DOI: 10.61782/fa.2023.0698

SET UP AND PRELIMINARY VALIDATION OF A SMALL SPATIAL SOUND REPRODUCTION SYSTEM FOR CLINICAL PURPOSES

Angerla Guastamacchia[1*] **Michele Ebri**[2] **Andrea Bottega**[1]
Enrico Armelloni[2] **Angelo Farina**[2] **Giuseppina Emma Puglisi**[1]
Fabrizio Riente[3] **Louena Shtrepi**[1] **Marco Carlo Masoero**[1]
Arianna Astolfi[1]

[1] Department of Energy, Politecnico di Torino, Italy
[2] Department of Engineering and Architecture, University of Parma, Italy
[3] Department of Electronics and Telecommunication, Politecnico di Torino, Italy

ABSTRACT

Hearing research and audio technology have met to cope with hearing impairment issues under multiple aspects. Among them, spatial sound reproduction systems have been used for both clinical and research purposes to optimize signal processing algorithms of Hearing Aids (HAs) and for the assessment of hearing loss under complex acoustic conditions. Furthermore, spatial sound reproduction systems are also well suited for the administration of listening tests properly designed to optimize HAs fittings, for which ecological validity is crucial to achieve effective hearing improvement in daily life. Based on well-grounded 3D sound systems, this work discusses the procedure of installation, signal network set-up and evaluation of a cost-effective Virtual Sound Environment (VSE) reproduction system that is meant to be replicated and used indoors in small settings for clinical purposes. The system, aimed at reproducing sound fields starting from 3rd-order ambisonics encodings, is based on a spherical array of 16 commercial 2-way active loudspeakers installed inside of a small acoustically dampened room of 35.5 m^3. Results of this work can be summarized as follows: (i) a small spatial sound reproduction system was tuned and (ii) a preliminary investigation of the accuracy of the reproduced VSEs compared to the real environments was performed.

[*] ***Corresponding author***: angela.guastamacchia@polito.it.

Keywords: *virtual sound environment, ecological listening tests, 3rd-order ambisonics.*

1. INTRODUCTION

Despite advancements in hearing device technologies, they still result inefficient when used in ordinary acoustically competitive sceneries, like conversations in noisy reverberant environments with multiple interfering talkers [1], [2]. Research has found a major issue in the current procedures employed to test hearing devices effectiveness and predict, during the fitting phase, the real-life benefit brought to users. On the one hand, clinical practice laboratory procedures lack ecological validity, relying on listening tests in quiet environments or with stationary noise using plain loudspeaker setups [3]. On the other hand, in-field procedures lack reproducibility, making it difficult to detect slight changes in individuals' hearing abilities [4]. Thus, aiming at retaining control over the tested acoustical stimuli and conducting listening tests matching real-life auditory demands, audiology research has increasingly embraced the use of spatial sound playback systems, which reproduce Virtual Sound Environments (VSEs) that mimic everyday listening payloads involving multiple sound sources from different directions in reverberant settings [5]. In particular, based on several multi-loudspeaker array arrangements, VSE reproduction systems can rely on different reproduction techniques to auralize a target sound field around the listener, in the center of the array, through either a physical (High-Order Ambisonics (HOA) [4], [6]–[8], Wave Field Synthesis (WFS) [9]) or a perceptual

(Vector-Base Amplitude Panning (VBAP) [4], [7], [8], Nearest Loudspeaker panning (NLS) [5], [6], [10]) sound field reconstruction in a defined region. Moreover, the sound field auralization through arrays of loudspeakers instead of headphones also allows considering real-world factors influencing auditory perception, such as the listener's head and its movements affecting sound localization perception [6], the listener's physical presence in the rendered sound field [3], and, finally, the actual impact of hearing device usage on the resulting degree of hearing loss [11]. However, most of the existing auditory research laboratories [8,5,[12] are based on very expensive and complex loudspeaker systems, often installed inside wide anechoic chambers, which hardly translate into setups that can be easily replicated in standard clinical environments, where both indoor spaces and costs are usually constrained.

