2020 IMEKO TC-4 International Conference on
Metrology for Archaeology and Cultural Heritage
Trento, Italy, October 22-24, 2020

Space & sound characterisation of small-scale architectural heritage: an interdisciplinary, lightweight workflow.

J.Y Blaise[1], I.Dudek[1], A.Pamart[1], L.Bergerot[1]
A.Vidal[2], S. Fargeot[2], M. Aramaki[2], S. Ystad[2], R. Kronland-Martinet[2]

[1] UMR CNRS/MC 3495 MAP 31 chemin Joseph Aiguier 13402, Marseille, France.
(jean-yves.blaise, et al.)@map.cnrs.fr
[2] Aix Marseille Univ, CNRS, PRISM, Marseille, France.
vidal@prism.cnrs.fr

Abstract **– This contribution reports on a data acquisition and processing chain the novelty of which is primarily to be found in a close integration of acoustic and visual/metric data. Its outputs pave the way for proportion-as-ratios analyses, as well as for the study of perception aspects from the acoustic point of view. Ultimately, "perceptive" data will be related to "objective" data such as acoustic descriptors or architectural metrics. The experiment is carried out on a set of fifteen "small-scale" rural chapels, which is a corpus intended at fostering cross-examinations and comparative analyses. The specificity of this corpus in terms of architectural layout, of use, and of economic and access constraints, will be shown to have had a significant impact on the technical and methodological choice made all along the acquisition and processing.**

I. INTRODUCTION

Architecture is perceived not only through vision but also through audition - and other senses – hence characterising it is likely to require more than studying its physical envelope. This fact is increasingly acknowledged, including in heritage studies, as illustrated in initiatives focusing on "places" as such [1] [2] or on their use [3].

Yet in the context of small-scale architectural heritage scientists and local communities face a specific challenge: studying, documenting, releasing end-user information sets about assets that are left aside from large, well-funded, heritage programmes. Hence a necessity to minor as much as possible the complexity and cost of workflows.

Our global objective is to support a multidimensional, interdisciplinary characterisation of small-scale architectural heritage. In that context we report on the programme's initial milestone: a data acquisition and processing chain integrating visual and auditory data. The contribution is above all about methodology: we basically combine and extend in a novel way pre-existing technologies and tools. For instance, the photogrammetric survey bases on a 360 panoramic camera (a tech discussed in [4]), and 3D point clouds are exploited inside the *Potree* renderer (well known in the application field) [5].

On the other hand, the influence of the rooms on the sound rendering has been studied for decades with for instance the seminal works made by Sabine on Reverberation [6]. Recent improvements in the field of 3D sound makes it now possible to accurately reproduce previously recorded rooms sound field, thanks to an array of loudspeakers. This allows for an experimental analysis of the induced perception, a key issue as far as this research is concerned. Here, the acquisition step builds on a 3D microphone released on the consumer market, the mh acoustics Eigenmike, a spherical array of 32 microphones, already used for sound field analysis and for sound perception studies [7], [8].

The originality of the research is rather to be found in a combination of techs and methods, with a twofold ambition:

- An interdisciplinary approach that runs all along the data acquisition, processing, and analysis chain (the word interdisciplinary should be understood as defined by [9]: mutual integration of concepts, methodology, procedures).
- A grid of metrics (space + sound) intended at fostering cross-examination.

The experiment is conducted on fifteen interiors of rural chapels in south-eastern France, a collection that brought to light a significant set of feedbacks in terms of methodology and open issues. This paper will focus on the survey step (section 3) and on the data processing step (section 4). But the former and the latter steps hinge on a set of critical choices in terms of corpus and practical constraints, as well as of analytical needs (perception analysis, extraction of proportions, etc.) - aspects briefly debated in sections 2 and 5.

II. CONTEXT AND REQUIREMENTS

The data acquisition and processing chain we present builds on a series of constraints and choices that ensue from both the corpus under scrutiny (small scale rural architecture), and the overall objective of the research (interdisciplinarity, reproducibility, comparability).

