Acquisition and integration of spatial and acoustic features: a workflow tailored to small-scale heritage architecture

Jean-Yves Blaise[1], Iwona Dudek[1], Anthony Pamart[1], Laurent Bergerot[1], Adrien Vidal[2], Simon Fargeot[2], Mitsuko Aramaki[2], Sølvi Ystad[2], Richard Kronland-Martinet[2]

[1] *UMR CNRS/MC 3495 MAP 31 chemin Joseph Aiguier 13402, Marseille, France*
[2] *Aix Marseille Univ, CNRS, MC, PRISM UMR 7061, 31 chemin Joseph Aiguier 13402, Marseille, France*

ABSTRACT
This paper reports on an interdisciplinary data acquisition and processing chain, the novelty of which is primarily to be found in a close integration of acoustic and spatial data. It provides a detailed description of the technological and methodological choices that were made in order to adapt to the particularities of the corpus studied (interiors of small scale rural architectural artefacts) keeping in mind the backbone objective of the research: facilitate comparisons (among buildings, among spatial and acoustic features). The research outputs pave the way for proportion-as-ratios analyses, as well as for the study of perceptual aspects from an acoustic point of view. Ultimately, "perceptual" acoustic data characterised by acoustic descriptors will be related to "objective" spatial data such as 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 in the context of an architectural programme acting as a constant. The specificity of this corpus, in terms of architectural layout, usage, and economic or access constraints, will be shown to have had a significant impact on choices made during the acquisition and processing chains.

1. 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].

However in the specific context of small-scale architectural heritage, often left aside from large, well-funded, heritage programmes, scientists and local communities face a specific challenge. Indeed, in that context, studying, documenting, and enhancing such buildings requires new methods minoring as much as possible the complexity and cost of workflows.

The above statements correspond to both sides of the equation this research aims to solve: acquiring and integrating spatial and acoustic characteristics, while maintaining a level of simplicity suited to buildings without prestige, often neglected or at risk.

Our global objective is to support a multidimensional and interdisciplinary characterisation of small-scale architectural heritage. This contribution is centred on the programme's initial milestone: a data acquisition and processing chain integrating visual and auditory data. It is (above all) about methodology: as will be shown we mainly combine and extend pre-existing technologies and tools in a novel way.

The photogrammetric survey is based on a 360 panoramic camera (a technology 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 effects of the room's configuration on the sound rendering has been studied for decades with for instance the seminal works on Reverberation

made by Sabine [6]. Recent improvements in the field of 3D sound make it now possible to accurately reproduce previously recorded sound fields, 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.

The originality of the research lies in a combination of technologies and methods, with a twofold ambition:

- to develop an interdisciplinary approach that should be maintained all along the data acquisition, processing, and analysis chain (the word interdisciplinary should be understood as defined by [7]: *mutual integration of concepts, methodology and procedures*),
- to single out a grid of metrics (space + sound) aiming at helping analysts to cross-examine data on and between buildings.

The study is carried out on fifteen interiors of rural chapels in south-eastern France. This collection has shed light on a significant set of feedbacks in terms of methods and open issues.
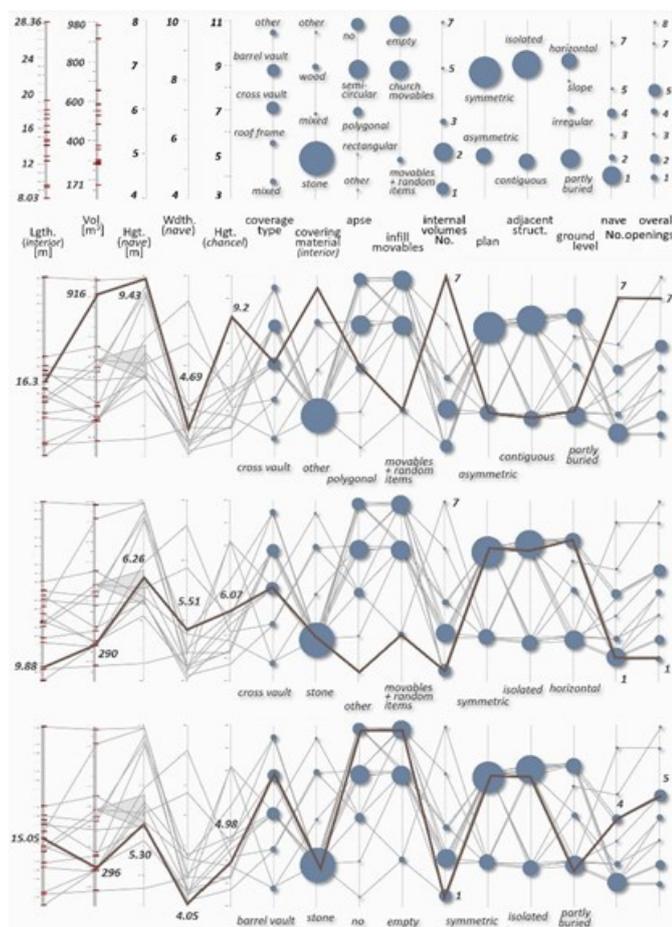

Figure 1. A "parallel coordinates" diagram illustrating the diversity of the corpus – the legend (top) followed by three examples: N.D de la Salette (Tourves), Saint-Roch (La Verdière) and Saint-Roch (Les Mées). Each line - running from left to right - corresponds to one of the fifteen selected chapels. Each column corresponds to one of the "diversity factors" (partial view). The first five bars (on the left) correspond to quantitative data. Other factors are related to categorical data - e.g., covering types (barrel vault, cross vault, roof frame, other vaulting), apse shape (semi-circular, polygonal, rectangular, lack of apse, other), volume complexity (single regular nave vs. buildings with transept or other ruptures in the continuity of the nave), empty spaces vs. buildings with furniture, vicinity factor (isolated buildings vs. buildings with adjoining structures), plan symmetry, etc. Blue circles represent the number of chapels corresponding to a given value (diameter of a circle represents a number, *min* = 1, *max* = 11).

We opted for such a corpus because of its consistency in terms of their initial intended function, and its variability in terms of actual architectural layout. This collection opens an opportunity to analyse and compare the acoustic responses of buildings that initially share a common usage scenario (the Christian ritual) although they differ physically in many ways, as shown in Figure 1 (type of covering, materials, vicinity, *etc.*).

This paper focuses on the survey step (section 3), the data processing step (section 4), and comments on obstacles and limitations of the protocol (section 5). The former and the latter steps hinge on a set of critical choices in terms of corpus, practical constraints, and surveying technologies of architectural interiors. They also revolve around analytical needs (perception analysis, extraction of proportions, etc.). These aspects are debated in sections 2 and 6 so as to put the experiment in perspective.

2. RESEARCH CONTEXT AND REQUIREMENTS

The data acquisition and processing chain presented herein builds on a series of constraints and choices that ensue from the corpus under scrutiny (small scale rural architecture) and the overall objective of the research (interdisciplinarity, reproducibility, comparability). This section comments on these specific constraints and then positions choices made with regards to the state of the art.

2.1. The corpus, the analytical needs

The setup and protocol was designed to address a set of constraints that are key to list if wanting to assess the approach's reproducibility and relevance. Briefly speaking the initial priorities were the following:

- collect a consistent corpus (comparability issue),
- tailor the acquisition phase to that corpus (small scale, poorly funded heritage),
- design a multimodal protocol that would open on repurposable outputs. Those outputs should be valuable for local actors in their effort to favour a better recognition of their heritage assets. They should also be relevant for scientists and scholars in their analytical tasks (on the architectural analysis side as well as on the acoustics analysis side).

The first step of the programme was an interdisciplinary debate about the survey step, including feasibility tests in laboratory conditions. An architectural interior relatively consistent with the corpus was picked up, and named "fake chapel". It served as a substitute for "real" chapels during the early stages of the survey protocol's development (see section 2.2).

Architects, surveyors and acousticians confronted their views on their specific technical requirements in terms of survey, but also on how the data would be exploited in subsequent phases.

That first step ended on the co-design of *strategic, organisational, technical* specifications:

1) a unique setup and protocol that would be reproduced identically across the whole collection.
2) a selection of architectural layouts adapted to the analysis of compositional patterns,
3) a protocol respectful of a chapel's original function – hence a spatial distribution of speakers and microphones tailored to specific usage modalities (celebrant vs. listener opposition). The interior spaces need to be analysed in the light of the way they are or were used. Some of the chapels were indeed repurposed

(modification of function), but their reason to be - ~~and~~ that is the basis on which comparisons can be made - lies in the service they initially offered,

4) a need of reliability and accuracy concerning major dimensions of a chapel and positions of the acoustic devices, and relatively low requirement level in terms of density and quality of the 3D point cloud.