Thus, the proposed work focuses on the realization of a simple loudspeaker-based VSE, which can be rapidly assembled inside clinical settings and easily handled by non-experts, suitable to perform more efficient listening tests, particularly aimed at running ecological speech intelligibility tests. Specifically, HOA is chosen as the spatial audio reproduction technique since (i) it aims at optimizing the sound field reconstruction in a small specific region of the space (sweet spot), using a reasonable number of loudspeakers (depending on the order number, directly linked to the reconstruction accuracy), instead of a larger area that would require a higher loudspeaker number and would be overly dimensioned for the clinical application where only one listener at a time has to be tested, (ii) the physical reconstruction of the sound field instead of the perceptual one best suits the usage of hearing devices, being their processing considerably different from the human auditory one [13], (iii) it involves a scalable sound field encoding that is fully decoupled from the decoding, entailing a simple reproduction of the sound field, independent on the loudspeakers array arrangement. According to the [13] study, 7OA should ensure an accurate sound field reconstruction for a frequency range adequate for testing hearing aids when realistic reverberant environments are auralized; however, it would require optimally spatially arranged 64 loudspeakers, which would be a too expensive and complex system to be easily adapted for clinical settings. Moreover, even if hybrid techniques are used to achieve the same accuracy by rendering with decreasing ambisonic order the early reflections and the reverberant tail as in [6], 7OA would still be required to auralize the direct sound field, entailing the same high number of loudspeakers. Thus, to standardize the treatment of the impulse responses, a non-hybrid auralization system that equally treats direct sound, reflections, and reverberation results more convenient than the hybrid approach, independently of the actual ambisonic order employed. Nevertheless, in [13] the authors themselves remark that the study quantified the HOA errors compared with the ideal real-life sound field, which could be a too strict metric, especially thinking of the clinical tests currently used to assess how hearing device performances translate in real-life auditory benefit for hearing-impaired individuals. Indeed, even with a low-order HOA system, the boosted realism and complexity in the reproduced test acoustical stimuli may be more than enough to take a step forward in the fitting practice of hearing devices.

Therefore, the present work reports the installation and tuning procedure of a simple 3OA reproduction system inside a small room, further proposing a preliminary evaluation of the system accuracy based on the comparison of standard room acoustics parameters values measured in the auralized environment compared with the values measured in the corresponding real environment. In conclusion, the paper aims to be a guideline in the case of cost-effective clinical lab for the assessment of hearing-impairment and effects of hearing-aiding. The proposed VSE reproduction system is meant to be replicated full-scale in clinics. To the authors' knowledge only two papers detailed such a guideline, which are [5] and [7], but they report more complicated and expensive set-ups that cannot be easily implemented full-scale.

2. METHODOLOGY

2.1 Installation of the VSE reproduction system

The VSE reproduction system is installed inside an acoustically damped room, referred to as the Audio Space Lab, of the Energy Department of the Polytechnic of Turin. The small listening room of 35.36 m^3 volume (length of 5.45 m, width of 2.67 m, height of 2.43 m) is located on the first floor, overlooking an inner courtyard, and was sound treated previously the sound system installation in accordance with the criteria stated in the standard ITU-R BS.1116-3 recommendation for the subjective assessment of small impairments in audio systems [14]. In particular, the room is characterized by reverberation time values around 0.17 s that fall within the optimal range for all octave band frequencies from 0.25 to 4 kHz and by background noise level values measured in the listening position that fall between NR 10 and NR 15 for frequencies up to 1 kHz and values lower than 16 dB for the highest octave bands. Choosing to install the loudspeaker-based

VSE inside an acoustically damped room instead of an anechoic chamber had a twofold reason [7]. First, finding an available room large enough to develop an anechoic chamber in standard clinical environments is a challenging task that further entails high costs. Then, to some extent, a few reflections could also be helpful to mask reproduction inaccuracies when the system goal falls outside of exact objective evaluations.