The setup and protocol we present was initially designed to address a set of general constraints : a limited time spent in situ (3 hours as a maximum, all included, 2 edifices surveyed per day), soundfield recording along with live recording of usage scenarios, a distribution of speakers and microphones tailored to specific usage modalities (celebrant vs. listener), a lightweight instrumentation (accessibility issues), adaptability to a corpus chosen in order to maximise diversity (in terms of architectural layout, but also in terms of dimensions – volumes ranging from 171 m3 to 981 m3 – *cf.* Fig 1.), *etc.*

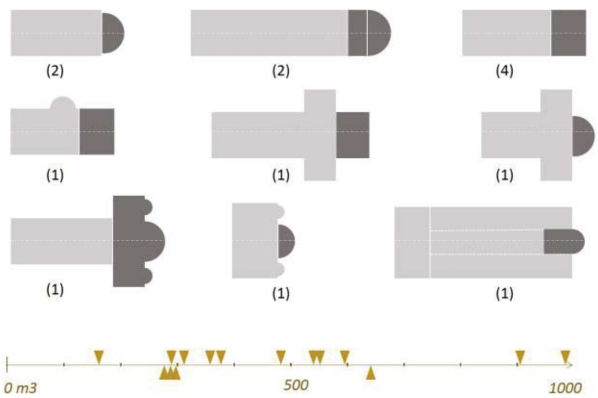

Fig. 1. Diagrams illustrating the diversity of the corpus (spatial layouts in plan, in dark grey chancels, and below figures amount of building corresponding to each "type"). Bottom: volumes, in cubic meters. Triangles above the axis correspond to edifices located in remote areas, at a distance from villages, triangles below the axis correspond to edifices inside or in the vicinity of villages.

The acquisitions were conducted on real cases that introduced yet more constraints: the whole setup had to be chosen so that it could be carried in backpacks (remote sites), it needed to be autonomous in terms of energy (no power supply in situ), and it had to be adapted to interiors that in some cases could be congested – hence a difficulty to maintain the geometry of the grid of instruments. Finally, adaptation to lighting conditions was a recurring problem – conditions varied from large openings, sunny days, to almost no openings, rainy days – hence a necessity to think out solutions in situ. Four LED grids were used when needed, but their correct positioning can be time consuming if wanting to avoid too strong contrasts during the photogrammetric survey.

The acquisition process is now mature, although still improvable, and can be considered as reproducible. But it has to be said that there will always be a series of "expert" choices to make in situ (lighting, number and distribution of stations - from 14 to 93 stations in our experiments, positioning of the rangefinder, analysis of the interior enveloppe, etc.).

III. A MULTIMODAL SURVEY PROTOCOL

The protocol's key components are in fact two low-cost 3D Cross Line Self-Leveling Laser level (instruments often in use in the building activity), that project green laser beams on surfaces. The laser beams are combined in such a way that they mark four planes: they are used to build a reference system and to position auditory devices. Intersections of beams on the walls, ceiling and floor are called "named reference points" and act as markers in the scaling of the photogrammetric model: their relative positions are measured using a Leica rangefinder. Auditory instruments form a grid allowing for a systematic relative positioning of instruments with regards to one another (Fig 2.).

The grid's positioning in the reference system is also done using the laser rangefinder (except for two microphones positioned thanks to the photogrammetric model). Sound devices are mounted on tripods positioned relatively to the named reference points, and reused (once the emitting / recording tasks are over) to install the 360 camera we use in the photographic acquisition, hence allowing for a double checking of the sound devices' positions.

