of the left ear, the impulse response is delayed compared to the ipsilateral one and has a lower amplitude. Moreover, the left ear impulse response presents two peaks with very similar amplitudes and this may cause a mismatching in the peak detection algorithm, as it will be widely discussed below after a description of the state-of-the-art algorithm used as reference.

3 Reference Algorithm

The implemented algorithm aims at obtaining an interpolated impulse response starting from the knowledge of two measured impulse responses h_1 and h_2. In the case of BRIR, the interpolation process is applied to both impulse responses of the left ear and right ear. The approach presented in [7] is considered as the reference algorithm and Figure 3 shows the total scheme. The two impulse responses h_1 and h_2 are split in two parts: the direct and early reflections (h_{e}) and the reverberant tail (h_{r}). Assuming that the significant information of the BRIRs is condensed only into the direct and early reflections, the two parts are processed separately. The first part is divided in two frequency bands using low-pass (LP) and high-pass (HP) filters. Low frequency responses $h^{\text{L}}_{\text{e}_1}$ and $h^{\text{L}}_{\text{e}_2}$ are linearly interpolated following the Equation (1). Instead, the high frequency signals $h^{\text{H}}_{\text{e}_1}$ and $h^{\text{H}}_{\text{e}_2}$ first pass through a peak detection and alignment algorithm, and then, are linearly interpolated. The two calculated responses $h^{\text{L}}_{\text{e}_{\text{int}}}$ and $h^{\text{H}}_{\text{e}_{\text{int}}}$ are simply summed, obtaining the interpolated early reflections part $h_{\text{e}_{\text{int}}}$. The reverberant tails h_{r_1} and h_{r_2} are interpolated according to the Equation (1). Finally, the interpolated impulse response h_{int} is reconstructed as the concatenation of the direct and early reflections part with the reverberant tail. A detailed description of each part of the diagram of Figure 3 follows.

3.1 BRIR Splitting

The first block of the algorithm consists of the windowing of the BRIRs. The impulse responses are split in two parts: the direct and early reflections h_{e} and the reverberant tail h_{r} as follows:

$$h_{\text{e}} = h[1 : n_{\text{t}}], \tag{5}$$

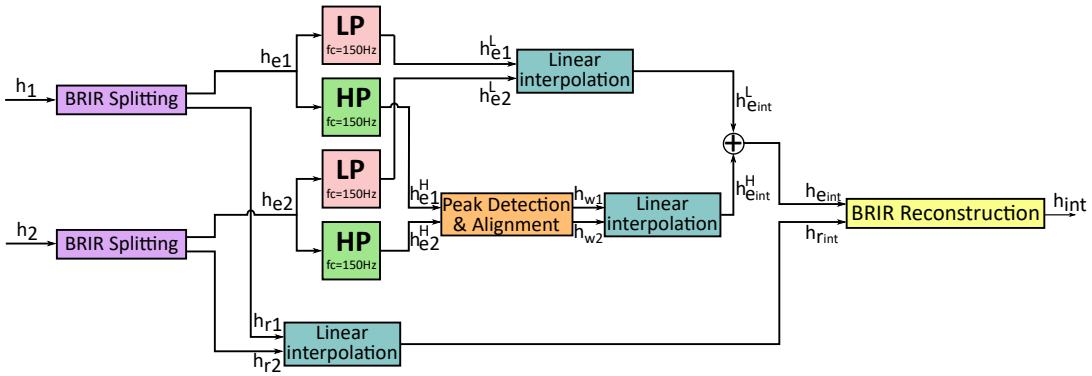

Fig. 3: Overall scheme of the implemented algorithm.

$$h_{\mathrm{r}} = h[n_{\mathrm{t}} + 1 : N], \tag{6}$$

where N is the total number of samples of the impulse response and n_{t} is the transition point between the early and late reflections. Starting from the transition time T_{t} and the sampling frequency f_{s}, the constant n_{t} is computed as follows:

$$n_{\mathrm{t}} = \lfloor T_{\mathrm{t}} \cdot f_{\mathrm{s}} \rceil, \tag{7}$$

where the transition time is set to $T_{\mathrm{t}} = 5$ ms. After this division of the BRIRs, the two parts are treated differently, as explained in the following sections.