The Audio Space Lab was set up with an ambisonics 16.2 system, relying on commercial hardware and mounting construction easy to find and a very low-cost software DAW. The system consists of a spherical array of 16 Genelec 8030B 2-way active monitors, used as sound sources from the 90 Hz to 20 kHz frequency range, and 2 Genelec 8351A 3-way active monitors used for the lowest frequencies, i.e., from 30 to 90 Hz, which are actually wasted given the narrow frequency range for which they are used, so they could be substitute with more appropriate and cheaper speakers, as the Genelec 7040A subwoofers in future implementations. The 16 speakers are distributed on a sphere of 120 cm radius, having the center, i.e., the sweep spot, at 121.5 cm height from the floor. Specifically, the speakers were arranged on 3 rings at different elevations, as shown in Fig.1, i.e., one at -45° elevation angle from the center of the sphere, then one at 0° and one at +45° elevation angle. Fig. 2(b) illustrates the vertical section of the Audio Space Lab, where the rings at the 3 different heights are visible.

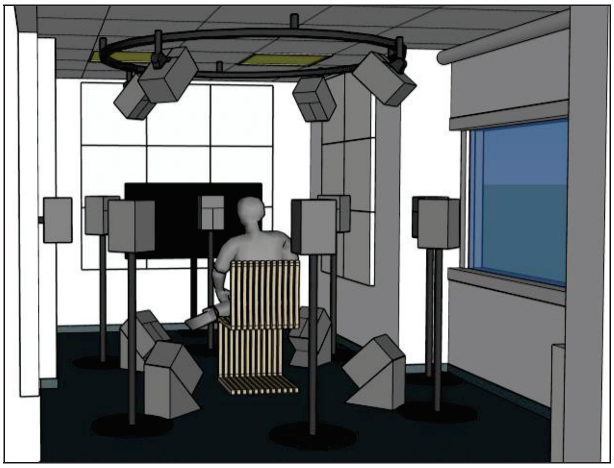
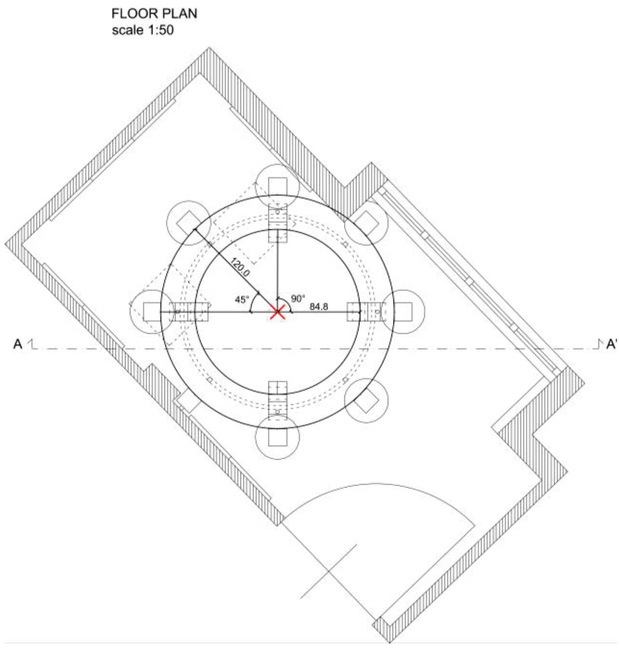
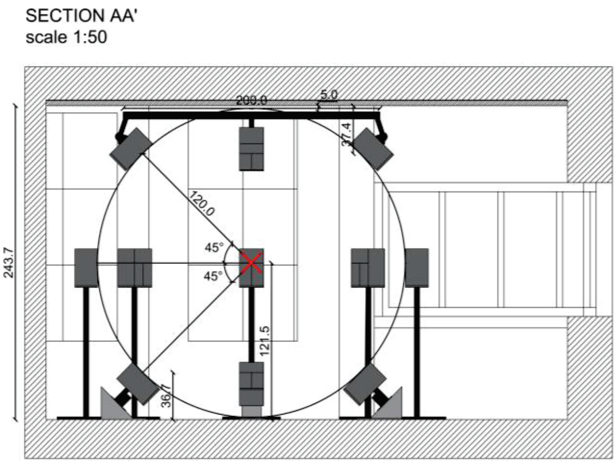


Figure 1. 3D model of the 3OA reproduction system inside the Audio Space Lab.



(a)



(b)

Figure 2. Audio Space Lab floor plan with projections of the speakers composing the 3OA reproduction system Fig. 2(a) and vertical section of the system Fig. 1(b).