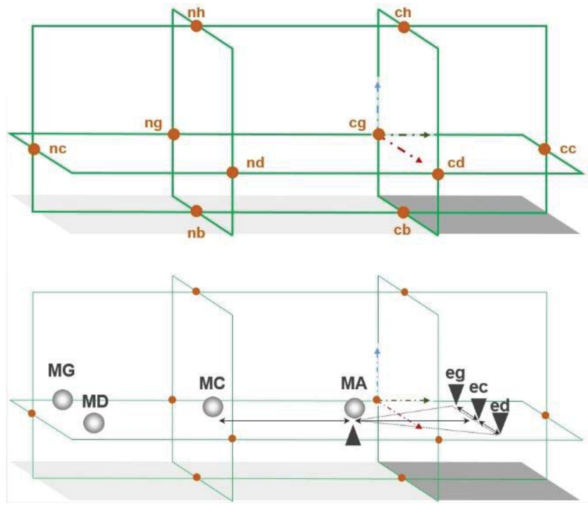

Fig. 2. Top, laser beams (green lines) and their intersections form "named reference points" (brown circles). The light grey parallelogram is the edifice's nave, the dark grey parallelogram is its chancel. Bottom, audio devices positioned relatively to the horizontal plane marked by laser beams. Three speakers are located right behind the altar, in the chancel (dark downwards triangles), at a given distance from one another. Microphones are then positioned relatively to the three speakers at systematic distances (equilateral triangle).

The grid's positioning in the reference system is also done using the laser rangefinder (except for two microphones positioned thanks to the photogrammetric model). Sound devices are mounted on tripods positioned relatively to the named reference points, and reused (once the emitting / recording tasks are over) to install the 360 camera we use in the photographic acquisition, hence allowing for a double checking of the sound devices' positions.

A. Metric and visual data acquisition

The photogrammetric data acquisition protocol for indoor space still suffers from several obstacles related the architectural context (narrow and dark spaces, occlusions) of the chapels. In order to gain in velocity, reproducibility and overall efficiency 360 cameras (dual sensors with fisheye lenses) were chosen as the expected accuracy is centimeter-sized. The choice of this camera type seems also relevant with perceptive analysis purposes foreseen and have interesting technological similarity (spherical projection) with the sound data acquisition devices. A flexible acquisition layout has been conceived aiming to be combined with telemetric surveys (using DXF feature of Leica S910) of architectonic points and instrument positions. This built-in feature is used to extract precise measurements (used as Ground Control Point) and allow orient all the 3D models in a consistent and constant absolute Cartesian coordinates. The technical setting of the metric survey protocol is bounded by economic constraints on one hand (preferably low-tech, low-cost), and compactness on the other hand (compatibility with remote sites) – main components are (see Fig 3.):

- Two Huepar 3D Cross Line Self-Leveling Laser levels,
- YI 360 VR Panoramic Camera - 5.7K HI Resolution, Dual-Lens - each lens is 220° with an aperture of f/2.0 (360° coverage , produces two unstitched hemispherical photograph for each shooting position),
- a laser Rangefinder Leica DISTO S910 (outputs DXF files),
- A Manfrotto tripod (055 series) allowing for horizontal/vertical shootings. The rotational mechanism of centre column is used to perform a pyramidal-based capture combining the benefits of faster survey (5 positions for a single tripod station) and better reconstruction (from a short baseline cameras network with variable orientations).

The main steps of the protocol are as follows:

1. Positioning the laser levels, starting by the one located at then entrance of the chancel.
2. Positioning of the grid of instruments (7 tripods), aligned with the levels vertically, and relatively to one another horizontally (the reference point being the theoretical position of a celebrant behind the altar).
3. Positioning of the rangefinder so that each and every intersection of laser beams is visible and can be surveyed.
4. Survey, using the rangefinder, of the grid of instruments – outputs a polyline connecting tripods to 5 points on the edifice.
5. Scaling protocol, using the rangefinder: a polyline that connects all the laser beams intersections (see Fig 5).
6. Dimensioning protocol, using the rangefinder: a polyline that connects laser beams intersections to elements of the enveloppe considered as significant (a keystone, the entrance's level, a cornice, etc.).
7. Photogrammetric survey, using the panoramic camera positioned on each tripod forming the grid, and then on its own tripod, decided in situ, moved in positions decided in situ.

