2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Advanced Audio Spatializer combined with a Multipoint Equalization System

Stefania Cecchi, Andrea Primavera, Francesco Piazza, Ferruccio Bettarelli, Junfeng Li

***Abstract*— The paper deals with the development of a real time system for the reproduction of an immersive audio field considering the crosstalk cancellation and the room response equalization issues. In particular, the real-time system is composed of two parts: a crosstalk cancellation network and a combined multipoint equalization structure. The former is required in order to have a spatialized audio and it is based on the free-field relationship in order to not introduce a timbre alteration. The latter is used to improve the objective and subjective quality of sound reproduction systems by compensating the room and loudspeakers transfer function. Both steps are based on a-priori analysis of the real environment using real impulse responses measured in different positions. In particular, an offline procedure capable of determining the tuning parameters for the crosstalk network and of deriving the final filters for the equalization structure, is adopted. Several results are presented in order to show the effectiveness of the proposed algorithms considering objective and subjective evaluations and comparing the presented approach with the state of the art.**

I. INTRODUCTION

In the field of immersive audio, there are two main approaches for 3D audio rendering: headphones reproduction and loudspeakers reproduction. The first approach attempts to reproduce through headphones at each eardrum of the listener, the sound pressure of virtual sources using head related transfer functions (HRTFs). In the second case, the binaural signal is delivered to the ears using two or more loudspeakers. In this case, the sound emitted from each loudspeaker is heard by both ears and a network of filters known as crosstalk canceller (CTC) is usually adopted to avoid this problem. Firstly introduced by Bauer in 1961 [1] and patented by Atal in 1966 [2] the crosstalk cancellation technique has received great interest in the academic world and several efforts have been accomplished in the last decades in order to improve the performance of crosstalk canceller taking into account different issues [3]. Indeed, although the fundamental procedure used to perform crosstalk cancellation has not changed since its original statement, several approaches have been presented in literature mainly focusing on improved CTC filter design techniques, and methods for increasing robustness and sweet spot dimension (i.e., the limited area where the crosstalk canceller is active).

Focusing on the CTC filter design, two categories can be considered: the first is represented by those approaches that are based on the use of measured HRTFs, while the second refers to the use of free-field model.

Regarding the first point, in [4] a fast deconvolution approach based on Fast Fourier Transform has been presented to derive the crosstalk matrix, on similar basis in [5] a time domain deconvolution method using both statistic and deterministic Wiener filters is described based on the used of measured HRTFs. Others well-known approaches involve minimax approximation [6] and least mean square (LMS) based solutions [7], [8]. Recent approaches have been based on the control of acoustic energy density [9] or IIR filters designed in frequency warping domain [10]. All these approaches exploit the use of real HRTFs, increasing the model reliability, thus decreasing the generalization of the system: different listeners with different ears could perceived differently the effect of the crosstalk cancellation.

Considering the second category, a system based on free-field model for the impulse response definition has been presented in [11], [12], [13], [14]. In this case, a very simple model that gives an estimation of the system impulse response along x-y coordinate is designed, introducing a low accuracy of the impulse responses but providing good performance in terms of audio quality since low alterations of the timbre are introduced. On similar basis, a heuristic approach named RACE (i.e., Recursive Ambiophonics Crosstalk Elimination) has been presented in [15], representing a simplification of the free-field model approach. In particular, RACE consists in cancelling the crosstalk component at a given contralateral ear by a delayed and attenuated copy of the signal that caused the crosstalk. Differently from the system based on the measured HRTFs, these approaches allow a generalization of the system thus providing better performance in terms of feasibility and effectiveness. However, considering measured HRTFs allows to have an intrinsic equalization of the environment and loudspeakers due to the the crosstalk cancellation matrix structure, while considering the free-field model, the correction of the environment is avoided. Therefore, a room response and loudspeaker equalization is needed in order to support the spatializer system, in the case of free-field modelling approach.

Room response equalization has been studied in theory and applied in practice for improving the objective and subjective quality of sound reproduction systems in different environment [16]. In a room response equalization system, the room transfer function characterizing the path from the sound reproduction system to the listener should be compensated with a suitably designed equalizer, in order to enhance the listening experience avoiding excessive reverberation

Stefania Cecchi, Andrea Primavera and Francesco Piazza are with the Department of Information Engineering (DII), Universitá Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona, Italy (email: a.primavera@univpm.it). Ferruccio Bettarelli is with Leaff Engineering, Via Pastore 10, 60027, Osimo, AN. Junfeng Li is with the Institute of Acoustics, Chinese Academy of Sciences, No 21, BeiSiHuan XiLu, Haidian, Beijing, China

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

and artefacts due to loudspeakers response. In particular, a multipoint equalization that implies the use of room impulse responses measured in more than one position around the listener, has been proved to be an effective approach to enhance the overall listening experience [17], [18], [16], [19], [20].

In this paper, an efficient implementation of an advanced audio spatializer based on acoustic crosstalk cancellation is proposed taking into account the free-field model [11]. In particular, since the proposed crosstalk cancellation algorithm lead to low timbre alterations due to the adopted model, an efficient equalization procedure useful to compensate the loudspeakers and room response transfer functions is here introduced taking into consideration a combined multipoint approach [21], [16], [19]. Then, since the approach of [11] deals with a manual tuning of the system, an automatic procedure is here proposed in order to determine the parameters values employed in the real-time processing. In the end, in order to highlight the effectiveness of the proposed algorithm, objective and subjective comparisons have been done comparing the presented approach with the state of the art.