5) a contact-free survey protocol, privileging a lightweight instrumentation (accessibility issue – some buildings can only be accessed by foot),

6) a severe time pressure in situ (three hours as a maximum per building, all surveys included, i.e. maximum two chapels surveyed per day),

7) a need to record each building's soundfield in order to allow for the restitution of its acoustics in an audio cave to perform remote perceptual evaluations and comparisons between chapels,

8) a live recording of usage scenarios: speech (human voice facing the apse or the nave) and sounds produced by a human walking within a chapel according to a specific protocol,

9) a recording of the soundscape, both interior and exterior.

At first glance architectural interiors of the corpus that was selected can be seen as a consistent architectural paradigm. This is basically a misconception [8], and we accordingly chose a collection of fifteen buildings that reflects the diversity of the corpus (in terms of their architectural layouts and dimensions – volumes ranging from 171 m³ to 981 m³ (*cf.* Figure 2).

The actual acquisitions introduced yet more technical 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 in some cases made it difficult to maintain the geometry of the grid of the instruments.

Finally, adaptation to lighting conditions proved to be recurring problem. Conditions varied from sunny summer days in well exposed buildings with large openings, to rainy winter days in chapels with very small windows located in a shady area.
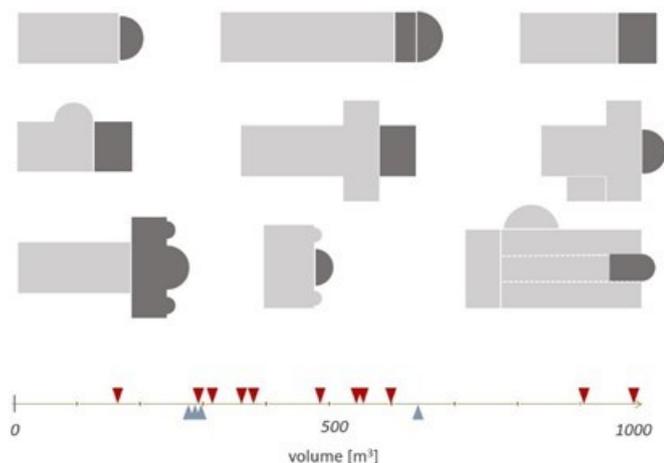

Figure 2. Diagrams illustrating the variety of spatial arrangements of the selected corpus (top - dark grey elements represent chancels) and the variation of the chapels' volume (bottom - triangles above the axis correspond to buildings located in remote areas, at a distance from villages; triangles below the axis correspond to buildings inside or in the vicinity of villages.

Four LED panels were used when needed, but their correct positioning can be time-consuming if one wants to avoid too strong contrasts during the photogrammetric survey.

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

Although this contribution insists on reproducibility, the experiment is tailored to a quite specific set of conditions and constraints. In no way do we claim that our approach is an "all purposes all situations" one. On the contrary, we consider that one of the key outcomes of such an experimental study is to act as food for thinking, especially at a time when technologies often affect methodological choices. It clearly illustrates that attention should be driven towards the "WHAT FOR" question before addressing the "WHAT WITH" question, in particular when targeting a lightweight approach, and minor heritage architecture assets.

2.2. The survey protocol: technological and methodological constraints

The metric data acquisition protocol proposed follows a twofold objective: a fast and versatile geometric survey of small scale indoor spaces, combined with the data integration need of acoustic measurements (from on-site positioning to combined visualisation). The choice to rely on panoramic-based photogrammetry converge to a single solution (by sharing 360° capture) to address the purpose and the challenge of the multimodal survey.

The use of low-cost spherical cameras for photogrammetric reconstruction has been discussed and evaluated in the literature [9], [10], [11], [4]. A critical review of those works provides useful insight on such cameras' potential efficiency and relevance with regards to our case-studies. Most of those works relate possible failings or limitations to the technological limitation of low-cost spherical cameras. And indeed, this technique's potential weakness is mainly related to the low-resolution of the sensor, and to the low quality of optical components for the hardware part. On the software side, spherical stitching and projection does increase the uncertainty above the standard of pin-hole-based photogrammetric reconstruction. With knowledge of these limitations in terms of accuracy we nevertheless chose to build and optimize an experimental setup upon this technique, in the light of this research's priority n°2: *a simplicity suited to unprestigious buildings*.

The "fake-chapel" (see section 2.1) acted as a calibration space to test, evaluate and improve the data acquisition strategy before the actual acquisition campaigns conducted on the fifteen chapels. Because the capture of panoramic images and some photogrammetric rules are contradictory (e.g. parallax condition), panoramic photogrammetry is often used in very specific contexts [12]. Nowadays, spherical photogrammetry can be performed in three different ways:

- the raw images of a constrained multi-camera rig [13] are used separately as a single frame picture usually with highly distorted fisheye lenses,
- an image set is derived to composite picture using stitching algorithm to get a spherical map (usually equirectangular projection),

- a stitched panorama is converted to cubic projection from which six cameras corresponding to each face are extracted to be processed as pin-holes.

The above-mentioned solutions remain sub-optimal in terms of photogrammetric processing, leaving the optimization of data acquisition as a relevant solution. Panoramic-based photogrammetry usually translates the Terrestrial Laser Scanning acquisition canvas with a straightforward linear sequence assuming that omni directional images behave like range imaging. While constant and uniform quality is expected with lasers both raw images and projective maps generated are imperfect for IBM (Image Based Modelling) purpose. The Tissot's indicatrix used in cartography to describe map distortion shows that any projection system, including spherical or cubic projection is non-destructive at pixel level (*cf.* Figure 3a). Therefore, panoramic-based photogrammetry is altered from distortion and aberration at many levels, from the optical system of the camera, during stitching step to the projection mapping. We assumed that only a small part of a panoramic capture is optimal for 3D reconstruction. At the raw level (*cf.* Figure 3d), only on the central part of the fisheye picture (focal axis shown in red in Figure 6a) is performing [14]. At stitching level (*cf.* Figure 3b), the azimuthal plane has artefacts due to low overlapping of raw images. At the mapping level (*cf.* Figure 3c), in the case of the equirectangular projection, poles on zenith axis are inconsistent, and only the horizontal plane is exempt of strong deformation.

Based on this observation spherical photogrammetry cannot be considered as omnidirectional (from a qualitative point of view) because the orientation of the camera system and its sensor is not insignificant. In other words, a simple, linear sequence of images, taken from a camera that would be systematically oriented the same way, impacts in a negative way the quality of the geometrical reconstruction.

In response to that issue, an original data acquisition sequence was developed, in order to compensate the fact that only the
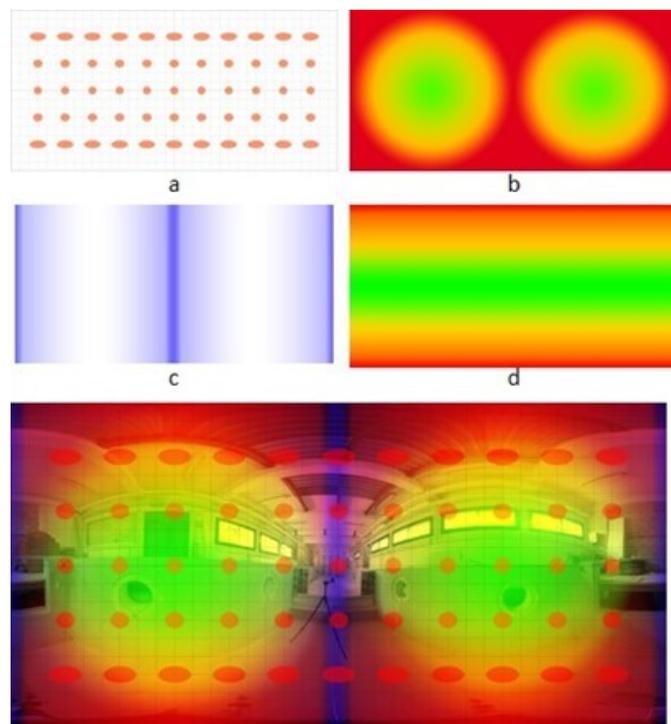

Figure 3. Synthetic image illustrating quality variance of an equirectangular projection composed of : Tissot's indicatrix (a), stitching artifact (b), gradient of equirectangular distortion (c) and gradient of double-fisheye distortion.

longitudinal axis and the horizontal plane of each shot provide optimal features for photogrammetric reconstruction. Instead of walking along the space with a linear sequence, the protocol is based on a dense network of pyramidal sequences (detailed in section 3) composed of translated and rotated camera positions.

This protocol basically reintroduces elements recognized to improve photogrammetric reconstruction like: short-baseline and high overlapping stereo-pairs, roto-translation variance in a dense camera network or feature redundancy to reinforce the panorama bundle adjustment.

In addition, because the principle of spherical capture is shared between 360° cameras and the 3D sound-field microphone used in this study (the mh-acoustics Eigenmike 32), we anticipated a seamless data integration all along the operative chain (i.e. capture, registration, fusion and visualisation stages). The technological analogy of our tools indeed provide similarities to ease data fusion steps and is intended to further facilitate the perceptual analysis, providing immersive environments for both reality-based image and sound captures.