Steps 1, 2, 4, 5 are systematic, steps 3, 6 and 7 require an adaptation to conditions found in situ. Steps 4 to 7 are conducted before or after the acoustic measurements, depending on the lighting conditions.

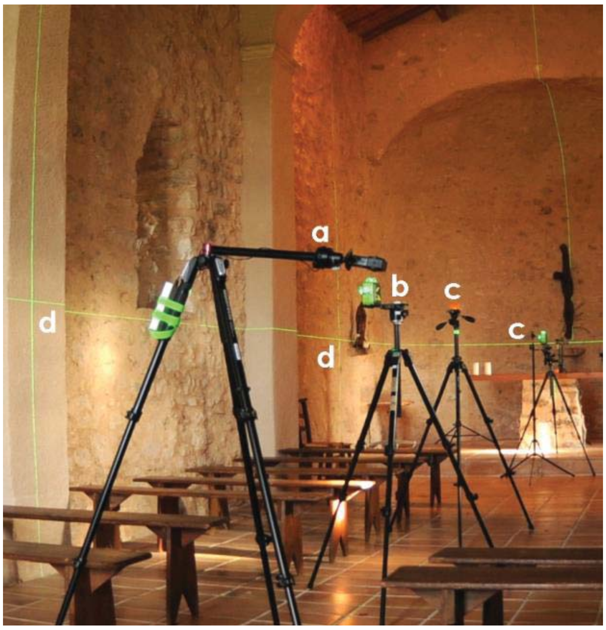

Fig. 3. A sample setup: **a** *the 360 camera; oriented horizontally,* **b** *a laser level,* **c** *tripods on which acoustic devices wuill bemounted,* **d** *intersection of laser beams on the enveloppe.*

Image-Based Modelling (IBM) has been chosen for the cost efficiency and user friendliness aspects, however due to the type of camera chosen (low-cost, low-res) and to the variety of the interior spaces we also used in two of the fifteen sites a Terrestrial Laser Scanner (Range-based modelling) in order to support future qualitative evaluations of results, and as a backup solution in complex case studies.

B. Acoustic measurements

From an acoustical point of view, the main goal of this research is to study the influence of rooms on sound perception: in that context getting consistent results requires the same listeners to perform the same perception assessment tasks for every chapels. However, since human immediate auditory memory is short, this prevents listeners to compare a collection of remote chapels in situ. For that reason, we chose to capture the acoustics of the chapels to reproduce them within a 3D sound platform. To do that, we used a 3D sound technology based on the Higher Order Ambisonics formalism. It is well known that building materials and furniture have a major influence on the sound field. Our measurements aimed to characterize the current state of the chapels: we did not characterize the influence of a change in materials or furniture.

Measurements consisted in characterizing the so-called Room Impulse Responses (RIR). An impulse response corresponds to the sound transformation between a sound source (generated by a loudspeaker) and the sound measured at a microphone level. The RIR allows to proceed, in a second step, to the so-called "auralization". Thanks to the convolution operation of an arbitrary sound stimulus with the RIR, one can play any stimulus as if it was played in situ. To estimate the RIR we used a method presented in [10]. Emitted sounds were three logarithmic sine sweeps from 20 Hz to 20 kHz with a duration of 10s each and separated by 10 s of silence.

The objective is to study the perception in accordance with the sites' initial use: a celebrant near the altar speaking to listeners in the nave. Thus, we placed a loudspeaker in the middle of the chancel (celebrant's privileged position - point ec, see Fig 2.). Two lateral loudspeakers (eg, ed) are then aligned with ec at a distance of 1.25 m (Epistle side vs. Gospel side in terms of initial use, or if thinking about contemporary reuses of chapels simulation of the rendering of a musical trio).

We invariably placed the microphone at point **MC** at a distance of 5.5 m from **ec**, and at the same height. At this distance, the angular spacing between the lateral loudspeakers and the frontal loudspeaker is only 13°. We therefore repeated the same measurements at a closer distance (point MA, apex of an equilateral triangle eg - ed - MA). An "invariable" placement has been chosen instead of a "proportional" placement since the source-receiver distance plays a major impact on the rendering in room acoustics. Indeed, listening at a fixed distance allow to assess only the sound field in the room independently of the measuring distance.