3.2 Early reflections interpolation

The impulse responses elaborated by the BRIR splitting are further divided in two subbands using 3^{rd} order low-pass and high-pass Butterworth IIR filters with a cut-off frequency of 150 Hz. The low frequency parts $h^{\mathrm{L}}_{\mathrm{e}_1}$ and $h^{\mathrm{L}}_{\mathrm{e}_2}$ are linearly interpolated in time domain following the Equation (1) and obtaining the response $h^{\mathrm{L}}_{\mathrm{e_{int}}}$. The high frequency signals $h^{\mathrm{H}}_{\mathrm{e}_1}$ and $h^{\mathrm{H}}_{\mathrm{e}_2}$ must be elaborated, as shown in Figure 3, before being linearly interpolated. In particular, these signals have to be aligned in time. To do that, a peak detection and matching algorithm is required as described below.

3.2.1 Peak detection and matching

The peak detection and matching algorithm is engaged only for the high frequency band of the early reflections of the two impulse responses $h^{\mathrm{H}}_{\mathrm{e}_1}$ and $h^{\mathrm{H}}_{\mathrm{e}_2}$. The algorithm consists of the peak detection in both responses with a series of restrictions and the construction of energy blocks that will have to be aligned. Initially, the peaks of the two impulse responses are found imposing a minimum peaks distance of $M = 100$ samples. The second restriction is related to the peaks magnitude. In fact, the peaks below a threshold G_{th} are dropped out because they are considered not significant. The threshold is calculated as follows:

$$G_{\mathrm{th}} = max(h^{\mathrm{H}}_{\mathrm{e}})/\delta, \tag{8}$$

where δ is a constant value that could be changed to reach the best performance of the algorithm. In this case, the value has been selected as $\delta = 10$, after several experimental tests. Finally, N_{p1} and N_{p2} peaks are found in $h^{\mathrm{H}}_{\mathrm{e}_1}$ and $h^{\mathrm{H}}_{\mathrm{e}_2}$ respectively, with $N_{\mathrm{p1}} \neq N_{\mathrm{p2}}$. The next step consists of relating the peaks in $h^{\mathrm{H}}_{\mathrm{e}_1}$ with the ones in $h^{\mathrm{H}}_{\mathrm{e}_2}$ by calculating a matrix where coefficients are computed as follows:

$$C_{i,j} = \frac{1}{(1+\Delta s_{i,j})(1+\Delta p_{i,j})}, \tag{9}$$

where $\Delta s_{i,j}$ and $\Delta p_{i,j}$ are the absolute values of the difference in samples and amplitude, respectively, between the i-th peak of $h^{\mathrm{H}}_{\mathrm{e}_1}$ and the j-th peak of $h^{\mathrm{H}}_{\mathrm{e}_2}$, with $i = 1,...,N_{\mathrm{p1}}$ and $j = 1,...,N_{\mathrm{p2}}$. The coefficients with high values represent potentially related peaks and if one peak of $h^{\mathrm{H}}_{\mathrm{e}_1}$ is related with more than one peak of $h^{\mathrm{H}}_{\mathrm{e}_2}$, only the one with the highest value is chosen and the rest is dropped. As additional constraints, peaks spaced more than 100 samples are discarded. At the end, an equal number N_{p} of peaks in $h^{\mathrm{H}}_{\mathrm{e}_1}$ and $h^{\mathrm{H}}_{\mathrm{e}_2}$ are selected with one-to-one relationship and energy blocks will be made starting from the found peaks. These blocks are made up around the selected peaks and their length is calculated according to the number of peaks.

3.2.2 Alignment and interpolation

The alignment is based on the concept of gravity points which are used to warp the signal in the interpolation position. Therefore, the gravity point describes the exact sample index where both peaks of $h^{\mathrm{H}}_{\mathrm{e}_1}$ and $h^{\mathrm{H}}_{\mathrm{e}_2}$ must coincide. The gravity points (GP) are calculated for each k-th pair of blocks by the following formula:

$$\mathrm{GP}_k = \left\lfloor s_{1,k} + (s_{2,k} - s_{1,k}) \frac{x_{\mathrm{int}} - x_1}{x_2 - x_1} \right\rceil, \qquad (10)$$