The middle ring has 8 of the 16 speakers equally spaced of 45° azimuth angle, so as to maximize the spatial definition in the horizontal listening plane, i.e., the plane at which the listener has the highest level of resolution in terms of spatial separation of sound sources [13]. Then, the other two rings, upper and lower, are composed of only 4 speakers since the human sound spatial resolution is poorer at those altitudes. In both rings, the 4 loudspeakers are separated of 90° one another and are tilted so that their acoustical axis points toward the center of the speaker sphere. Furthermore, their position was chosen to favor stereophonic listening, as shown in Fig. 1. Furthermore, the listening position, matching the sweet spot, inside the room and the radius length of the speakers sphere was chosen so as to fully exploit the maximum width of the room as shown in Fig. 3(a), where the floor plan of the Audio Space Lab with the projection of 3 speakers rings is detailed. Each of the 8 speakers of the middle ring is mounted on one Genelec 8000-409B adjustable solid steel floor stand, while both upper and lower speakers make use of custom-made supports.

In particular, the upper speakers are hung up through simple brackets to a 2 m diameter aluminum ring (thick 5.5 cm) secured to the ceiling, while each lower speaker is attached to a custom-made iron tilted plane of 45° (thick 8 mm) through the Genelec 8000-402B adjustable wall mounting bracket. Overall, the mounting system was conceived to be as flexible as possible so that the speakers could be easily moved and adjusted to match the wanted position while being as unobtrusive as possible to limit sound field distortions that could potentially entail biased reproduction error metrics and perceptual test results [7]. Moreover, the two Genelec 8351-B are placed on the floor in the bottom of the room facing the listening position. Finally, all speakers are connected, through Canare Star Quad cables with Neutrik XLR connectors, to the Antelope Orion32 32-channel sound card directly driven by the high-end Desktop PC (CPU: Intel® CoreTM i7-12700F, GPU: NVIDIA GEFORCE RTX 3080 Ti, RAM: 32 GB) running the real-time signal processing for the decoding of the 3OA audio tracks into the single signal feeding the loudspeakers implemented on the commercial Plogue Bidule block-and-wire DAW. Fig. 3 shows the real system picture implemented inside the Audio Space Lab, comprising, on the right, the control station and, on the left, the spherical speakers array in the center of which an adjustable chair where the listener should be seated is placed.

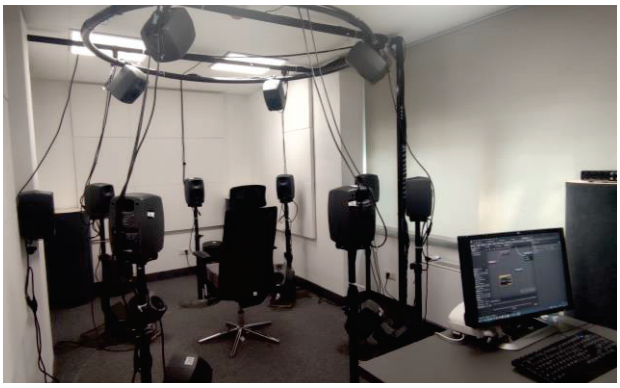


Figure 3. Picture of the Audio Space Lab hosting the designed 3OA sound reproduction system.

In particular, the listener's chair was chosen, making sure that the height could be adjusted so that subjects of different heights could be properly centered in the sweet spot. The fabric chair has a solid steel base that can be moved to make the listener's ears center the sweet spot, and that allows a rotation around the listener's longitudinal axis when the listening test allows it. Furthermore, the chair is provided with an unobtrusive breathable mesh headrest, used to keep the subject's head still in the sweet spot. Overall, the budget required for the installation of the entire system (mounting construction, electronic hardware, and software components), excluded the room acoustical treatment, which strongly depends on the structural characteristics of the available room, hovers at around 20000 €.