In Section II the proposed approach is introduced taking into consideration the advanced spatializer developed considering a crosstalk cancellation (Section II-A) algorithm and an innovative equalizer (Section II-B) both based on a-priori analysis of measured impulse responses. The obtained experimental results are reported in Section III. Finally, conclusions and future works are described in Section IV.

II. PROPOSED APPROACH

An efficient implementation of an advanced spatializer is described in this section. Based on an improved version of the free-field crosstalk canceller algorithm [11] combined with a multipoint equalization procedure, the presented approach allows obtaining improvements in terms of robustness, channel separation, and achieved audio quality. More in detail, while free-field crosstalk canceller bring a remarkable spatialization effect introducing low timbre alteration and artefacts, the equalization procedure has been introduced in order to enhance the audio quality experience compensating the loudspeaker and room transfer functions. On this basis, an offline procedure is presented in order to achieve the parameters values employed during the real-time processing (i.e., the crosstalk cancellation matrix and the coefficients used in the equalization procedure). The overall diagram of the proposed technique is visualized in Figure 1 reporting the main operations. First of all, a measurement of the real impulse responses is achieved in different positions, then an offline procedure is performed in order to obtain the final filters for the equalization stage and the tuning parameter for the audio spatializer. These steps are widely described in the following subsections.

A. Audio Spatializer based on IRs analysis

A typical two-loudspeakers listening situation is shown in Figure 2: x_L and x_R are binaural signals sent to the loudspeakers while e_L and e_R are signals perceived at the listener's ears. The system can be described by the following equation

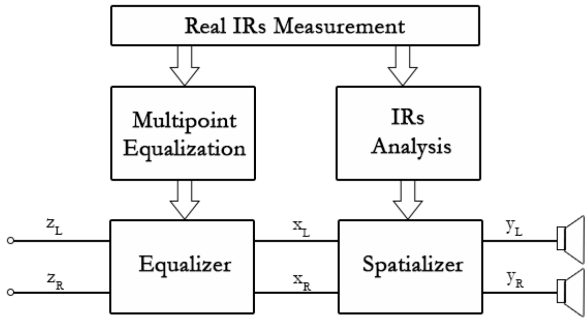


Fig. 1. Overall block scheme of the presented algorithm algorithm.

$$\begin{bmatrix} e_L \\ e_R \end{bmatrix} = H \cdot W \cdot \begin{bmatrix} x_L \\ x_R \end{bmatrix} = \begin{bmatrix} x_L \\ x_R \end{bmatrix} \tag{1}$$

where H is the matrix of the head related transfer functions and W is the crosstalk canceller matrix, defined as follows:

$$H = \begin{bmatrix} H_{LL} & H_{LR} \\ H_{RL} & H_{RR} \end{bmatrix} \tag{2}$$

$$W = \begin{bmatrix} W_{LL} & W_{LR} \\ W_{RL} & W_{RR} \end{bmatrix}. \tag{3}$$

For optimal result, the matrix product between H and W should be the identity matrix, i.e.,

$$W \cdot H = [I] \tag{4}$$

in order to deliver the binaural signal x_L to the left ear and x_R to the right ear. In this way, unwanted crosstalk terms will be eliminated.

However, considering the free-field model [11], [13], [14], the matrix H can be defined as follows:

$$H = \frac{\rho_0}{4\pi} \begin{bmatrix} \frac{1}{l_{LL}} e^{-jkl_{LL}} & \frac{1}{l_{LR}} e^{-jkl_{LR}} \\ \frac{1}{l_{RL}} e^{-jkl_{RL}} & \frac{1}{l_{RR}} e^{-jkl_{RR}} \end{bmatrix}, \tag{5}$$

where $k = (2\pi f)/c_0$ represents the wave number with c_0 the velocity of sound, $\rho_0 = 1.21 kg/m^3$ the density.

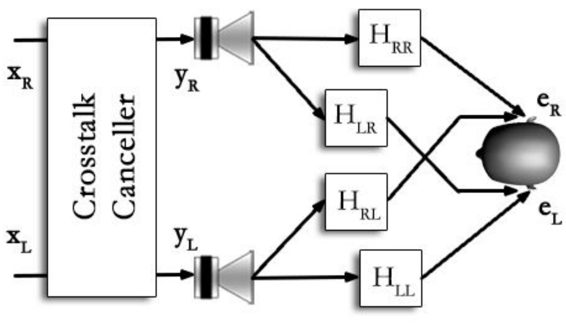


Fig. 2. A block diagram of an audio system that can adjust the sweet spot with relation to the listener position.

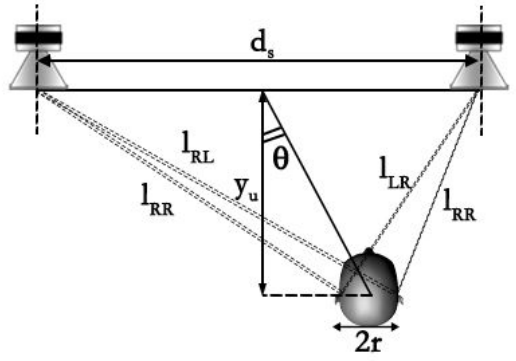


Fig. 3. Geometrical representation of the stereophonic reproduction environment using two loudspeakers.