Finally, we placed a fourth loudspeaker 40 cm below the microphones in MA and MC. This specific measurement aims at recording the soundfield as if both the transmitter and the receiver were the same person. This will later allow to study the influence of rooms acoustic's on musicians gestures.

The loudspeakers we used were Genelec 8020C (compact loudspeakers with a frequency range at +/- 2.5 dB, 66 Hz to 20 kHz) and the main microphone we used was a mh Acoustics Eigenmike (spherical array of 32 microphones, see Fig 4.).

In complement to the above measurements, we also positioned two omnidirectional microphones (Neumann U87 Ai) at point MD and MG (in order to capture the soundfield at the entrance of the chapel) and recorded a speaker positioned behind the altar and telling a given sentence while facing the West (nave side) and the East (Chancel side). Main measurements were the diffusion of sine sweeps. Additionally, footsteps of a person walking into the chapel and the speech of a person were also recorded. Finally, 5 min of the soundscape inside and outside the chapel were recorded using a Zoom H3VR.

Fig. 4. The 32-way eigenmeike and speakers positioned for reference tests in an anechoic room.).

IV. DATA PROCESSING AND OUTPUTS

Results of the acquisition step act as inputs processed independently at first and then pulled together again as combined outputs in a twofold way: "consumer" products and analytical overlays to the *Potree* 3D pointcloud renderer.

A. Metric and visual data processing

The first step is to produce 3D point clouds from the YI360 panoramas: named points are transferred into a csv-formatted list and used in order to scale the photogrammetric model (Fig 5.). The resulting point cloud is then exported and integrated in the *Potree* renderer, a free open-source WebGL based point cloud renderer developed at TUWien [5]. One of its most valuable aspects is that it allows for the development of "overlays" - additional functionalities that can be tailored to specific user needs.

We have developed several add-ons concerning either metric or sound data such as display of images corresponding to named points, measurement on recalibrated and concatenated DXF inputs (imported from the Laser rangefinder), representation of each camera inside the 3D scene and link to the corresponding panorama (viewed using the panolens js library), User-monitored selection of sub-clouds (sections corresponding to the laser beams, segmented upstream – see Fig 5.), marching cube method volume calculation, exploratory 3D representation of sound data (clarity and energy indicators), *etc.*

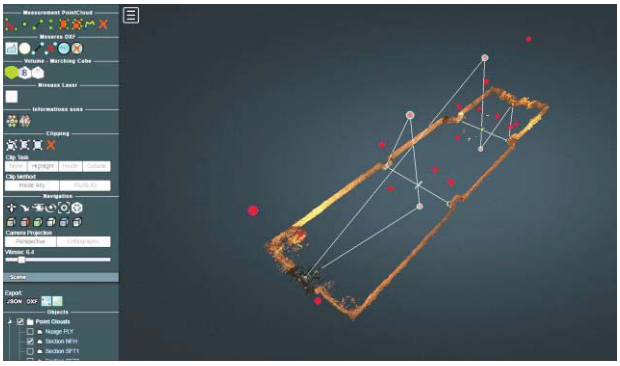

Fig. 5. The polyline that corresponds to the scaling protocol (DXF outputted by the Leica Disto), represented inside the Potree point cloud renderer.

At the end of the day the idea is to use the renderer, in complement with the Leica Disto data, in order to extract significant dimensions / proportions, and to develop analytical models such as rhythmic of contours [11].

B. Sound data processing

The first step of the processing is the computation of the impulse response from the recorded sine sweeps. For that, a standard deconvolution process was used, as explained in [10]. Then, 32 channels impulse responses from the Eigenmike were processed using the Ambisonics formalism. Impulses responses were projected on the basis of spherical harmonics up to the fourth order, leading to 25 components. Then, impulse responses were used in two ways: to create sound stimuli dedicated to the listening, and to compute acoustic descriptors.