2.2 Signal processing and tuning procedure

As aforementioned, the real-time signal processing used to properly decode the loudspeaker-independent HOA tracks into driving signals for the speakers, and to simultaneously make the system provide a flat frequency response in the sweet spot, is implemented inside a single patch of the Plogue Bidule DAW. Fig. 5 outlines the schematic of the processing patch made up of the following interconnected blocks:

- 3OA Player: where the 3OA audio tracks must be loaded and played to make the processing and the sound reproduction start;
- AllRA Decoder IEM spatial plugin: used to properly decode the 3OA track into signals fitting given the current loudspeakers array arrangement (see Fig. 4);
- Gain blocks: used to adjust the channels gain;

- MultiEQ IEM plugin blocks: which are frequency filters used to equalize each single speaker;
- Delay blocks: needed to delay the signal for each channel;
- Orion32 ASIO Driver: used to route the processed signals to the sound card.

All these blocks were then set during the system tuning procedure. In particular, the tuning procedure, used to ensure a flat frequency response in the sweet spot, was carried out following 3 steps:

1. Tuning of the acoustical response of each speaker individually;
2. Tuning of the acoustical response of the whole system when playing ambisonics tracks;
3. Fine-tuning based on subjective evaluations.

The tuning procedure was carried out inside the same patch of Bidule, synchronizing the acquisition of a class-1 omnidirectional microphone placed in the center of the loudspeaker array. The procedure also made use of the following VSTs:

- X-MCFX-CONVOLVER: necessary for processing the measured IRs for each speaker by sine sweep method [15];
- Wave observer: needed to observe the IRs in the time domain in order to adjust the delays for each speaker;
- Voxengo span: which is the spectrum analyzer block needed to observe the spectrum of the signal generated by each speaker.

In the first phase, i.e., the individual speaker tuning, a reference speaker was chosen (the one at 0° azimuth and 0° elevation angles), which was equalized with IIR filters so as to obtain a flat response (by smoothing in 1/3 octaves) from 90 Hz to 20 kHz, avoiding the use of filters with Q > 5. Then, all other speakers were equalized so as to have a response as close as possible to that of the reference speaker.
In the second phase, the overall loudspeaker system was equalized during the playback of ambisonic tracks. First, equalization of the two speakers used for low frequencies was performed to ensure proper balance with the 16-speaker array. Next, using common IIR filters on all channels, the system was equalized so as to correct for spectral cancellations and emphases due to the simultaneous use of multiple speakers.
During the last phase, the final fine-tuning was carried out based on subjective listening of ambisonics tracks. In this phase, a circumscribed adjustment of the ambisonics decoder parameters was made so as to maximize the naturalness and spatial definition of the entire ambisonics system (decoder weights: maxrE).

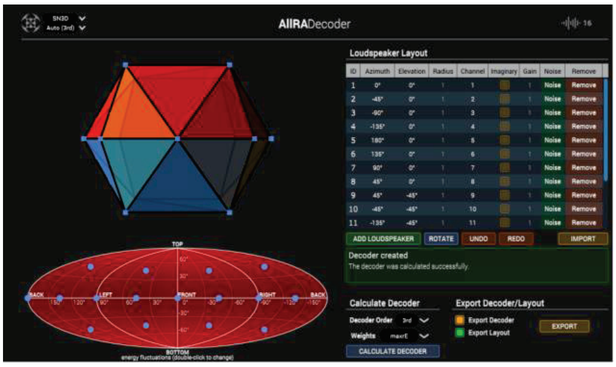


Figure 4. AllRA Decoder IEM spatial plugin used to decode the 3OA track.

2.3 Preliminary system evaluation

To retrieve a first evaluation of the goodness of the implemented system, preliminary measurements of the main room acoustics parameter (i.e., reverberation time (T_{20}), Early Decay Time (EDT), and speech Clarity (C_{50})) were performed on a virtual sound environment reproduced inside the Audio Space Lab and were compared with the values measured in the corresponding real environment. An 800 m^3 volume classroom was chosen as the benchmark. At first, Room Impulse Response (RIR) measurements were performed in the classroom through recordings of exponential sine sweep signals emitted by the NTi Audio Talkbox acoustic signal generator (flat frequency response from 0.1 to 10 kHz).