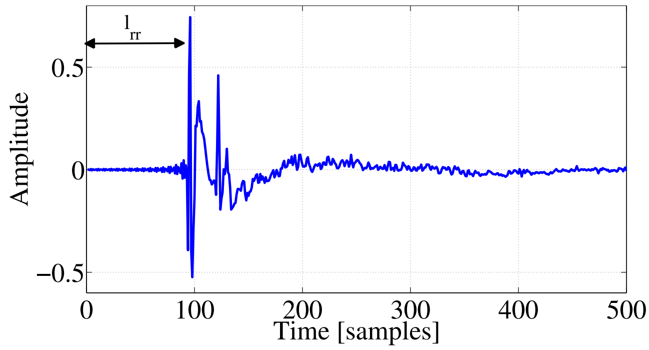


Fig. 4. Evaluation of the distances between the loudspeakers and the listener (l_{RR}) considering the impulse response IR3.

While l_{LL}, l_{LR}, l_{RL}, l_{RR} represent the distances between the loudspeakers and the listener as reported in Figure 3. Considering the approach of [11], [14], these parameters can be calculated as follows:

$$l_{LL} = \sqrt{\left(\frac{d_s}{2} + y_u tan\theta - r\right)^2 + y_u} \quad (6)$$

$$l_{LR} = \sqrt{\left(\frac{d_s}{2} + y_u tan\theta + r\right)^2 + y_u}$$

$$l_{RL} = \sqrt{\left(\frac{d_s}{2} - y_u tan\theta + r\right)^2 + y_u}$$

$$l_{RR} = \sqrt{\left(\frac{d_s}{2} - y_u tan\theta - r\right)^2 + y_u},$$

where y_u denote the distance of the listener from the central axis, d_s is the distance between the two loudspeakers, θ is the look direction of the listener, r is the radius of the listener head that can be set equal to $0.10m$.

However, the application of Equation (6), required a-priori knowledge of the listener's and loudspeakers distances. Therefore in order to avoid this step, an automatic procedure is proposed in order to automatically have the parameters required for the real-time processing operations, starting from an IRs measurement. In particular, focusing on the crosstalk cancellation, the matrix H used in the crosstalk canceller and previously reported in Equation (5) are computed taking into account the arrival time (i.e., the direct path from each loudspeaker to the microphone) obtained from the evaluation of the measured IRs. Different algorithms were tested in order to obtain the best performance in terms of accuracy of the arrival time measure. Among these approaches, the one used is based on a threshold evaluation [22]. In particular, taking into account a given impulse response, the arrival time is calculated considering the time instant in which the impulse response reach the 10% of the maximum peak value as reported in Figure 4. Starting from this analysis, it is possible to obtain the relative distance value l_{LL}, l_{LR}, l_{RL}, l_{RR} useful to compute the HRTF matrix as reported in Equation (5), avoiding the parameters calculation of Equation (6).

At this point, after the definition of Equation (5), the crosstalk cancellation matrix have to be derived. Several approaches [7] [8] [6] have been presented in literature to compute the crosstalk cancellation matrix obtaining the filters value in order to satisfy the condition of Equation (4). Among these, one of the best approach in terms of robustness, separation between the right and left channel, and obtained audio quality, has been proposed by Kirkeby in 1998 and based on a fast deconvolution operation [4]. This approach provides for a frequency regularization in order to control ill-conditioning in the matrix inversion. In particular, exploiting this approach, the crosstalk cancellation matrix is obtained as reported in the following equation:

$$W = H^{-1} = \frac{1}{D}\begin{bmatrix} H_{RR} & -H_{LR} \\ -H_{RL} & H_{LL} \end{bmatrix}, \quad (7)$$

where

$$D = H_{LL}H_{RR} - H_{RL}H_{LR} \quad (8)$$

and $H_{LL}, H_{LR}, H_{RL}, H_{RR}$ are the HRTF coefficients obtained from impulse response measurements, and the inverse operation $\frac{1}{D}$ is performed using the fast deconvolution approach with regularization .

B. Equalizer based on Multipoint room response equalization

The proposed approach is based on a multipoint techniques [16] capable to enlarge the listening sweet spot and takes into consideration a quasi-anechoic approach [21], that is

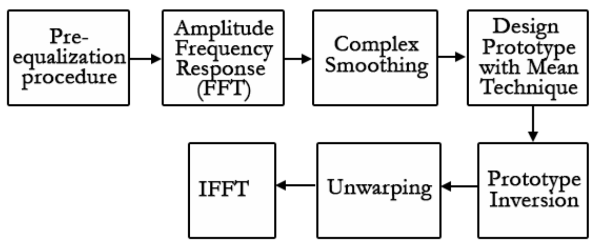


Fig. 5. Room response equalization diagram.

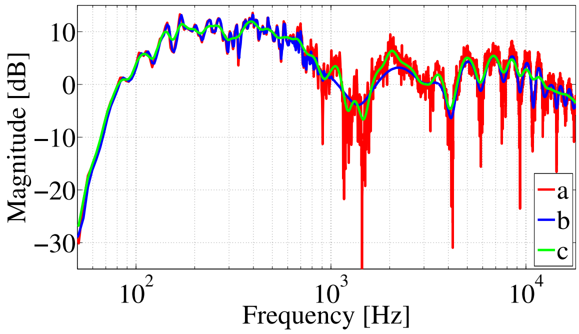


Fig. 6. Comparison between (a) the measured IR and (b) the pre-equalized IR and (c) one third octave smoothed IR.

capable to produce also a general equalization of the used loudspeakers. The novel prototype function is derived from the combination of quasi-anechoic IRs, derived from a gated version (up to the first reflection) of the responses, with the IRs recorded in the real environment.