Acoustic descriptors were computed using spherical harmonics components of the impulse responses. Most of the indicators were computed using the 0-order component of the spherical harmonics (omnidirectional component). These indicators were the reverberation time (T20 and EDT), the central time, the clarity, the acoustic strength, the Schroeder frequency, the centroid spectral, the bass ratio and the trebble ratio. Few indicators were computed using all spherical harmonics component to take into account the spatial: the InterAural Cross-Correlation (based on a binaural reduction), the Lateral Gain and the Lateral Fraction. As an illustration, Fig. 6 represents the reverberation time T20 in the octave band centered on 1 kHz according to the measurement position. Measurement position had no major influence on the T20, whereas the church had a high influence: values of T20 ranged 1s to 4.5s..

V. NEXT STEPS

Next steps are bounded by a backbone objective: better understanding, characterising an edifice and the way it is perceived, though vision and audition. This implies building on the interdisciplinary nature of the research, including in the analysis steps. Accordingly we have launched a series of experiments aimed at exploiting 3D representations to position and analyse sound data, and at using sound to represent dimensions and geometric features.

As far as metric and visual data is concerned our approach, at first, can be summed up as a "feature extraction" effort: dimensions, ratios-as-proportion [12], *etc.* - as opposed to approaches where a 3D point cloud is analysed as such (segmentation, classification, etc.) [13]. Features can then be compared, trends spotted, exceptions raised and analysed, basing on methods and practices from the infovis (information visualisation) community. Concerning sound data, the next steps of this work is to use the collected data, to map the chapels. In particular, several listening tests will be conducted. The 3D sound field perception is a complex process, leading to several specific experimentations.. For instance, a recent sound sources localisation protocol [14] will be experienced as well as "sound coloration" evaluation. Furthermore, we intend at visually and acoustically immersing the participants, in order to check for the coherency between vision and acoustics.

VI. CONCLUSION

This contribution reports on a data acquisition and processing chain the novelty of which is primarily to be found in a close integration of acoustic and visual/metric data. The overall protocol is first intended at helping actors to characterise and correlate in a consistent way acoustic and morphological features of heritage architecture. It is, secondly, intended at opening up new analytical biases, building on the comparative analysis mantra (Fig 6.). We make no claim this second ambition has yet been reached: what has actually been done is tailoring the data acquisition and processing chain to an interdisciplinary list of requirements, in order to allow for a series of analytical tasks that are now being carried out.

We make no claim this second ambition has yet been reached: what has actually been done is tailoring the data acquisition and processing chain to an interdisciplinary list of requirements, in order to allow for a series of analytical tasks that are now being carried out. The workflow has been applied to a collection of fifteen small-scale edifices, with keeping the constraints linked to that type of heritage asset. The approach does open up new trails research trails, typically in terms of perceptual experiences combining sound and space, or in the 3D visualisation of acoustic data and the sonification of dimensional data.