3. A MULTIMODAL SURVEY PROTOCOL

The protocol's key components are in fact two low-cost 3D Cross Line Self-Leveling Laser levels (instruments often in use in the building activity). These levels project green laser beams on surfaces. The laser beams are combined so as to mark four planes constituting a reference system. They are also exploited to position sound measuring instruments. 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 S910 rangefinder equipped with its so-called "smart base" and its integrated tilt sensor. Sound recording instruments (microphones and loudspeakers) form a grid allowing for a systematic relative positioning of the instruments with regards to one another (*cf.* Figure 4.). The grid's positioning in the reference system is also done using the laser rangefinder (except for two microphones, MG and MD, positioned thanks to the photogrammetric model). Microphones and loudspeakers are mounted on tripods positioned relatively to the named reference points (seven positions, *cf.* Figure 4).

These tripods are reused (once the emitting / recording tasks are over) to install the 360 camera and to capture photogrammetric data in these specific points. It hence allows for a double checking of the acoustic (sound recording) instruments' positions. These shots are also, later on in the process, repurposed as the visual components of online immersive panoramas inside which sound tracks corresponding to the named point MA MC MD and MG are displayed according to the corresponding panoramas (see section 6).

3.1. 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 centimetre-sized.

The pyramidal data acquisition protocol has been conceived, evaluated and optimized to face with issues discussed in Section 2.2. A flexible acquisition layout has then been framed so as to couple up the photogrammetric acquisition to telemetric surveys of significant points on the building's surface (using the DXF feature of Leica S910), and to the acoustic instruments positions.
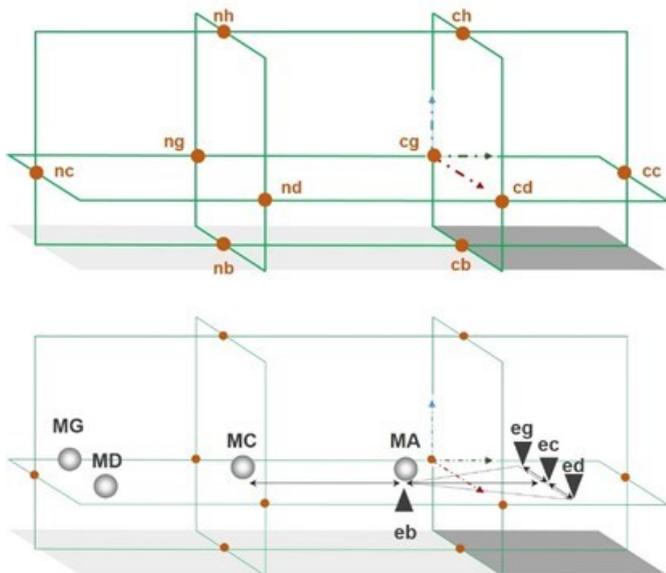
Figure 4. Top, laser beams (green lines) and their intersections form "named reference points" (brown circles) that are visible on the building's surfaces and surveyed using the rangefinder. The light grey parallelogram is the chapel's nave, the dark grey parallelogram is its chancel.

Bottom, auditory instruments positioned relatively to the horizontal plane marked by laser beams. Three loudspeakers **eg, ec, ed** are located right behind the altar, in the chancel (dark downward triangles), at a given distance from one another, with **ec** aligned on the main longitudinal axis. A fourth loudspeaker, **eb**, is positioned right under the first microphone, and tilted so as to face the covering of the building.

Microphone **MA** is then positioned relatively to loudspeakers **eg,** and **ed** at a systematic distance (**MA, eg, ed** form an equilateral triangle). Microphone **MC** is aligned with **MA**, on the main longitudinal axis, and positioned at a fixed and systematic distance from **MA**. Microphones MG and MD are positioned at the very beginning of the nave, at a fixed distance (one meter) from the walls.

The DXF built-in feature is used to extract precise measurements (used as Ground Control Point). It allows to 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 (*cf.* Figure 5):
- two Huepar 3D Cross Line Self-Leveling Laser levels,
- a 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 (this instrument outputs DXF files, it is used to survey *named reference points on one hand and* significant points on the surface of the edifice),
- a Manfrotto tripod (055 series) allowing for horizontal/vertical shootings. The rotational mechanism of the 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, cf. section 2.2).

The main steps of the protocol are as follows:
1) Positioning of the laser levels, starting by the one located at the entrance of the chancel.
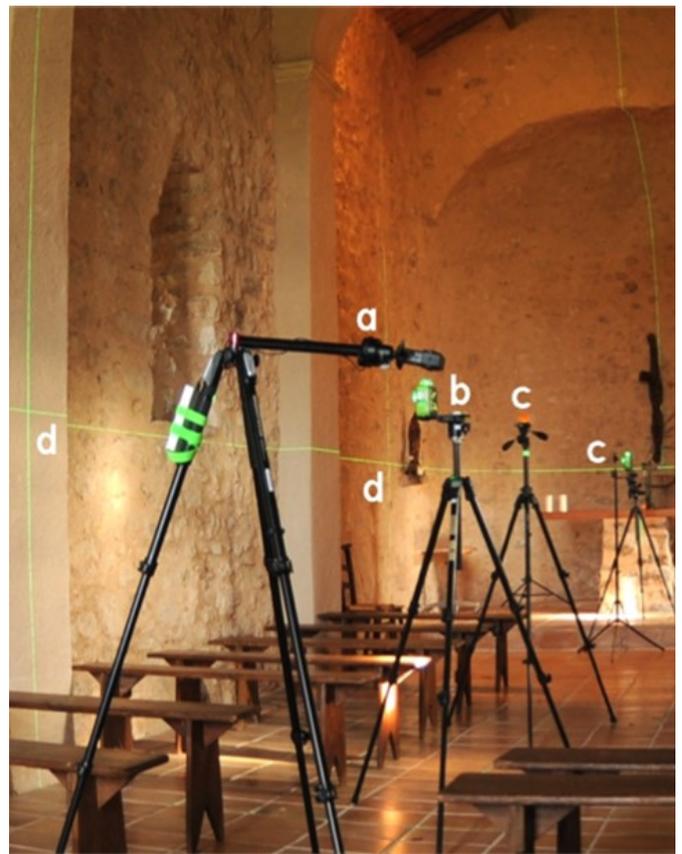

Figure 5. A sample setup: **a** - the 360 camera; oriented horizontally, **b** - a 3 axes laser level, **c** - tripods on which acoustic devices will be mounted, **d** - intersection of laser beams on the interior's enveloppe. (St Pancrace's Chapel, Pyloubier).

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, can be pointed at and surveyed.
4) Survey, using the rangefinder, of the grid of instruments – outputs a polyline connecting tripods to 5 points on the building.
5) Scaling protocol, using the rangefinder: a polyline that connects all the laser beam intersections.
6) Dimensioning protocol, using the rangefinder: a polyline that connects laser beam intersections to elements of the envelope considered as significant (a keystone, the entrance level, a cornice, etc.).
7) Photogrammetric survey, using the panoramic camera positioned on each tripod forming the grid, and then on its own tripod, moved in different positions decided in situ. For each position, the pyramidal sequence (*cf.* Figure 6) is repeated and modulated according to the architectural morphology and environmental constraints.
8) Acoustic survey: instruments – loudspeakers and microphones - are positioned on the grid of tripods, a sine sweep is emitted from each speaker and recorded on the microphones (several times iteratively in order to spot and eliminate "outliers").

9) Live recordings: a given sentence is pronounced systematically by the same person, positioned behind the altar and facing either the chancel or the nave, and footsteps of a person walking from the entrance to the altar and back are recorded.

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, steps 8 and 9, depending on the lighting conditions.

The pyramidal protocol consists of a complete series of 6 pictures, 5 at the summits of a square-based pyramid and one in the centre. The roto-translation between each camera position is fast and easy to reproduce, from a single tripod position, thanks to the rotating arm of the tripod and to a rotative head (see Figure 6, b). Opposed positions are generally arranged by common orientations to create stereo-pairs for which:
- longitudinal capture (camera positioned vertically or horizontally and oriented on the long axis of the chapel shown by the red axis in Figure 6a) helps the alignment along the sequence and correlation of elements perpendicular to side walls,
- zenithal capture (camera positioned horizontally and oriented up and down with blue axis in Figure 6b) improves the cover of floor and ceiling,
- transversal capture (camera positioned horizontally or vertically and oriented on side walls with green axis in Figure 6b) is globally used for close-range reconstructions.

On top of the general improvement for 3D reconstruction this pyramidal protocol turned out fully versatile against multiple in-situ constraints. From our experience on the 15 chapels that were surveyed, several benefits can be noted in terms of :
- completeness: minimization of occluded areas concerning architectural, structural or ornamental elements,
- velocity: Gain in acquisition time by minimizing the number of tripod stations,
- accuracy: increase in redundancy for pointing, scaling and extracting coordinate positions or measurements,

- security: enabling sufficient back-up data in case of a mistake or error,
- practicality: avoidance of obstacles in the sequence (including our own equipment that must remain fixed during data acquisition).