Figure 5 shows each step of the equalization approach. In particular, given a set of impulse responses measured in the to-be-equalized zone, a pre-equalization procedure is applied considering the quasi-anechoic IR spectrum for frequency greater than the transition frequency, and the original (un-gated) IR spectrum below the transition frequency. This is done in order to equalize the direct sound only in mid-high frequency range, while applying full equalization in the modal frequency range. Figure 6 shows the behaviour of a measured impulse response, and the obtained impulse response following the proposed procedure in comparison with the third octave complex smoothing, considering a transition frequency of $700Hz$.

Then, starting from the magnitude spectra of the IRs, a smoothing operation is applied: this method simulates a well-known property of the auditory system which presents a poorer frequency resolution at higher frequencies. In this way, it is possible to consider a non-uniform resolution, which decreases with increasing frequency, to obtain a less precise equalization at higher frequency resulting in a broader equalized zone.

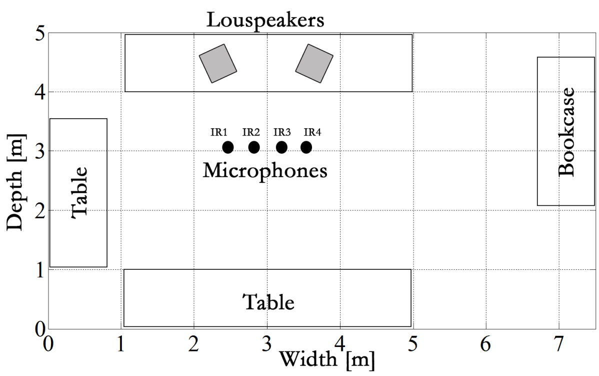


Fig. 7. Microphones and loudspeakers position for the experimental results.

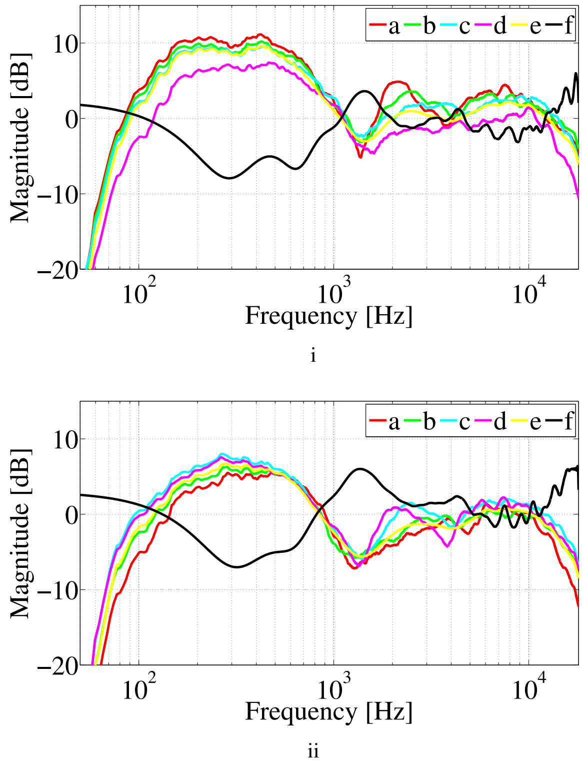


Fig. 8. Magnitude frequency response of (a), (b), (c), (d) smoothed impulse response, (e) prototype function, (f) equalizer considering (i) left and (ii) right channel.

At this point, a representative response of the considered acoustic environment is derived taking into account all smoothed IRs. The prototype frequency response is obtained using an arithmetic mean of the smoothed frequency responses, as follows

$$H_p\left(k\right) = \frac{1}{M}\sum_{i=1}^{M} H_i\left(k\right) \tag{9}$$

with $k = 0, \cdots, K-1$. Some tests have been performed considering the Fuzzy c-means approach [18] for the prototype extraction, however the same results of the mean approach, in terms of prototype definition, have been obtained considering the aforementioned pre-equalization approach. This could mean that the pre-equalization method is already capable of extracting the main characteristics of the analyzed environment. However, future works will be oriented to perform a deep investigation taking into account several microphones positions, in order to verify the enlargement of the equalized zone. Then, the inverse model of the prototype function is obtained using a frequency deconvolution with regularization [4] that is capable to avoid excessive gains, especially at high frequencies. It is applied to the prototype as follows:

$$H_{inv}\left(k\right) = \frac{H_p^*\left(k\right)}{\left|H_p\left(k\right)\right|^2 + \beta} \tag{10}$$

where β is the regularization factor, $k = 0, \cdots, K-1$ and $(^*)$ indicated the conjugate of the filter. For the experimental