where $k = 1, ..., N_{\mathrm{p}}$, with N_{p} number of gravity points (that corresponds also to the number of blocks and the number of peaks), $s_{1,k}$ and $s_{2,k}$ are the samples at which the k-th peak is located in $h^{\mathrm{H}}_{\mathrm{e}_1}$ and $h^{\mathrm{H}}_{\mathrm{e}_2}$, respectively. Once the gravity points are calculated, each k-th block signal can be stretched or compressed consistent with the relative position between GP_k and the k-th peak. Therefore, if the peak is located on the left of the GP, a certain number of points are added by interpolating the first samples of the block, while in the opposite case the samples are discarded. After this warping procedure, the blocks are re-assembled in order to obtain the two aligned impulse responses h_{w_1} and h_{w_2} as shown in Figure 3. These two responses are linearly interpolated following Equation (1) obtaining the impulse response $h^{\mathrm{H}}_{\mathrm{e_{int}}}$. Finally, the early reflections of the interpolated impulse response $h_{\mathrm{e_{int}}}$ is computed as the sum between the high frequency and the low frequency part as follows:

$$h_{\mathrm{e_{int}}} = h^{\mathrm{L}}_{\mathrm{e_{int}}} + h^{\mathrm{H}}_{\mathrm{e_{int}}}. \qquad (11)$$

3.3 Reverberation tail interpolation

The interpolation of the reverberation tails of the two impulse responses h_{r_1} and h_{r_2}, is quite simpler than the elaboration of the early reflections. In fact, this part of the BRIR carries limited information. Therefore, the final reverberation tail $h_{\mathrm{r_{int}}}$ is obtained by the linear interpolation between h_{r_1} and h_{r_2}, according to the Equation (1).

3.4 Final BRIR reconstruction

The final interpolated impulse response h_{int} is obtained as the concatenation of the early reflections with the reverberation tail, as follows:

$$h_{\mathrm{int}} = [h_{\mathrm{e_{int}}}; h_{\mathrm{r_{int}}}]. \qquad (12)$$

The whole interpolation process shown in Figure 3 and widely explained above is applied separately to both left ear and right ear impulse responses. Therefore, two interpolated responses are obtained: $h_{\mathrm{int,left}}$ related to the left ear and $h_{\mathrm{int,right}}$ related to the right ear, as depicted in Figure 1.

4 Proposed peak detection and matching algorithm

The reference peak detection and matching algorithm, described in Section 3.2.1, shows good performance when the impulse responses exhibit a well defined main peak. However, in some cases the BRIRs present several peaks with similar amplitudes, especially when the contralateral ear or the positions behind the listener are considered. For this reason, a different algorithm is here proposed for the peak detection and matching algorithm. The aim of the peak detection and matching algorithm is to obtain N_{p} peaks of h_1 related to N_{p} peaks of h_2. In Figure 4, the proposed algorithm flowchart is shown. Firstly, N_{p} peaks of h_1 are found in the same way of the reference algorithm, as described in Section 3.2.1. In this way, a vector $\mathbf{s}_1 = [s_{1,1}...s_{1,k}...s_{1,N_{\mathrm{p}}}]$, containing the samples at which the peaks of h_1 are located, is obtained. Secondly, N_{p2} temporary peaks of h_2 are looked for with no constraint regarding the peaks distance, but with a threshold defined as $G_{\mathrm{th2}} = G_{\mathrm{th}} \cdot \gamma$, where $\gamma = 0.8$. Therefore, the vector $\mathbf{t}_2 = [t_{2,1}...t_{2,j}...t_{2,N_{\mathrm{p2}}}]$, containing the samples at which the temporary peaks of h_2 are located, is defined. The lengths of the vectors $\mathbf{s}_1$ and $\mathbf{t}_2$ are N_{p} and N_{p2} respectively, with $N_{\mathrm{p2}} > N_{\mathrm{p}}$. Consequently, the proposed algorithm allows to extract N_{p} peaks from $\mathbf{t}_2$ that match with the peaks of h_1. For each k-th peak of h_1, a window is defined as follows:

$$\mathbf{w}_{1,k} = h_1 \left[\left(s_{1,k} - \frac{L_{\mathrm{w}}}{2} + 1 \right) : \left(s_{1,k} + \frac{L_{\mathrm{w}}}{2} \right) \right], \qquad (13)$$