The overall protocol is relatively fluid, and has been systematically reproduced, with as a result, a good feedback now on stop points, *i.e.* key aspects or moments that can result in failures (*cf.* Figure 7).

Positioning of the 3D Cross Line Self- Levelling Laser levels: one at the entrance of the chancel, one half-way along the nave

Check that the two horizontal lines perfectly overlap Check the surface condition where beams intersect

Installing tripods supporting the grid of instruments: three speakers aligned behind the altar.

Check alignment of the three speakers; Check horizontality of the tripod heads using the laser beam and targets

Installing tripods supporting the grid of instruments: two microphones at a systematic distance from the speakers, two microphones at a 1 meter distance from corners on the entrance side.

Check positons : eg ed mc, eg ed ma should form regular triangles

Installing the Leica S910 rangefinder on its tripod (equipped with its orientable head).

Check if each and every intersection of laser beams is visible given potential occultation or dead angles of the head

Applying survey protocol 1: nine points surveyed (movable targets on the grid of tripods, plus four points of the reference system)

Check the angle of incidence of the rangefinder' pointer on intersection of beams. Check the DXF output

Applying survey protocol 2: twelve points (scaling protocol) distributed across the edifices' horizontal and vertical envelopes

Check the angle of incidence of the rangefinder' pointer on intersection of beams. Check the DXF output

Applying survey protocol 3: seventeen predefined points, with in situ adaptations to specific spaces

Check the angle of incidence of the rangefinder' pointer on intersection of beams. Check the DXF output

Remove the rangefinder, and launch the photogrammetric survey itself, adapted in situ to match the spaces

Adapt to the lighting conditions. Check the result

Remove the tripod used for the photogrammetric survey, the movable targets, and start the sound acquisition phase
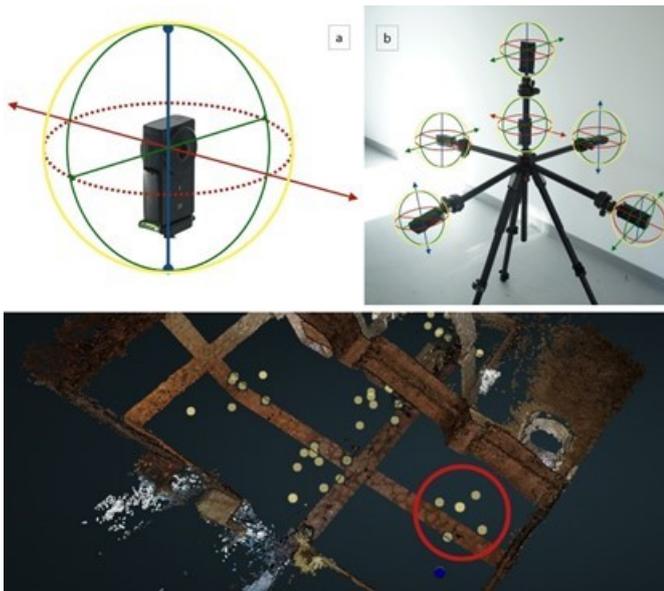

Figure 6. Schema of the world camera system (a) and an example of a complete pyramidal sequence composed of 6 different positions, (b). Bottom, real case application: each yellow sphere corresponds to a camera position.
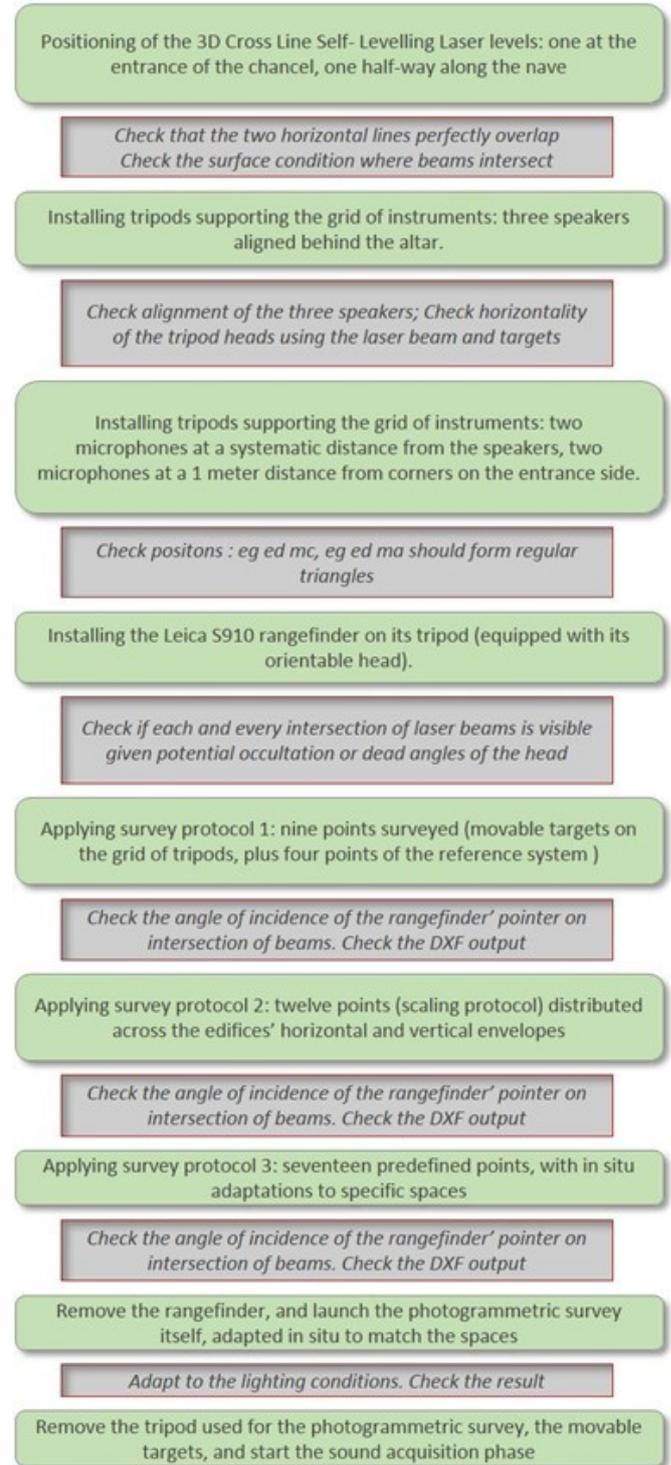
Figure 7. A decision diagram positioning key steps of the survey process (steps 1 to 7, grid installation and metric survey), with possible pitfalls and success factors.

3.2. 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 perceptual assessment tasks for each chapel. However, since human immediate auditory memory is short, it is not possible to compare a collection of remote chapels *in situ*. As a work around, the 3D acoustics of the collection of chapels can be measured and rendered in laboratory conditions. For this purpose, a 3D sound technology (microphone and loudspeaker arrays) based on 3D recordings (Eigenmike 32) and Higher Order Ambisonics (HOA) restitution [15], [16] was used.

Measurements consisted in characterising the so-called Spatial Room Impulse Responses (SRIR). An impulse response corresponds to the sound transformation between a sound source (generated by a loudspeaker) and the sound measured at the microphone level. The SRIR enables, in a second step, to proceed to the so-called "auralization". Thanks to the convolution operation of an arbitrary sound stimulus with the SRIR, one can play any stimulus as if it were played *in situ*.

In this paper, SRIRs were derived from the measurement of sine sweeps, as proposed by [17]. Emitted sounds were logarithmic sine sweeps from 20 Hz to 20 kHz with a duration of 10 s followed by 10 s of silence. This method has many advantages, including fast measurements, good signal to noise ratios (SNR) and immunity to source distortions [18]. It is well suited for quiet closed spaces such as rural chapels. The main drawback is that this technique is sensitive to impulsive noise. To overcome this problem, each measurement is repeated three times in a row and the SRIR is derived from take with the best signal to noise ratio.

Since the aim was to investigate the acoustics related to the sites' initial use, i.e. a celebrant near the altar speaking to the audience in the nave, we placed a loudspeaker in the middle of the chancel (point ec, *cf.* Figure 4). Two lateral loudspeakers (eg, ed) were 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 systematically placed the microphone at point MC at a distance of 5.5 m from ec, and at the same height (*cf.* Figure 8). This distance was constrained by the smallest chapel's

dimensions and corresponded to the largest source-to-microphone distance that can be obtained in the configuration presented in Figure 4. 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, *cf.* Figure 4).

An "invariable" placement has been chosen instead of a "proportional" placement since the source-receiver distance plays a major impact on the room acoustics rendering. 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. It can be used in psycho-physical experiments requiring real-time auralization of autophonic stimuli. As an example, we plan to study the influence of room acoustics on musicians' gestures.