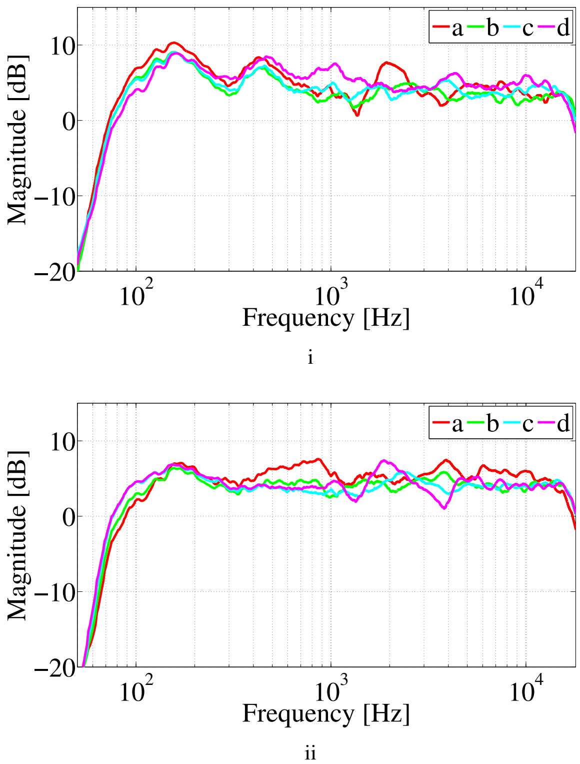


Fig. 9. Magnitude frequency response of (a), (b), (c), (d) impulse responses after equalization procedure considering (i) left and (ii) right channel.

results, a regularization factor with value 0.00001 is considered. The computational complexity of the fast deconvolution method is essentially that of the inverse FFT which is an $O(K \log K)$ algorithm [4]. The corresponding time domain filter is truncated by an appropriate window function with a length L, in order to have a more compact representation. Its length can be limited due to the beneficial effect of the pre-equalization procedure.

III. EXPERIMENTAL RESULTS

Several tests have been carried out in order to evaluate the effectiveness of the presented algorithm taking into account a real environment, and also providing comparison with the existing techniques in terms of objective and subjective measures. Loudspeakers and microphones positions are shown in Figure 7 together with room size: the distance between each microphone has been set to 17 cm, taking into consideration the median distance between human ears. The distance between the loudspeakers has been set to 60 cm with an angle of 30°, representing a typical listening condition. Measurements have been performed using a professional ASIO sound card and microphones with an omnidirectional response. The loudspeakers were a Genelec 6010A with free field frequency response 74 Hz - 18 kHz. A personal computer running NU-Tech platform has been used to manage all I/Os [23]. The impulse responses have been derived using a logarithmic sweep signal excitation [19] at 48 kHz sampling frequency.

A. *Objective Analysis*

First of all, the performance of the equalization has been proved considering the aforementioned scenario. Figure 8 shows the magnitude responses of the identified IRs at different microphone positions, of the prototype function, and of the equalizer, considering the left and right channel. It is evident that the prototype follows the behaviour of the identified room impulse responses, in both cases. Figure 9 depicts IR magnitude spectra resulting after equalization procedures at the different positions using the proposed equalizer. The equalization procedure should ideally lead to a flat curve around zero. Obviously, this cannot be achieved since the equalizer is derived from a set of IRs and considering a quasi-anechoic procedure. However, Figure 9 shows good results with the proposed equalizer for both channels, with a considerable uniformity, especially with respect to Figure 8.

Considering the audio spatializer, first of all an analysis of the impulse responses has been considered in order to calculate the relative distance value l_{LL}, l_{LR}, l_{RL}, l_{RR} useful to compute the HRTF matrix as reported in Equation (5). In particular, Figure 4 represents an example for the calculation of one parameter, i.e., l_{RR} using IR2 of Figure 7, since IR2 and IR3 represent the ideal position of the listener's head. With relation to the crosstalk cancellation procedure, Figure 10i shows the function $1/D$ derived using Equation (8), while Figure 10ii shows the frequency response of $1/D$ comparing the results obtained for the proposed approach and considering real IR. It is worth notice that time

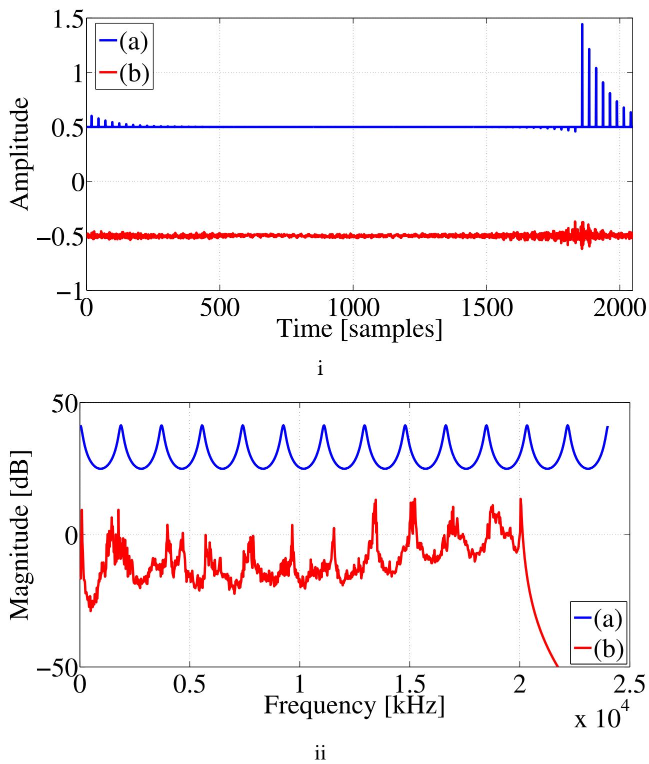


Fig. 10. 1/D function derived using the Kirkeby inversion [4] for (a) free-field model [11] and (b) real IRs: (i) time response, (ii) frequency response.