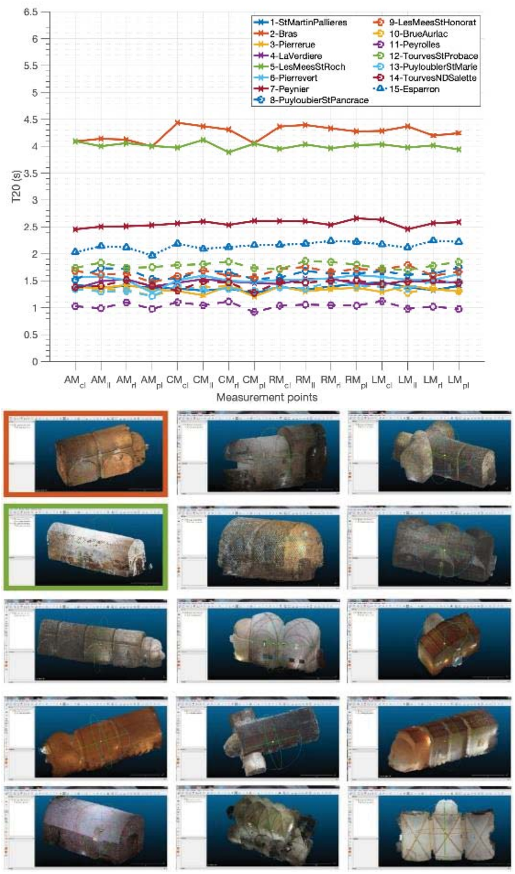

Fig. 6. Comparative analysis of a time-based sound data indicator (reverberation, top) across the collection: two items in the collection stand out significantly (corresponding point clouds are bordered in colour).

Acknowledgments : The project is funded by the ANR (ANR-18-CE38-0009-01).

REFERENCES

[1]L.A. Morales, T. Zamarreño, S. Girón, M. Galindo A methodology for the study of the acoustic environment of Catholic cathedrals: Application to the Cathedral of Malaga Building and Environment 72 (2014) 102-115

[2]Z. Karabiber, The conservation of acoustical heritage. In Cultural heritage research: a Pan-European challenge. Proc. 5th EC conference, 2002, Cracow, Poland. EC- ICSC pp 286-290 ISBN 92-894-4412-6

[3]B. Boren, D. Abraham, R. Naressi, E. Grzyb, B. Lane, et al.. Acoustic simulation of Bach's performing forces in the Thomaskirche. EAA Spatial Audio Signal Processing Symposium, Sep 2019, Paris, France. pp.143-148, 10.25836/sasp.2019.39.

[4]L. Barazzetti, M. Previtali, F. Roncoroni. Can we use low-cost 360 degree cameras to create accurate 3d models? In 2018 ISPRS TC II Mid-term Symposium "Towards Photogrammetry 2020", 4–7 June 2018, Riva del Garda, Italy . https://doi.org/10.5194/isprs-archives-XLII-2-69-20

[5]M. Schütz, Markus. Potree: Rendering Large Point Clouds in Web Browsers. PhDThesis Vienna University of Technology (2016)

[6]W.C. Sabine. Collected Papers on Acoustics. Harvard Univ. Press. Reprinted by Peninsula Publishing, Acoustical Society of America, 1922

[7]A. Farina, A. Amendola, A. Capra, & C. Varani (2011, May). Spatial analysis of room impulse responses captured with a 32-capsule microphone array. In Audio Engineering Society Convention 130. Audio Engineering Society.

[8]D. A. Dick, & M. C. Vigeant (2019). An investigation of listener envelopment utilizing a spherical microphone array and third-order ambisonics reproduction. The Journal of the Acoustical Society of America, 145(4), 2795-2809.

[9]Council of Arts Accrediting Associations 2009 Disciplines in Combination: Interdisciplinary, Multidisciplinary, and Other Collaborative Programs of Study CAAA Briefing paper online http://nast.arts-accredit.org

[10]A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," presented at the 108th AES Convention, France, 2000.

[11]J.Y. Blaise, I. Dudek "Identifying and Visualizing Universal Features for Architectural Mouldings" IJCISIM, Vol. 4, 2012, ISSN2150-7988, pp. 130-143.

[12]Cohen, M.A., 2014 Conclusion: Ten Principles for the Study of Proportional Systems in the History of Architecture. Architectural Histories, 2(1), p.Art. 7. DOI: http://doi.org/10.5334/ah.bw

[13]E. Grilli, F. Menna, F. Remondino. A review of point clouds segmentation and classification algorithms. ISPRS Archives Volume XLII-2/W3, 2017 3D Virtual Reconstruction and Visualization of Complex Architectures, 1–3 March 2017, Nafplio, Greece. doi:10.5194/isprs-archives-XLII-2-W3-339-201

[14]S. Fargeot, O. Derrien, G. Parseihian, M. Aramaki, R. Kronland-Martinet. Subjective evaluation of spatial distorsions induced by a sound source separation process. EAA Spatial Audio Signal Processing Symposium, Sep 2019, Paris, France. pp.67-72.