As far as the equipment is concerned, the main measurement was recorded with a 3D microphone released on the consumer market, the mh-acoustics Eigenmike 32 (em32). This spherical array of 32 microphones has already been used for sound field analysis and for perceptual studies [19], [20]. It allows precise spatial recording of the sound-field that can further be converted for restitution purposes to HOA format up to the 4th order. The loudspeakers used were Genelec 8020C. This loudspeaker is not omni-directional (as required for measurement of acoustics parameters [21]), but as mentioned earlier the main goal of these measurements was to proceed to auralization. For this purpose, omni-directionality was not required since the sources to be auralized have their own directivity pattern (*e.g.* voice, guitar, etc.). Moreover, this loudspeaker was chosen here for its compactness, and its relatively low-cost (compared to dodecahedron omni-directional sources) while having fair frequency characteristics (+/- 2.5 dB, 66 Hz to 20 kHz).

This protocol was deployed on the fifteen chapels of the corpus. Additionally, a reference measurement based on the same layout was performed in an anechoic chamber (*cf.* Figure 9). This reference measurement allowed to characterise the setup in free field (*i.e.* without any room effects).

While the main measurement consisted in recording the SRIR, we also recorded the voice of a person positioned behind
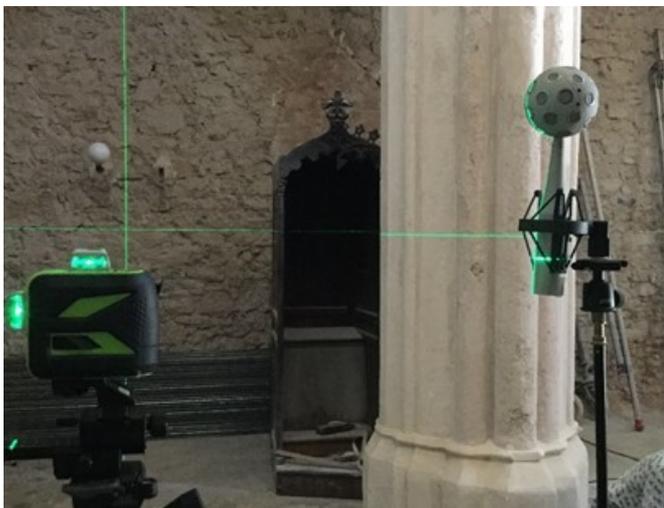
Figure 8. Vertical alignment of the Eigenmike 32 microphone, once positioned on the tripod in point MC, using the laser beams (N.D. de la Salette, Tourves).

Figure 9. The Eigenmike 32 and speakers positioned for reference measurement in an anechoic room.

the altar and pronouncing a given sentence while facing the chancel and then facing the nave. The idea was to open up on a qualitative measurement of the impact of the Vatican II council's reforms on the way the celebrant's discourse is perceived when he faces the people (current ritual) and when he faces the altar (the way it used to be prior to the reform). Additionally, footsteps of a person wearing systematically the same shoes and clothes and walking from the chapel's entrance to the altar were also recorded. Finally, 5 min of the soundscape inside and outside the chapel were recorded using a Zoom H3VR. The soundscape is related to the recording moment (day vs night, summer vs winter, etc.), but for practical reasons we could not record at the same time in the different places. To quantify the temporal influence, for one specific chapel we recorded the soundscape during 2-min each 30-min during 10 days and repeated the process twice (in February and in July). For these measurements we used the LabMaker AudioMoth, a low-cost acoustic monitoring device used for monitoring wildlife. We repurposed it here to monitor the soundscape with a twofold prospective aim, *i.e.* characterising (qualitatively) the variations of the soundscape over time and characterising the way the building acts as an acoustic "filter" (both the exterior and the interior soundscapes were recorded at the same time).

4. INITIAL DATA PROCESSING

Results of the acquisition step act as inputs processed independently at first and then pulled together again as combined outputs in a twofold way: end-user products and analytical overlays to the Potree 3D pointcloud renderer. This section focuses on describing the way raw data is processed, and a discussion on how the data is repurposed with regards to this or that objective is proposed in section 6.

4.1. Metric and visual data processing

The first step is to produce the 3D point clouds from the YI360 panoramas. The photogrammetric processing is done using the Agisoft Metashape suite, with MicMac as a prospective alternative solution. To do so named points acquired with the rangefinder are transferred into a csv-formatted list. They are then used as control points in order to scale the photogrammetric model (*cf.* Figure 10). 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.

The first add-ons introduced focus on viewing the input data resulting from the survey protocol itself, for each chapel, *i.e.* on one hand the DXF input extracted from the rangefinder and on the other hand the panoramas extracted from the 360 camera.

Concerning the former, the Leica S910 rangefinder allows the surveying of a maximum of thirty points in a row, outputted as one single DXF file. This is why three different survey protocols had to be conducted (grid of auditory devices, scaling and direct measures), with from nine to seventeen points surveyed for each. As a result a step of realignment of the DXF outputs was necessary. They are automatically realigned geometrically in the same frame, when loading each chapel inside the viewer.

This is done from a manually created text file that identifies the first two reference points for each of the three DXF files associated with each chapel. From these two points, all the other DXF points are readjusted by translations, then rotations, and finallly displayed in the renderer. This adjustment also makes it possible to display, in their correct positions, each thumbnail image associated with each DXF point. These images are recorded by the Leica distortion camera all along the protocol.

Concerning the panoramas extracted from the Yi 360 camera, they are materialised in the renderer by spheres (*cf.* Figure 6), textured with the stitched 360 panoramic photos. Spheres are positioned from a text file associating each image file name with its XYZ position extracted during the photogrammetric processing. They give access to the corresponding panorama (viewed using the panolens js library, *cf.* section 6).

Finally, other specific add-ons allow on-the-fly measurement on DXF points, the naming of these points, the visualization of the laser levels, and the representation, based on an on-the-fly computation of the approaching volume of chapels by voxel-based segmentation of the dense point cloud.

The renderer is used to display a complete point cloud, but also allows for user-monitored selections of sub-clouds (sections corresponding to the laser beams, segmented upstream − *cf.* Figure 10).

4.2. Acoustic data processing

For all measurements using the Eigenmike a few operations were applied to the recorded signals. First, the 32 input channels were encoded in the spherical harmonics domain using a VST plugin provided by mh acoustics. The Eigenmike allows an encoding up to the 4th order on the spherical harmonics basis, corresponding to a 25-channels signal. This 25-channels signal is then decoded in two ways: (i) for a restitution through a 42-loudspeakers; and (ii) for a restitution through headphones with a binaural conversion [22].

For measurements using other microphones (Neumann and Zoom), no specific operation was required after the acquisition.

Sounds reproduced were either directly the ones recorded on site (speech and footsteps) or the characterised SRIR convolved with monophonic stimuli (as mentioned in section 3.2).

In the metrology field in general, measurements are subject of uncertainties. In this work in particular, the instruments' positions could differ, and the variability of positions is quantified in the following. Tripod positions of points MC, EG, EC and ED were measured using the Leica Disto. The distance MC-EC over the 15 chapels was 557 +/- 5 cm. The standard deviation was 5 cm over the 15 chapels, corresponding to the centimeter-sized accuracy targeted. The mean distance was 557,
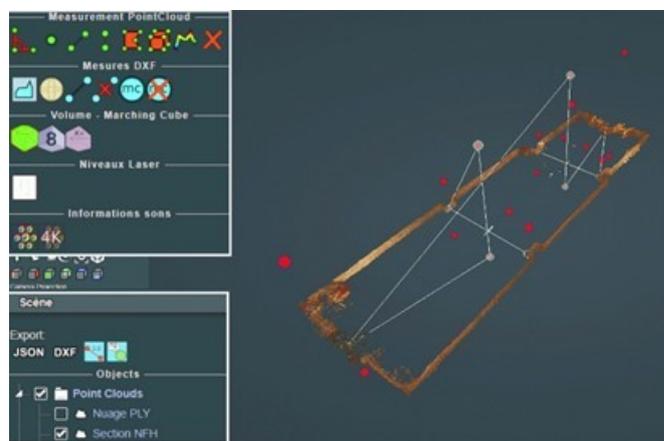

Figure 10. The polyline that corresponds to the scaling protocol (DXF outputted by the Leica Disto, twelve control points), represented inside the *Potree* point cloud renderer. Here only a sub-coud correspoding to the horizontal laser beam is shown.
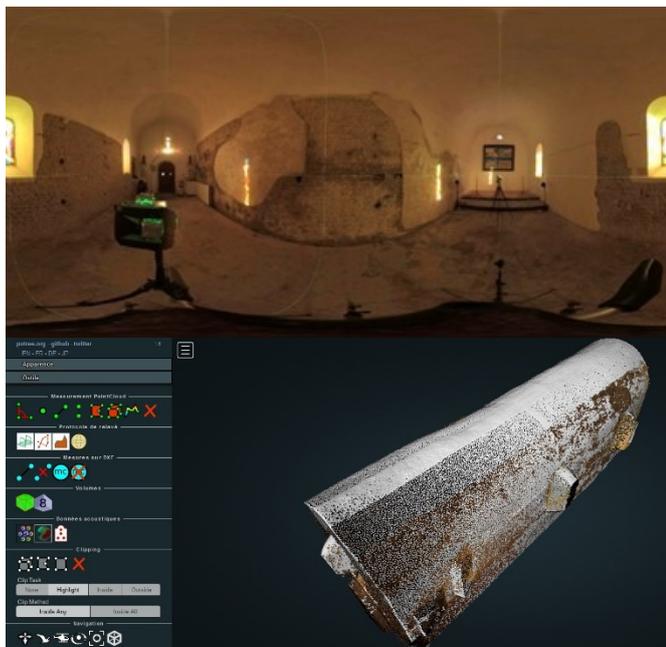
Figure 11. A case found to be critical for IBM (Saint-Roch chapel, Les Mées). Top, a stitch showing the predominance of white surface in the edifice. Bottom, the reconstruction, combining the photogrammetric model (brown points) and the laser scanner's output (white points).

slightly higher than the 550 cm targeted, but this disto measurement corresponded to the top of the tripods, while the instruments placed on top of it (loudspeaker and microphone) have a centimetre-thickness explaining this difference.