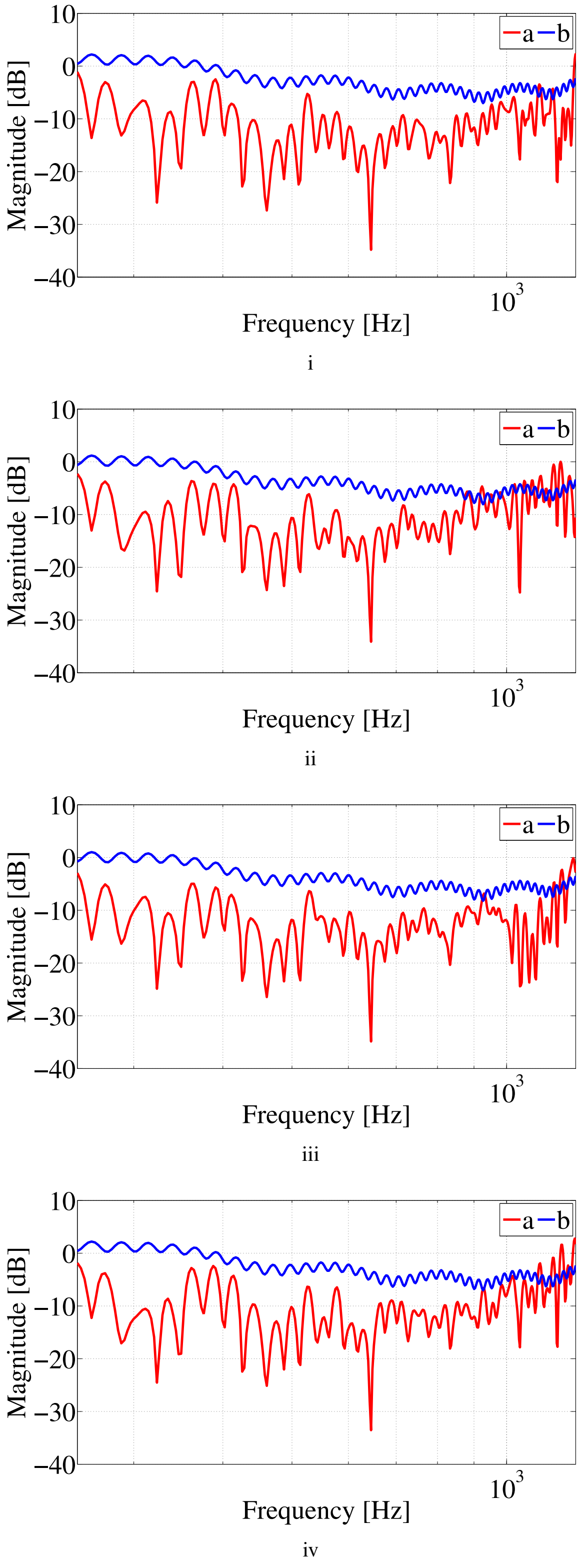


Fig. 11. Magnitude response of the crosstalk cancellation matrix W considering (a) the real IR and (b) the proposed approach in the frequency range of application: (i) W_{ll}, (ii) W_{lr}, (iii) W_{rl} and (iv) W_{rr}.

and frequency responses have been shifted on the y-axis to enhance the readability. Figure 11 shows the coefficient of the crosstalk cancellation matrix W in the the frequency domain considering the proposed approach based on free-field model and the filter derived using the measured impulse response. They represent the filters used in the spatializer block for the subjective listening tests session.

TABLE I
LIST OF ATTRIBUTES USED FOR THE LISTENING TESTS.

Attributes	Description
Main impression	The integrity of the total sound image and the interaction between the other parameters
Spatial impression	The performance appears to take place in an appropriate spatial environment

TABLE II
LIST OF SOUND TRACKS USED FOR THE LISTENING TESTS.

Genre	Author	Sound Track
Pop	M.Biondi	I Know It's Over
Jazz	Freak Power	Turn on, tune in, cop out
Country	J.Cash	Cry, Cry, Cry
Soul	Seal	A change is gonna come

B. *Subjective Analysis*

To assess the overall audio quality perception of the proposed procedure, subjective listening tests according to the ITU-R BS.1284-1 [24] were performed. This recommendation provides guidelines for the general assessment of sound quality. Following the recommendations, the subjects involved in the listening tests were 10 expert listeners (8 males and 2 females, ages from 21 to 35) with a technical background in acoustics. The subjective tests have been performed taking into account the spatialization and audio enhancement of a two channel system, positioned in a table in front of the listener as reported in Figure 7. Before the listening test, a training test was presented to the listeners to familiarize them with the test procedure, the test materials, and the test environment, as suggested in [24].