where $k = 1, ..., N_{\mathrm{p}}$, with N_{p} the number of peaks of h_1, L_{w} is the length of the window $\mathbf{w}_{1,k}$ set to $L_{\mathrm{w}} = 100$ and $s_{1,k}$ is the sample corresponding to the k-th peak of h_1. Similarly, for each j-th temporary peak of h_2, the window $\mathbf{w}_{2,j}$ is obtained as follows:

$$\mathbf{w}_{2,j} = h_2 \left[\left(t_{2,j} - \frac{L_{\mathrm{w}}}{2} + 1 \right) : \left(t_{2,j} + \frac{L_{\mathrm{w}}}{2} \right) \right], \qquad (14)$$

where $j = 1, ..., N_{\mathrm{p2}}$, with N_{p2} the number of temporary peaks of h_2, L_{w} is the length of the window $\mathbf{w}_{2,j}$

(that is the same of the window $\mathbf{w}_{1,k}$) and $t_{2,j}$ is the sample corresponding to the j-th temporary peak of h_2. Then, for each k-th peak of h_1, an error vector $\mathbf{E}_k = [e_{k,1}...e_{k,j}...e_{k,N_{p2}}]$ of length N_{p2} is defined. Therefore, for each j-th peak of h_2 an element $e_{k,j}$ of the vector $\mathbf{E}_k$ is calculated as follows:

$$e_{k,j} = \sum_{n=1}^{L_w} \left| \mathbf{w}_{1,k}[n] - \mathbf{w}_{2,j}[n] \right|. \tag{15}$$

The error function $\mathbf{E}_k$ is subject to a minimization process in order to find the index of the peak of h_2 that has the best match with the k-th peak of h_1, i.e.,

$$j_{opt} = \arg\min_j \{\mathbf{E}_k\}. \tag{16}$$

Finally, the k-th peak of h_2 related to the k-th peak of h_1 is derived as follows:

$$s_{2,k} = t_{2,j_{opt}}, \tag{17}$$

where $s_{2,k}$ is the k-th element of the vector $\mathbf{s}_2 = [s_{2,1}...s_{2,k}...s_{2,N_p}]$ that is N_p long and contains the samples of the final peaks of h_2 that have the best match with the N_p peaks of h_1.

To validate the peak detection and matching algorithm, several binaural impulse responses have been considered from the MIT Media Lab database in [8]. Figure 5 shows the resulting matching between peaks of $h^H_{e_1}$ and peaks of $h^H_{e_2}$, comparing the reference algorithm described in Section 3.2.1 with the proposed algorithm. Referring to the Figure 1, three different angles have been considered, i.e., 95°, 135° and 165°. The selection of these angles is justified by the fact that the differences of the two algorithms are more evident for the positions behind the listener. Considering the angle of 95°, the two algorithms have the same behaviour in the case of the right ear (Figures 5(vii) and 5(x)). Instead, for the left ear, a peak mismatching occurs with the reference algorithm as shown in Figure 5(i), while the Figure 5(iv) shows better performances of the proposed algorithm. Regarding to the angle of 135°, the impulse response of the right ear has an higher amplitude but it is more dispersive than the left ear, as shown in the second column of Figure 5. In fact, for the left ear, the two algorithms have the same performance (Figures 5(ii) and 5(v)), while for the right ear the proposed algorithm works better than the reference one, as displayed in Figures 5(viii) and 5(xi), respectively. Similar considerations can be applied for the

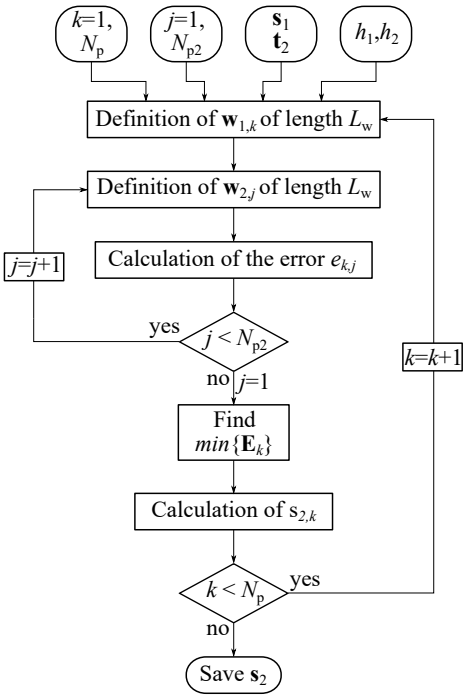

Fig. 4: Flowchart of the proposed peak detection and matching algorithm.