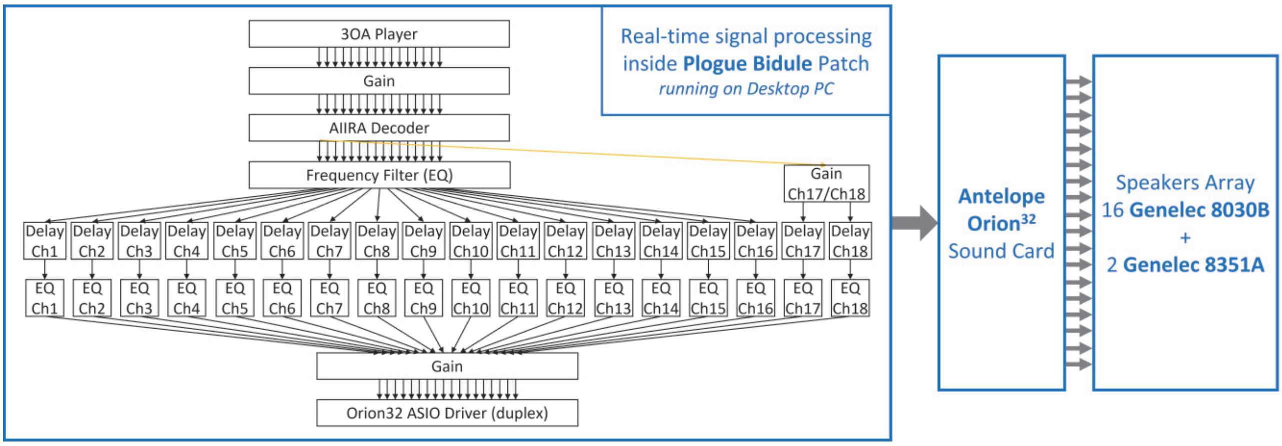


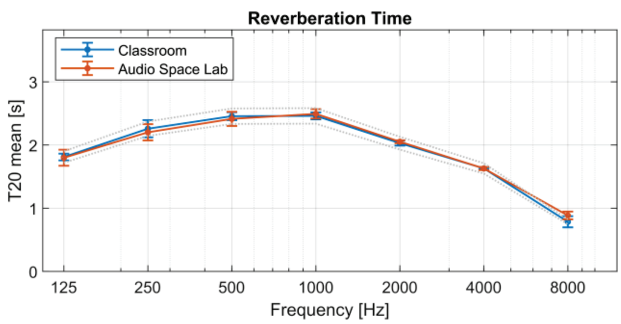
Figure 5. Schematic of the signal routing, including real-time processing, of the built 3OA reproduction system.

The recordings were taken in the same 5 random positions inside the room (2 repetitions for each location placed more than 2 m far from the source and all other receivers and positions) with both the NTi Audio XL2 calibrated omnidirectional class-1 sound level meter (slm) and the Zylia ZM-1 Spherical Microphone Array (SMA) at a 48 kHz sample frequency (32-bit float). Furthermore, the background noise level measurement was performed in one of the 5 positions through the slm, which showed values less than 37 dBA for each octave band from 125 Hz to 8 kHz. Afterward, the 3OA tracks, acquired convolving the SMA recordings with the A2B-Zylia-3E-Jul2020 19x16 filter matrix[1], were reproduced inside the Audio Space Lab and re-recorded by placing the XL2 microphone in the center of the loudspeakers array. Then, all NTi recordings acquired both in the classroom and in the Audio Space Lab were processed using a Matlab routine first to compute the actual RIRs and then measure all room acoustics parameters following the ISO 3382-2 standard [16] by applying the backward integrated impulse response method.

3. RESULTS AND DISCUSSION

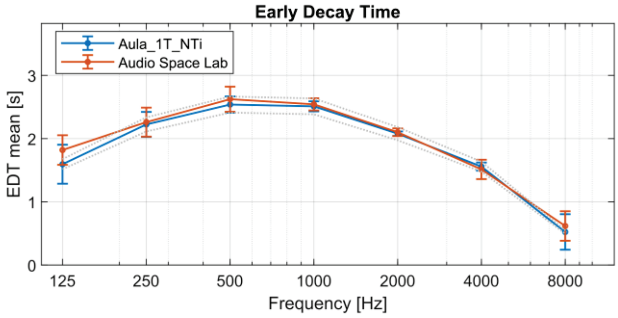
T_{20}, EDT, and C_{50} for both the real classroom and the classroom auralization inside the Audio Space Lab were measured in octave bands from 0.125 to 8 kHz and spatially averaged according to [16]. Frequency averages, standard deviations, and Just Noticeable Difference (JND) values [17] for all parameters are reported in Fig. 6. At first glance, both the real and the virtual environment measures follow the same trend for all parameters, with almost all average values falling within the JND and similar standard deviation values. However, the main results can be summarized as follows.