For the test session, three signals have been considered

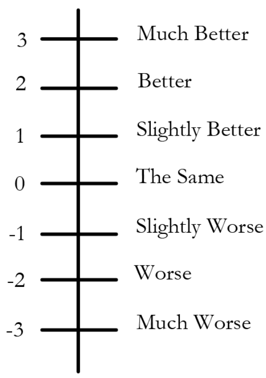
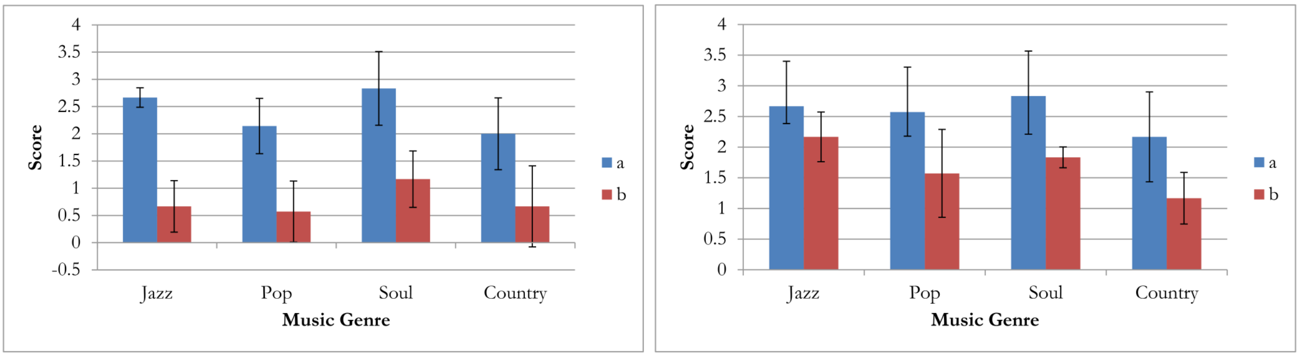


Fig. 12. Quality scale of [24] used for the listening tests.



i ii

Fig. 13. Results of listening test using the mean value and a confidence intervals calculated taking into consideration the standard deviation and a significance level of 0.05: (i) Main impression, (ii) Spatial impression for (a) proposed approach and (b) the approach of [4].

and compared with the original signal without crosstalk cancellation (i.e., the reference signal): a signal processed using the proposed approach, a signal obtained using the crosstalk cancellation procedure described in [4] using the measured IR, and a hidden reference signal to test the subject skill. For the listening test, the subjects were instructed to make a paired comparison between the original sound track (i.e., the reference signal) and one of the processed stimuli under test, scoring the spatial impression and the main impression, as described in Table I. Figure 14 shows the graphical user interface used in the listening test session. According to [24], the score was given using the continuous quality scale divided into six equal intervals (from +3 to -3), where negative values indicate a worsening of performance, positive values an improvement, and zero indicates that no audio difference is perceived, as reported in Figure 12. Each stimulus was 20 s long, as suggested in the recommendation, and the listeners had the possibility to repeat the presentation. The reference signals were chosen considering different music genres in order to test the proposed approach using different spectral contents. Table II shows the selected music genres and the sound tracks used as reference signal.

As suggested in [24], the subjective data have been processed to derive the mean values and the confidence intervals. Figure 13 shows the obtained results in terms of statistical evaluation taking into consideration all the music genre used in the listening test procedure. In particular, each vertical colored bar corresponds mean grade values while the confidence intervals are given by vertical black lines.

From Figure 13i, it is evident that the proposed approach is capable of obtaining good results, achieving values between "better" and "much better" while the approach used for the comparison obtains limited positive results placed below "slightly better" interval. This fact is more evident in the case of soul music genre, where all the listener have reported very good comments.

Considering just the spatial impression as reported in Figure 13ii, it evident that both proposed approach and approach based on the real impulses responses are capable of giving a better spatial impression to the reproduced sound. While the proposed approach shows the better performance considering the soul music genre, the other approach has reached good results for the jazz song. This phenomenon could be explained from the fact that the approach based on the real IR introduces a more strong spatialization, that are perceived annoying when the song is characterized from a high level of spectral content.

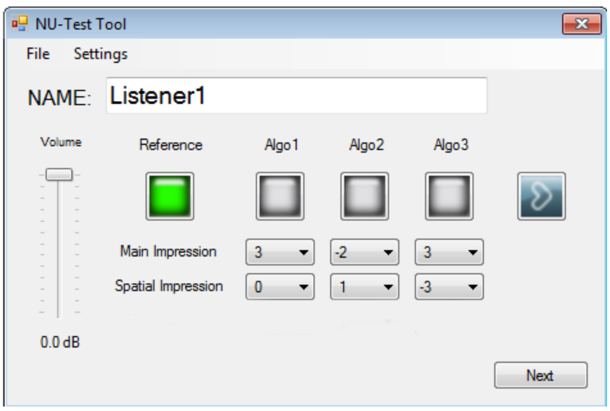


Fig. 14. Graphical User Interface used for the listening tests.

IV. CONCLUSION

In this paper an efficient real time system for the reproduction of an immersive audio system considering the crosstalk cancellation and the room response equalization issues has been presented. It has been realized using a crosstalk canceller based on the free-field relationship with an automatic calculation of the setting parameters combined with a multipoint equalization procedure for compensating the loudspeaker and room transfer functions of the real environment.

Several tests were performed and some results were presented comparing our approach with the state of the art. The experiments have shown that the proposed approach is capable of achieving better results in term of stability and perceived audio quality both in terms of objective and subjective measures. In particular, since the proposed approach derives from a combination of two methods, particular attention has been given to the subjective results. In this context, the proposed approach has shown that is capable of improving the overall audio quality thanks to the multipoint equalization procedure and the spatialization perception due to the crosstalk cancellation approach.