BIRIRs related to the angle of 165°, as shown in the third column of Figure 5. In this case, for the right ear the reference algorithm presents a mismatching for all found peaks (Figure 5(ix)), while the proposed one shows better performances (Figure 5(xii)). Instead, for the left ear the impulse responses are much simple and the algorithms behave in the same way, as shown in Figures 5(iii) and 5(vi).

5 Experimental Results

To evaluate the performance of the proposed algorithm, objective and subjective tests were made. MATLAB has been used to implement and test the algorithm and to make a comparison between the impulse responses interpolated with the reference, the proposed algorithms and the measured ones. Objective and subjective tests were carried out using three different angles, i.e., 95°, 135° and 165°. The impulse responses used to test the algorithm have been measured by MIT Media Labs [8].

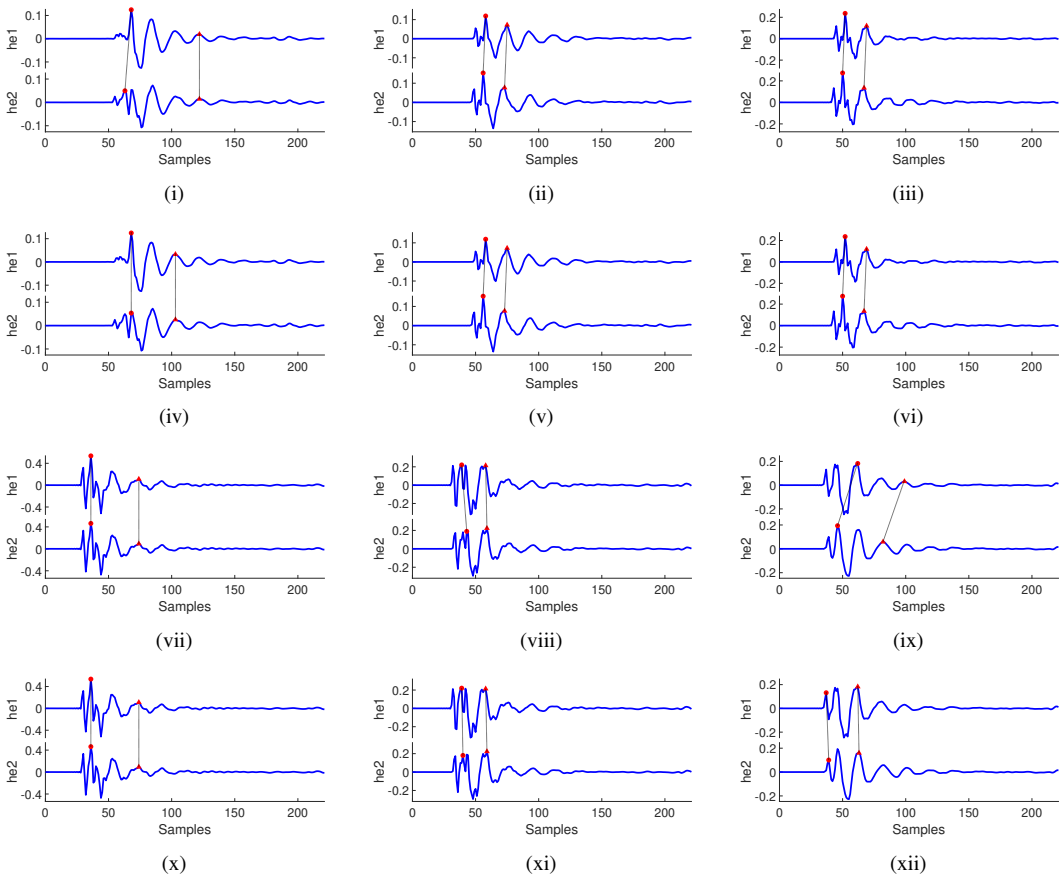

Fig. 5: Peak matching of the reference algorithm (i), (ii), (iii) for the left ear and (vii), (viii), (ix) for the right ear and of the proposed algorithm (iv), (v), (vi) for the left ear and (x), (xi), (xii) for the right ear, considering the angle of 95° (in the first column), 135° (in the second column) and 165° (in the third column).