5. OBSTACLES AND LIMITATIONS

Experimenting the multimodal acquisition and processing the chain on fifteen different chapels shows that the overall method is sound, with some clear strong points. It is fast, reproducible, lightweight enough to be applied in remote sites. Ultimately it does correspond to what were our analytical needs: easy extraction of architectural features, close integration of acoustic and metric data, and a ground for comparisons and correlations. Briefly said the protocol allows for a quick multimodal acquisition, with the scaling of the photogrammetric model facilitated by the combined use of self-levelling levels, of a rangefinder and of a low-cost 360 camera.

Yet the experimentation also highlighted some difficulties that one may have to overcome at different steps of the protocol when applying it to different use cases.

First and foremost the method is tailored to a quite specific set of requirements (and particularly in terms of corpus) – a limitation *per se* (see section 2.1). With a closer look at the acquisition time, the method does requires a step of adaptation to *in situ* conditions such as lighting or congested spaces, factors that affect the photogrammetric acquisition. Having to cope with local conditions is natural and not surprising, but the quantity of potential factors of failure to take into consideration is significant. Typically pointing at intersections of the laser beams with the rangefinder may be seen as a very straightforward task, but in practice potential occlusions, surface conditions or angles of incidence have to be dealt with.

Furthermore, one should keep in mind limitations due to the photogrammetric process itself in conditions where contrasts lack. Thanks to the global efficiency and robustness of our

protocol, the complementary use of Terrestrial Laser Scanner (Faro Focus 3D) was required in only two of the fifteen sites, exemplifying the limitation of the geometric survey method. One chapel was textureless, mostly covered with white painting which is prohibitive for IBM (*cf.* Figure 11).

For the second chapel, the low resolution of the camera was suspected not to cover efficiently the most difficult chapel of the corpus (in terms of dimensions, volume and architectural complexity). At processing time, the quality of the 3D point clouds produced varies noticeably due to the above mentioned factors. It can be seen as good enough in a research programme that targets services like extracting dimensions or positioning instruments in space relatively to a systematic reference system, but it obviously is not good enough if targeting a fine-grain 3D mesh reconstruction.

So at the end of the day comes the question of how to rate the "quality" and "reproducibility" of the protocol, and the corollary issue of "in how does the choice of a low cost 360 camera impact the final results". As a provisional answer an experiment has been conducted in the "*fake chapel*" to evaluate the potential gain of upgrading our protocol with professional VR camera developing up to 12K panoramas instead of low-cost devices.

Aware of the main limitation of the method, the aim was to discuss the scalability of the setup to complex case-studies, regarding resolution vs. accuracy aspect. Briefly said, the insight of this qualitative comparison confirms the hypothesis made in section 2.2, that the acquisition strategy can improve results in a more significant way than the technological component itself. Our preliminary tests show that the density but also the uncertainty actually increases proportionally to the resolution. Therefore a better resolution of image sources doesn't improve intrinsically the quality and the range of the reconstruction without an efficient data acquisition protocol.

As shown in Figure 12, the acquisition canvas (i.e. camera position) seems to be more effective on the result than the resolution of the panorama itself.

Concerning acoustic measurements, there are also some limitations. The placement of the Eigenmike microphone is tricky, in particular because it is a spherical object. Because its correct placement is important if wanting to ensure comparability, its positioning might be time consuming. This is why, due to the time limitation on site (two sites per day constraint), the number of microphone positions was limited to two (MA and MC, *cf.* Figure 4). A higher number of measurement positions would have allowed a dynamic auralization, so as to reproduce an exploration through the chapel. Reproducing such an exploration requires a numerical simulation of the room, and necessitates precise information on acoustic properties of the building materials [23].

In addition, it has to be said that a number of factors related to conditions found in situ (typically congestion of spaces, or simply time of the day) do impact the direct, raw comparability of the data (and in particular of live recordings). These factors should act as a reminder that such data sets should not be over-interpreted, but rather be considered as means to reveal a qualitative acoustic identity of the sites, and to uncover general trends and patterns within the collection of sites.

6. DATA EXPLORATION AND REUSE

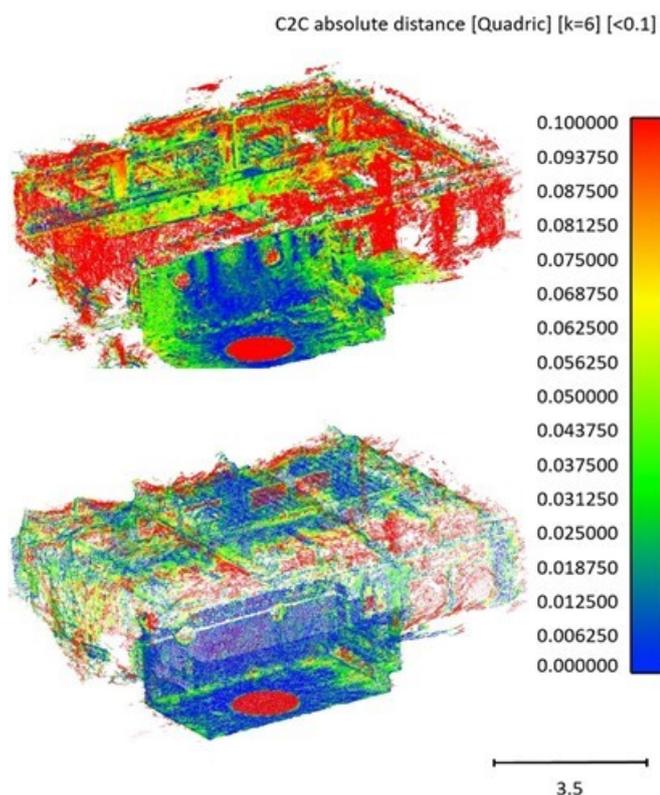The processing chains presented above lead to the production of a series of heterogeneous data sets:

C2C absolute distance [Quadric] [k=6] [<0.1]

	0.100000
	0.093750
	0.087500
	0.081250
	0.075000
	0.068750
	0.062500
	0.056250
	0.050000
	0.043750
	0.037500
	0.031250
	0.025000
	0.018750
	0.012500
	0.006250
	0.000000

3.5

Figure 12. Qualitative comparison tests conducted inside the "fake chapel" (see section 2.1): result of Cloud2Cloud (C2C) distance between laser scanner reference and (top) point cloud generated from a 12k panorama without pyramidal protocol and (bottom) compared to a pointcloud generated from the 5.7K panoramas with pyramidal sequence.

- - raw photographic material (unstitched hemispherical photograph – raw outputs of the 360 camera),
- - panoramas (stretched, little planet, and round views),
- - 3D point clouds, including localisations of the grid of the acoustic devices,
- - raw quantitative data (dimensions, volumes, XYZ coordinates of cameras and acoustic devices),
- - room impulse responses,
- - auralizations in 4 points (MA, MC, MD, MG).
- - live recordings,
- - quantitative acoustic indicators (e.g. reverberation time).

These outputs are integrated in various ways, in order to allow further exploration and cross-examination of the data sets, and the extraction of goal-bounded interpretations. Said differently, the above data sets are repurposed and combined with regards to usage scenarios that range from pattern analyses (e.g. proportions, reverberation, *etc.*) to the production of dissemination material. The following sub-sections illustrate the two main lines of development that are being followed, one targeting fine grain analysis and quantitative data correlation, one targeting perceptual analyses in contexts ranging from experimental setups to dissemination or edutainment activities.

6.1. Acoustic indicators

As mentioned before, the Spatial Room Impulse Response (SRIR) measures were primarily needed in order to process auralizations (see section 3.2). But they were also used to compute acoustic indicators [21], [24], [25], [26], listed hereafter. Acoustic indicators are a set of quantitative values that are used to characterise and differentiate spaces, roughly said on three

aspects (time-related indicators such as Reverberation time, tone-related indicators such as Bass Ratio, and space-related indicators such as Lateral Strength).