- Virtual environment T_{20} average values fall within the JND, computed referring to the real environment averages, for all frequencies up to 8 kHz, for which the virtual environment average exceeds the upper JND limit of 0.06 s. It follows that reverberation time is well-preserved inside the virtual environment auralized by the Audio Space Lab system and that the room, also due to the system tuning procedure, does not provide significant additional reverberation in the sweet spot.
- Similarly, for the EDT, the average values of the virtual environment fall within the JND for all frequencies, except for the frequencies at the extremes, i.e., 125 Hz and 8 kHz, exceeding the upper JND limit of 0.14 s and 0.06 s, respectively.

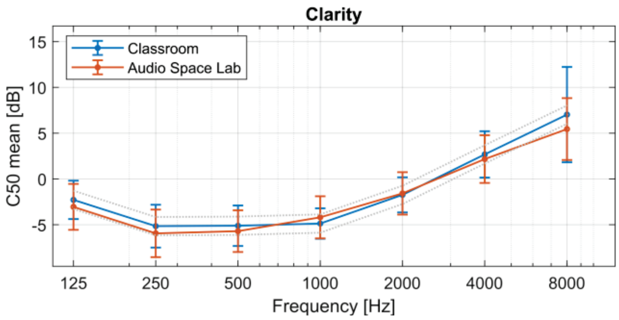


(a)

[1] freely available at http://pcfarina.eng.unipr.it/Public/Xvolver/Filter-Matrices/Aformat-2-Bformat/Zylia-Jul-2020/



(b)



(c)

Figure 5. Results (average and standard deviation) from octave frequency band room acoustics parameter measurements of the real environment (Classroom) and virtual classroom environment (Audio Space Lab). Grey dashed lines refer to the JND for each parameter. (a) T_{20}. (b) EDT. (a) C_{50}.

- Finally, as expected, the C_{50} shows the reverse trend of the reverberation time, being the ratio between the energy in the first 50 ms after the direct sound and the remaining portion of the decay curve. All virtual environment average values fall within the 1 dB JND for all frequencies up to the 8 kHz, at which the lower JND limit is exceeded by 0.5 dB. However, according to [18], a JND value of 3 dB would be a more reasonable estimate compliant with actual minimum clarity differences detectable in everyday life listening situations. Thus, if a 3 dB JND is considered for C50, all virtual environment average values would actually fall within the JND, pointing out that the created system achieves a convincing auralization at first notice.

4. CONCLUSIONS

This paper presents the installation, tuning, and first validation procedures of a simple 3rd order ambisonics virtual sound environment reproduction systems, installed in a small sound-treated room (i.e., the Audio Space Lab), which can be easily replicated and used inside clinical settings (system cost of around 20000 €) to implement ecological listening tests aimed at performing more effective hearing-loss assessments and hearing devices tunings. After the details on the set-up and procedure to implement the proposed system, a first preliminary evaluation of the sound reproduction accuracy of an auralized environment based on standard monoaural room acoustics parameter measurements, i.e., reverberation time (T_{20}), early decay time (EDT), and speech clarity (C_{50}), compared with the corresponding real environment was performed, which overall showed average values for all parameters computed for the virtual environment inside the Audio Space Lab that fell within the just noticeable difference computed starting from the real environment average values. From that, it seems that the proposed system achieves a convincing auralization, being able to recreate the same frequency trend of T_{20}, EDT, and C_{50} (at least from 125 Hz to 8 kHz) parameters that would be measured in the corresponding real environment. However, still, further investigations on the reproduction accuracy of the implemented system should be carried out in the future, evaluating the system spatial response, which may be accomplished through the measurement of early and late lateral energy (J_{LF} and J_L) parameters, and binaural parameters, such as the inter-aural cross correlation coefficients (IACC). Furthermore, also perceptual tests could be performed by comparing, for instance, the results from auralized speech intelligibility tests inside the Audio Space Lab and the ones from the same test carried out in the corresponding real environment in order to effectively verify how results from listening test performed in the proposed loudspeaker-based virtual sound environment actually translate in real-world environment settings. Finally, to further boost the ecological validity of the proposed system, the audio reproduction could be coupled with visual stimuli corresponding to the provided acoustical information. To this end, a head-mounted display, which would entail negligible added costs (around 500 €), could be synched with the audio playback system so as to return the listener a multi-perceptual 3D immersive scene. However, it would require the editing of the current software environment to properly add the management of the visual stimuli and the synchronization between the acoustical and visual scenes.