5.1 Objective Results

The objective results, shown in Figure 6, have been evaluated with a comparison, both in time and frequency domain, between three different impulse responses:

- the measured impulse response drawn by blue line marked with (a),
- the interpolated with the reference algorithm drawn by green line marked with (b),
- the interpolated with the proposed algorithm drawn by the dashed red line marked with (c),

with the aforementioned three different angles and considering left and right ear, separately.
The first column of Figure 6 shows the comparison for the angle of 95°. As mentioned before, the left ear (i.e., the contralateral ear) shown in Figures 6(i) and 6(iv), has a small value of sparsity. For this reason, the peak detection became a difficult task, but the proposed algorithm manages to interpolate an impulse response more similar to the measured one. Instead, the interpolated IR with the reference algorithm is different from the measured one.
In the second column of Figure 6 the results for the angle of 135° are shown. In this case, the contralateral response is well-estimated by both algorithms. Instead, the right ear impulse response is completely wrong when the reference algorithm is applied, while the proposed algorithm makes a perfect IR estimation.
The third column of Figure 6 shows the results considering the angle of 165°. Also in this case, the reference algorithm fails to interpolate the IR correctly when the

right ear is considered.
To sum up, objective results have highlighted better performances of the proposed algorithm than the reference one, especially for the impulse responses in which the main peak is not well defined.

5.2 Subjective Results

In order to evaluate the effectiveness of the algorithm, subjective listening tests were carried out according to ITU-R BS.1284-2 [9]. For the listening tests, the same angles of the objective ones were considered: 95°, 135° and 165°. A listening panel was asked to evaluate some attributes of several sound tracks using a pair of headphones as a playback device. The listening panel was composed of 23 listeners, 14 men and 9 women aged between 20 and 55, in which 11 were expert listeners and 12 not. Each test was based on the listening of three tracks with a length of 20 s, containing the same song filtered with:

A. the measured room impulse responses, taken from the MIT Media Labs database in [8];

B. the interpolated binaural impulse responses obtained from the reference algorithm;

C. the interpolated binaural impulse responses obtained from the proposed algorithm.

The "Unipolar discrete five-grade scale" [9] has been employed as grading scale and for each track, listeners was asked to judge:

- sound quality: the subjective impression of the general quality of the reproduced track;
- transparency: if all details of the performance are clearly perceived in relation to the real position of the source.

Four different songs were employed considering different music genres in order to test the algorithm using different spectral contents. Table 1 shows the selected music genres and the respective soundtracks. The results of the listening tests was elaborated and analyzed by mean with a 95% confidence interval. Figures 7(i) and 7(iv) show the results obtained considering an angle of 95° in terms of sound quality and transparency, respectively. In both attributes, the case of measured impulse response has higher scores, as expected. The reference algorithm presents an higher sound quality only in the case of rock music, while for the other genres the proposed algorithm is quite better in terms of sound quality. The transparency, instead, has similar high scores for all the methods. In particular, the reference algorithm presents higher transparency values for jazz music, while in the case of rock music the proposed one is better. The second column of Figure 7 shows the results of the listening tests considering the angle of 135°. Regarding the sound quality (Figure 7(ii)), the scores reached by the proposed algorithm are very similar to the ones presented by the case of the measured response, while the reference algorithm has a lower sound quality for all the music genres. Also the transparency, shown in Figure 7(v), is better for the proposed algorithm, especially in the case of rock and classical music, even if the difference between the two algorithms is smaller than the sound quality. The third column of Figure 7 shows the results obtained from the listening tests for the angle of 165°. In this case, the difference between the two algorithms is more evident. In fact, the sound quality of the proposed algorithm is much better than the reference one and very similar to the method with the measured impulse response (Figure 7(iii)). Regarding the transparency, Figure 7(vi) exhibits the obtained scores. In this case, the results are comparable to the sound quality ones. In fact, the proposed algorithm exhibits higher scores of transparency than the reference one for all the considered music genres. In conclusions, the impulse responses interpolated with the proposed algorithm seem to perform better than the reference ones in terms of both sound quality and transparency for all the considered music genres. The scores reached by the proposed algorithm are similar to the ones presented by the measured response. In particular, these characteristics are more evident when the angles of 135° and 165° are considered, that is when the sound source is behind the listener head.