Most of the indicators were computed using the 0-order component of the spherical harmonics (omnidirectional component). On the overall, 13 such indicators have been computed, among which for instance the reverberation time (RT_{20} and EDT), the central time, the C50 clarity (see section 6.2), the acoustic strength, the Schroeder frequency, the spectral centroid, the bass ratio, the treble ratio, and an approximation of the Speech Transmission Index (without considering the background noise). Some indicators were computed using all spherical harmonic components to take into account the spatial: the InterAural Cross-Correlation (based on a binaural reduction), the Lateral Strength and the Lateral Energy Fraction. In the next steps of the research programme, these indicators will be used, in correlation with quantitative and qualitative architectural features, to characterise the particularities of each chapel, and to analyse the collection as such.

6.2. 2D/3D visualisation of acoustic indicators

Two of the quantitative acoustic indicators produced are calculated for each position of the Eigenmike microphone, and provide values that correspond to a specific angle in space. This gives an opportunity to try and spot differences in the way the sound hits the microphone depending on its origin (emission point) and on reflectance patterns inside the building. These indicators correspond to transmission-reception pairs (four speaker positions and two microphone positions), and correspond to two different methods.

The C50 clarity indicator (relation of the early IR – 50 first milliseconds – to the late IR – after 50 milliseconds) is calculated on the 32 channels of the Eigenmike microphone: one quantitative value for each capsule, and for each of the seven frequencies (*cf.* Figure 13). It is important to mention that according to [21], C50 is calculated from an omnidirectional room impulse response. In our case the 32 microphones of the Eigenmike (em32) are not omnidirectional, especially for high frequency bands. We therefore chose to calculate the C50 on each capsule of the em32 to highlight the early energy differences with respect to the microphone directions.
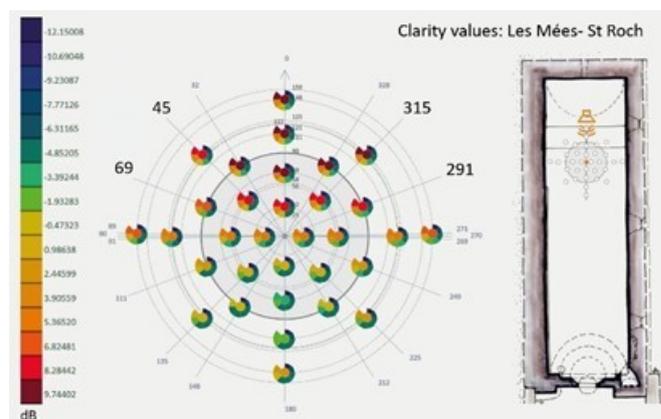

Figure 13. A 2D visualisation of the clarity values for one recording emitting tuple. Each symbol corresponds to one of the 32 capsules of the Eigenmike microphone, projected on a 2D plane. Sectors correspond to the seven frequencies, and colours to a quantitative value (in dB, distributed in a 16 values colour scale). Note here for instance a dissimilarity for the 8K frequency between values for angles 45 and 69 (top left) and for values 291 and 315, right) that cannot be explained by the layout – particularly simple and regular – of the edifice (Saint-Roch chapel, Les Mées).

Another space related indicator is a spatialized energy map (unrelated to the 32 channels) calculated using the PWD method (Plane Wave Decomposition) [27]: one value every 5 degrees of rotation, 2592 values for one microphone position. In this case the energy map evolves with time: twenty time-frames are considered for each microphone position.

Visualisation of such data sets raises two major and tricky challenges that go beyond the scope of this paper: having the analyst understand the relation of the data with space, and handling temporal aspects (sound evolving with time)- an ongoing research issue in the infovis (information visualisation) community [28].

What can be said at this stage is that several exploratory visual formalisms have been developed, both in 2D and 3D. In brief 2D solutions offer a better global and synthetic view (no occultation), but 3D solutions are more efficient in helping analysts spot the potential relation of a value or of a pattern to an architectural specificity in the chapels' interiors.

Concerning 3D solutions data is represented in the Potree-based renderer (see section 4.1) by spherical heat maps (*cf.* Figure 14), at the two microphone positions (MA and MC). This visualization enables to display the distribution of the reception data over 360 degrees, for both methods. It is interactive thanks to the two tools allowing to choose the transmission-reception pairs or to change the frequencies or the temporality according to the chosen methods.

Obviously what will come out of this part of the research is the main scientific added-value of the whole approach, but we are here still in an exploratory phase, with challenges that concern infovis methods rather that metrology *per se*.

6.3. Panoramas in the four recording points, with soundtracks

As mentioned before, two types of soundtracks are produced: basic live recordings (speech, footsteps and sound scape) and auralizations (simulations of how the same soundtrack would be perceived if played in the different recording positions of the various chapels). These outputs are used (and combined) inside online immersive panoramas corresponding to the four recording positions MA MC MG MD (*cf.* Figure 15). The panorama itself is viewed using the panolens.js JavaScript library,

Figure 15. An online panorama in point MC, with (bottom right) symbols used to give access to the various soundtracks. An exemple for Notre-Dame-du-Revest chapel (Esparron-de-Pallières).

in which users move from one position to another a bit like in 3D bubble worlds. On each position soundtracks are available and can be played, hence allowing listeners to spot differences as they would be perceived *in situ*.

Another somehow resembling set of outputs is a collection of interactive PDF flyers (*cf.* Figure 16) on which "little planet" views of interiors are combined with textual triggers that launch the soundtracks (6 audio tracks illustrating the acoustic identity of the building - clap, guitar, piano, steps, voice, exterior). These served as a basis for various dissemination initiatives or edutainment-like presentations of the research, typically in sound/space association games in which the audience must associate a building with its acoustics.

At the end of the day what can be said about the overall method (a combined acquisition procedure, parallel processing chains, and common data reuses or explorations) is that it does promote reproducibility and repurposability. In no way do we ignore or minor weaknesses such as metric accuracy – but on the other hand it has never been the core goal of our research knowing the severe economic constraints that one has to deal with in the context of minor heritage, and the impact on the objects themselves of multiple and undocumented transformation phases. Instead, we consider that minor heritage items can gain visibility and support when they are envisioned as part of a wider asset: a collection. Hence experimenting and better understanding how actors concerned could tailor the data acquisition, processing and reuse steps to collections of small-scale, minor heritage assets has been and remains the core result of the approach.
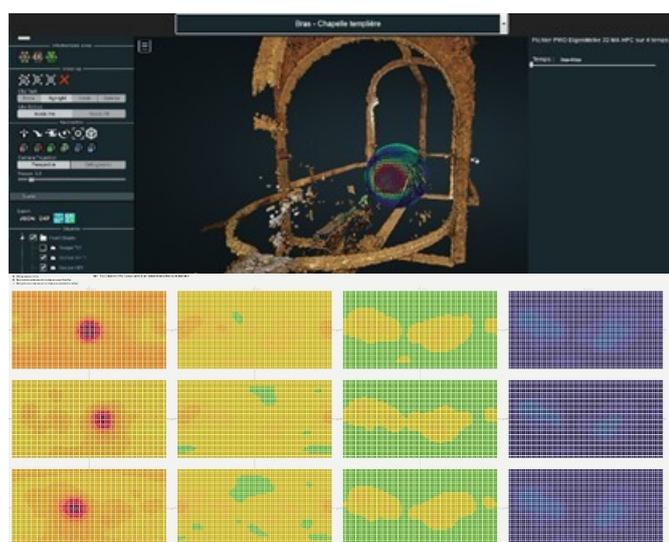

Figure 14. 2D and 3D visualisations of the PWD energy map (N.D de Bethléem, Bras): twenty time frames available in the 3D version, 4 time frames in the 2D version. Note for the latter differences between lines 2 and 3 (emission point EC or ED).

Figure 16. An example of an interactive sound + image PDF flyer with (on the left) text triggers that launch soundtracks for Notre-Dame chapel in Brue-Auriac.

7. PERSPECTIVES

At this stage there are still improvements that could be brought about in the data acquisition phase itself: typically a better monitoring of the lighting conditions in situ, or maybe an automatization of the rangefinder's movements or of the rotating arm and rotative head of the tripod used during the photogrammetric survey. The latter improvement would speed up the protogrammetric survey itself significantly.