5. REFERENCES

[1] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acta Acustica united with Acustica*, vol. 86, no. 1, pp. 117–128, 2000.

[2] F. Pausch, L. Aspöck, M. Vorländer, and J. Fels, "An extended binaural real-time auralization system with an interface to research hearing aids for experiments on subjects with hearing loss," *Trends in Hearing*, vol. 22, p. 2331216518800871, 2018.

[3] J. Cubick and T. Dau, "Validation of a virtual sound environment system for testing hearing aids," *Acta Acustica united with Acustica*, vol. 102, no. 3, pp. 547–557, 2016.

[4] G. Grimm, B. Kollmeier, and V. Hohmann, "Spatial Acoustic Scenarios in Multichannel Loudspeaker Systems for Hearing Aid Evaluation," *J Am Acad Audiol*, vol. 27, no. 7, pp. 557–566, Jul. 2016, doi: 10.3766/jaaa.15095.

[5] B. U. Seeber, S. Kerber, and E. R. Hafter, "A system to simulate and reproduce audio–visual environments for spatial hearing research," *Hearing research*, vol. 260, no. 1–2, pp. 1–10, 2010.

[6] S. Favrot and J. M. Buchholz, "LoRA: A loudspeaker-based room auralization system," *Acta acustica united with Acustica*, vol. 96, no. 2, pp. 364–375, 2010.

[7] F. Pausch, G. Behler, and J. Fels, "SCaLAr–A surrounding spherical cap loudspeaker array for flexible generation and evaluation of virtual acoustic environments." EDP Sciences, 2020.

[8] L. S. Simon, N. Dillier, and H. Wüthrich, "Comparison of 3D audio reproduction methods using hearing devices," *Journal of the Audio Engineering Society*, vol. 68, no. 12, pp. 899–909, 2021.

[9] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764–2778, 1993.

[10] M. M. Hendrikse, G. Llorach, V. Hohmann, and G. Grimm, "Movement and gaze behavior in virtual audiovisual listening environments resembling everyday life," *Trends in Hearing*, vol. 23, p. 2331216519872362, 2019.

[11] A. Weisser *et al.*, "The ambisonic recordings of typical environments (ARTE) database," *Acta Acustica United With Acustica*, vol. 105, no. 4, pp. 695–713, 2019.

[12] S. Fichna, T. Biberger, B. U. Seeber, and S. D. Ewert, "Effect of acoustic scene complexity and visual scene representation on auditory perception in virtual audio-visual environments," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, IEEE, 2021, pp. 1–9.

[13] C. Oreinos and J. M. Buchholz, "Objective analysis of ambisonics for hearing aid applications: Effect of listener's head, room reverberation, and directional microphones," *The Journal of the Acoustical Society of America*, vol. 137, no. 6, pp. 3447–3465, 2015.

[14] "RECOMMENDATION ITU-R BS.1116-3 – Methods for the subjective assessment of small impairments in audio systems," *R BS.*.

[15] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio engineering society convention 108*, Audio Engineering Society, 2000.

[16] "UNI EN ISO 3382-2:2008 - UNI Ente Italiano di Normazione." https://store.uni.com/uni-en-iso-3382-2-2008 (accessed Nov. 29, 2022).

[17] ISO/TC 43/SC 2 Building acoustics, "ISO 3382-1:2009," *ISO*. https://www.iso.org/standard/40979.html (accessed Nov. 29, 2022).

[18] J. S. Bradley, R. Reich, and S. Norcross, "A just noticeable difference in C50 for speech," *Applied Acoustics*, vol. 58, no. 2, pp. 99–108, 1999.