6 Conclusions

In this paper a novel method for BRIRs interpolation has been presented. Starting from a previous approach,

Genre	Author	Sound Track
Pop	Daft Punk	Get Lucky
Rock	Pink Floyd	Money
Jazz	Sarah Vaughan	Lullaby of Birdland
Classical	Tchaikovsky	The Nutcracker Op. 71 Act I

Table 1: List of sound tracks used for the listening tests.

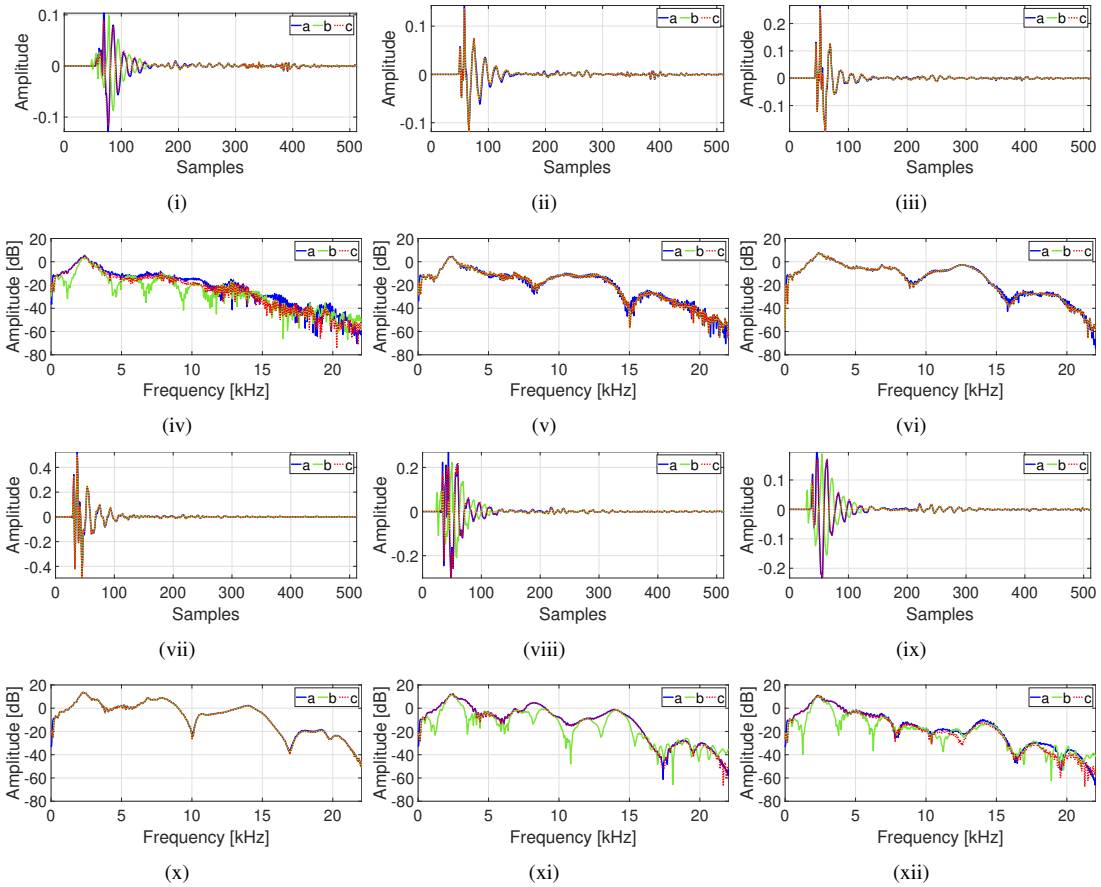

Fig. 6: Comparison between (a) the measured impulse response, (b) the interpolated impulse response following the reference algorithm and (c) the interpolated impulse responses with the proposed algorithm considering (i),(ii),(iii) left IR in time domain, (iv),(v),(vi) left IR in frequency domain, (vii),(viii),(ix) right IR in time domain, (x),(xi),(xii) right IR in frequency domain for the angle 95° (first column), 135° (second column), 165° (third column).

a method based on a new peak detection and matching algorithm has been presented. The resulting interpolated impulse responses are compared with the measured ones and with a method of the state of the art through objective and subjective tests, considering three different positions of the source. Analyzing the objective results, the proposed algorithm has shown better performances than the state-of-the-art algorithm of peak detection and matching, obtaining interpolated impulse responses more similar to the measured ones. Also through the subjective tests the proposed algorithm has ensured better sound quality and transparency, especially when the sound source is behind the listener head. Future works will expand the interpolation process considering more than one dimension and create a new BRIRs database employing different environments with different characteristics.