Future works will be focused on the refinement of real-time implementation also taking into account a DSP-based implementation and also on a deeper subjective investigation taking into consideration other type of loudspeakers and environments to show the effectiveness of the approach in creating an extended sweet spot.

References

[1] B. B. Bauer, "Stereophonic Earphones and Binaural Loudspeakers," *J. Audio Eng. Soc.*, vol. 9, no. 2, pp. 148–151, Apr. 1961.

[2] B. S. Atal and M. R. Schroeder, "Apparent Sound Source Translator," US Patent 3236949, 1966.

[3] B. Masiero, J. Fels, and M. Vorlander, "Review of the crosstalk cancellation filter technique," in *Proc. International Conference on Spatial Audio*, Detmold, GE, Nov 2011.

[4] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast Deconvolution of Multichannel Systems using Regularization," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 2, pp. 189–194, Mar. 1998.

[5] S. M. Kim and S. Wang, "A Wiener Filter Approach to the Binaural Reproduction of Stereo Sound," *J. Acoust. Soc. Amer.*, vol. 114, no. 6, pp. 3179–3188, Dec. 2003.

[6] H. Rao, V. Mathews, and Y.-C. Park, "Inverse Filter Design Using Minimax Approximation Techniques for 3-D Audio," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, May 2006, pp. 14–19.

[7] L. Lim and C. Kyriakakis, "Multirate Adaptive Filtering for Immersive Audio," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, Salt Lake City, UT, USA, May 2001, pp. 3357–3360.

[8] S. Cecchi, L. Palestini, P. Peretti, F. Piazza, and F. Bettarelli, "Sub-band Adaptive Crosstalk Cancellation: a novel Approach for Immersive Audio," in *Proc. 124th Audio Engineering Society Convention*, Amsterdam, The Netherlands, May 2008.

[9] J. Lee, Y. Park, and D. Youn, "Robust Crosstalk cancellation based on energy-based control," in *Proc. of AES 34th International Conference*, Aug. 2008.

[10] J.-w. Jeong, J.-T. Kim, J. Lee, Y.-c. Park, and D.-h. Youn, "Design and Implementation of IIR Crosstalk Cancellation Filters Approximating Frequency Warping," in *Proc. 118th Audio Engineering Society Convention*, Barcelona, Spain, May 2005.

[11] K.-S. Lee and S. pil Lee, "A real-time audio system for adjusting the sweet spot to the listener's position," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, pp. 835–843, 2010.

[12] J. Rose, P. Nelson, B. Rafaely, and T. Takeuchi, "Sweet spot size of virtual acoustic imaging systems at asymmetric listener locations," *J. Acoust. Soc. Am.* , vol. 112, pp. 1992–2002, Nov 2002.

[13] M. R. Bai and C.-C. Lee, "Comprehensive analysis of loudspeaker span effects on crosstalk cancellation in spatial sound reproduction," in *Proc. 120th Audio Engineering Society Convention*, May 2006.

[14] D. Ward and G. Elko, "A new robust system for 3d audio using loudspeakers," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, 2000, pp. II781–II784 vol.2.

[15] R. Glasgal, "360 localization via 4.x RACE processing," in *Proc. 123rd Audio Engineering Society Convention*, New York, NY, USA, Oct. 2007.

[16] A. Carini, S. Cecchi, I. Omiciuolo, F. Piazza, and G. L. Sicuranza, "Multiple Position Room Response Equalization in Frequency Domain," *IEEE Trans. Audio, Speech and Language Processing*, vol. 20, no. 1, pp. 122–135, Jan. 2012.

[17] S. Bharitkar and C. Kyriakakis, *Immersive Audio Signal Processing*. Springer, 2006.

[18] I. Omiciuolo, A. Carini, and L. Sicuranza, "Multiple Position Room Response Equalization with Frequency Domain Fuzzy c-Means Prototype Design," in *Proc. Workshop on Acoustic Echo and Noise Control*, Seattle, WA, USA, Sep. 2008.

[19] S. Cecchi, L. Palestini, P. Peretti, L. Romoli, F. Piazza, and A. Carini, "Evaluation of a Multipoint Equalization System based on Impulse Response Prototype Extraction," *J. Audio Eng. Soc.*, vol. 59, no. 3, pp. 110–123, Mar. 2011.

[20] S. Cecchi, L. Romoli, F. Piazza, and A. Carini, "Multipoint Room Response Equalization with Group Delay Compensation: Subjective Listening Tests," in *Proc. 7th Int'l Symposium on Image and Signal Processing and Analysis*, Dubrovnik, Croatia, Sep. 2011.

[21] B. Bank, "Combined quasi-anechoic and in-room equalization of loudspeaker responses," in *Proc. 134th Audio Engineering Society Convention*, May 2013.

[22] S. Hwang and Y. Park, "Time delay estimation from HRTFs and HRIRs," in *The 8th International Conference on Motion and Vibration Control*, Daejeon, Korea, Aug 2006.

[23] A. Lattanzi, F. Bettarelli, and S. Cecchi, "NU-Tech: The Entry Tool of the hArtes Toolchain for Algorithms Design," in *Proc. 124th Audio Engineering Society Convention*, Amsterdam, The Netherlands, May 2008, pp. 1–8.

[24] "General methods for the subjective assessment of sound quality," ITU-R BS. 1284-1, Geneva, 2003.