References

[1] Kearney, G., Masterson, C., Adams, S., and Boland, F., "Approximation of Binaural Room Impulse Responses," in *IET Irish Signals and Systems Conference (ISSC 2009)*, pp. 1–6, 2009.

[2] Mehrotra, S., Chen, W., and Zhang, Z., "Interpolation of combined head and room impulse response

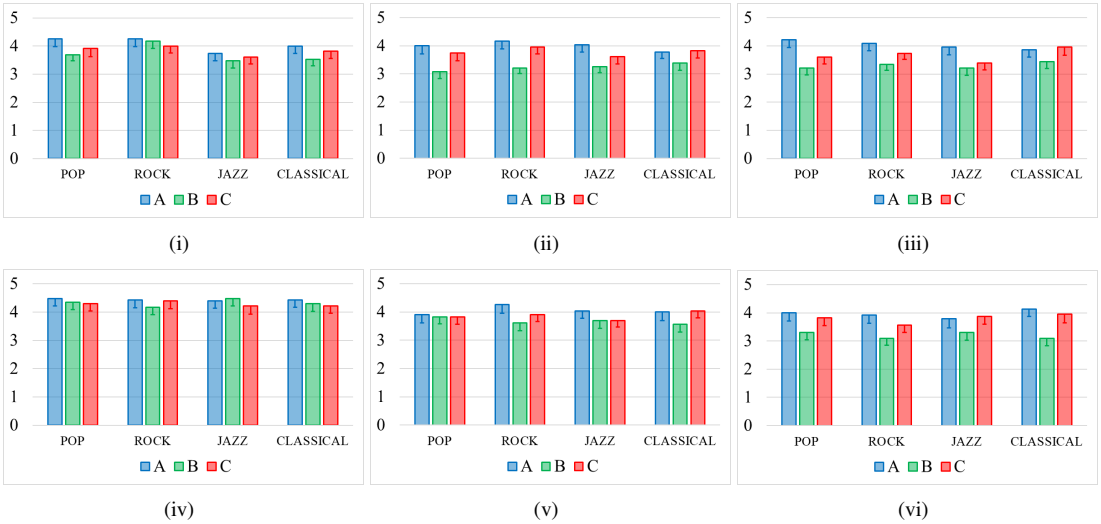

Fig. 7: Results of the listening test considering the angles of 95° (first column), 135° (second column) and 165° (third column), evaluating (i), (ii), (iii) Sound Quality and (iv), (v), (vi) Transparency, with (A) the measured impulse responses, (B) the reference algorithm and (C) the proposed algorithm.

for audio spatialization," in *IEEE 13th International Workshop on Multimedia Signal Processing*, pp. 1–6, 2011.

[3] Haneda, Y., Kaneda, Y., and Kitawaki, N., "Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function," *IEEE Transactions on Speech and Audio Processing*, 7(6), pp. 709–717, 1999.

[4] Kearney, G., Masterson, C., Adams, S., and Boland, F., "Dynamic Time Warping for acoustic response interpolation: Possibilities and limitations," in *17th European Signal Processing Conference*, pp. 705–709, 2009.

[5] Masterson, C., Kearney, G., and Boland, F., "Acoustic Impulse Response Interpolation for Multichannel Systems Using Dynamic Time Warping," in *35th AES Conference: Audio for Games*, 2009.

[6] Hartung, K., Braasch, J., and Sterbing, S. J., "Comparison of Different Methods for the Interpolation of Head-Related Transfer Functions," in *16th AES Conference: Spatial Sound Reproduction*, 1999.

[7] Garcia-Gomez, V. and Lopez, J. J., "Binaural Room Impulse Responses Interpolation for Multimedia Real-Time Applications," in *AES Convention 144*, 2018.

[8] Gardner, B. and Martin, K., "HRTF Measurements of a KEMAR Dummy-Head Microphone," *MIT Media Lab Perceptual Computing Technical Report 280*, 1994 (update on October 13, 2006).

[9] ITU-R BS. 1284-2, "General methods for the subjective assessment of sound quality ," 2019.