Combining an acquisition using a 360 panoramic camera and a more detailed photogrammetric acquisition for this or that architectural detail could also be a sound perspective. Indeed fine geometric details such as engravings on memorial stones, mouldings on altars, or sculptures cannot be acquired with enough resolution and quality using a low-cost 360 panoramic camera. Surveying such details could be done using a more "traditional" photographic captor. The resulting high density cloud of points, focused on one specific area, could then be integrated to the general 3D reference system and offer a sort-of eagle eye view on parts of the building's decor.
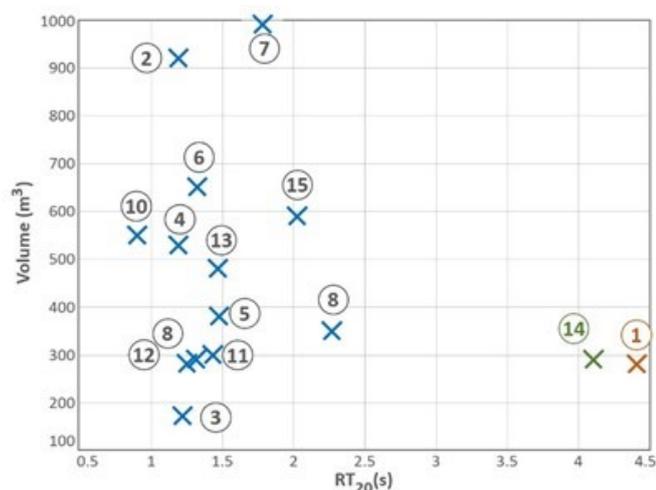

Figure 17. Top: Comparative analysis of data indicators (reverberation time vs. volume) across the collection (top): two items (n° 1 and 14) in the collection stand out significantly due to their high reverberation time and small volume. Bottom: The corresponding point clouds are bordered in colour in the bottom graphics: n° 1, red - N.D de Bethléem (Bras), n° 14, green - Saint-Roch chapel (Les Mées).

Later on in the processing chain, in terms of data fusion there is a very promising lead with the use of spherical depth map [29] combined with acoustical descriptors to enlarge cross-correlation analysis performed through image processing or signal processing algorithms [30].

More generally next steps are bounded by a backbone objective: better understanding, characterising such buildings and the way they are perceived, though vision and audition. This implies building on the interdisciplinary nature of the research, including in the analysis steps. Accordingly we currently launch a series of experiments aimed at exploiting 3D representations to position and analyse acoustic 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 [31], [32], etc. - as opposed to approaches where a 3D point cloud is analysed as such (segmentation, classification, etc.) [33]. Features can then be compared, trends spotted, exceptions raised and analysed, based on methods and practices from the infovis community. In that future effort, data extracted from traditional manual surveys (quantitative or qualitative) will complement dimensions and geometric features in order to widen the scope of differentiating factor when comparing chapels.

Concerning sound data, the next steps of this work is to use the collected data, to categorize and to distinguish the chapels in terms of acoustic descriptors and perceptual criteria. 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 source localisation protocol [34] 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. To take the acoustic simulations further, the integration of metric data acquisition, 3D point-cloud model estimation and acoustical measurements are of great interest. Indeed, acoustical simulation tools such as CATT or ODEON allow 12-DoF auralization of rooms based on 3D geometric models and impulse response measurements [23]. However, these tools are restricted to simple geometric models with a limited set of walls. Further work aims to derive simple geometric models, compatible with such tools, from complex 3D-point cloud models as suggested in [35].

8. CONCLUSION

This contribution reports on a research programme anchored in two prime concerns:
- considering interdisciplinarity as a mandatory requirement in the surveying of architectural interiors (during the co-design of the survey protocols, and in all subsequent phases of the operations, competences stemming from both architectural and acoustic studies were associated),
- tailoring the research's technological and methodological choices to the specific context of small-scale architecture (a collection of buildings with little prestige, often neglected or at risk). The specificity of the corpus under scrutiny undoubtedly shaped the overall survey and data processing strategy. In that sense, one of this research's originality is the effort to overcome the operational limits of a set of low-cost technologies.

The overall protocol intends to help actors to characterise and correlate acoustic and morphological features of heritage

architecture in a consistent way. Therefore it aims at opening up new analytical biases, built on the principles and potential of comparative analyses. As an example, Figure 17 shows cross modal representation of the studied collection, by representing the individuals as a function of their volumes V and reverberation times RT_{20}.

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 buildings (rural, often isolated chapels), with keeping the constraints linked to that type of heritage asset. The approach does open up new 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.

ACKNOWLEDGEMENT

REFERENCES

[1] L. Alvarez-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 vol 72, 2014, pp. 102-115.
DOI: 10.1016/j.buildenv.2013.10.015

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

[3] Boren, B. B, Acoustic simulation of J S Bach's Thomaskirche in 1723 and 1539, Acta Acustica vol 5, no. 14, 2021.
DOI: 10.1051/aacus/2021006

[4] L. Barazzetti, M. Previtali, F. Roncoroni, Can we use low-cost 360 degree cameras to create accurate 3d models? ISPRS TC II Mid-term Symposium "Towards Photogrammetry 2020", Riva del Garda, Italy, 4–7 06 2018. ISPRS Archives vol XLII-2, pp 69-75.
DOI: 10.5194/isprs-archives-XLII-2-69-2018

[5] M. Schütz, Markus, Potree: Rendering Large Point Clouds in Web Browsers. Ph D Thesis Vienna University of Technology, 2016.

[6] W. C. Sabine, Collected Papers on Acoustics. Harvard Univ. Press. Reprinted by Peninsula Publishing, Acoustical Society of America, Newport Beach, 1993, ISBN 9780932146601.

[7] Council of Arts Accrediting Associations 2009 Disciplines in Combination: Interdisciplinary, Multidisciplinary, and Other Collaborative Programs of Study CAAA Briefing paper. Online [Accessed 17 June 2022]
https://nast.arts-accredit.org

[8] J. Y. Blaise, I. Dudek, G. Saygi, Analysing citizen-birthed data on minor heritage assets: models, promises and challenges, International Journal of Data Science and Analytics, Springer Verlag, 2019, pp. 1-19.
DOI: 10.1007/s41060-019-00194-0

[9] L. T. Losè, F. Chiabrando, A. Spanò, Preliminary Evaluation of a Commercial 360 Multi-Camera Rig For Photogrammetric Purposes. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences vol. XLII-2, 2018, pp. 1113-1120.
DOI: 10.5194/isprs-archives-XLII-2-1113-2018

[10] C. Gottardi, F. Guerra, Spherical Images For Cultural Heritage: Survey And Documentation With The Nikon Km360. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences vol. XLII–2, 2018, pp. 385–390.
DOI: 10.5194/ISPRS-ARCHIVES-XLII-2-385-2018

[11] G. Fangi, R. Pierdicca, M. Sturari, E. S. Malinverni, Improving Spherical Photogrammetry Using 360° Omni-Cameras: Use Cases And New Applications. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences Vol. XLII-2, 2018, pp. 331-337.
DOI: 10.5194/isprs-archives-XLII-2-331-2018

[12] T. Luhmann, A historical review on panorama photogrammetry, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 34, 2008.

[13] L. Perfetti, C. Polari, F. Fassi. Fisheye Multi-Camera System Calibration for Surveying Narrow and Complex Architectures, ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XLII-2, 2018, pp. 877-883.
DOI: 10.5194/isprs-archives-XLII-2-877-2018

[14] L. Perfetti, C. Polari, F. Fassi, S. Troisi, V. Baiocchi, S. Del Pizzo, F. Roncoroni, Fisheye Photogrammetry to Survey Narrow Spaces in Architecture and a Hypogea Environment, Latest Developments in Reality-Based 3D Surveying and Modelling; MDPI: Basel, Switzerland, 2018, pp. 3-28.
DOI: 10.3390/books978-3-03842-685-1-1

[15] M. A. Gerzon, Ambisonics in multichannel broadcasting and video, Journal of the Audio Engineering Society, Vol. 33, no. 11, 1985, pp. 859-871.

[16] J. Daniel, Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia. Ph. D. Thesis, University of Paris VI, France, 2000.

[17] A. Farina, Simultaneous measurement of impulse response and distortion with a swept-sine technique, 108th AES Convention, Paris, France, 2000.

[18] S. Müller, P. Massarani, Transfer-function measurement with sweeps, Journal of the Audio Engineering Society, vol. 49, no. 6, 2001, pp. 443-471.

[19] A. Farina, A. Amendola, A. Capra, C. Varani, Spatial analysis of room impulse responses captured with a 32-capsule microphone array, 130th Audio Engineering Society convention, London, 2011.

[20] D. A. Dick, M. C. Vigeant, An investigation of listener envelopment utilizing a spherical microphone array and third-order ambisonics reproduction, The Journal of the Acoustical Society of America, vol. 145, no. 4, 2019, pp. 2795-2809.
DOI: 10.1121/1.5096161

[21] ISO 3382-1 (2009). Acoustics – Measurements of room acoustic parameters – Part 1: Performance parameters

[22] H. Moller, Fundamentals of Binaural Technology, Applied Acoustics, vol. 36, no. 3/4, 1992, pp. 171–218.
DOI: 10.1016/0003-682X(92)90046-U

[23] B. N. J Postma, B. F. G Katz, Perceptive and objective evaluation of calibrated room acoustic simulation auralizations, The Journal of the Acoustical Society of America, vol. 140, no. 6, 2016.
DOI: 10.1121/1.4971422

[24] A. Gade, Acoustics in Hall for Speech and Music, Springer Handbook of Acoustics, ISBN 978-0-387-30446-5, 2007, p. 301.
DOI: 10.1007/978-0-387-30425-0_9

[25] F. A. Everest, K. C. Pohlman, Master Handbook of acoustics, 5th edition, McGraw Hilll, New York, 2009, ISBN 9780071603331.

[26] G. Peeters, A large set of audio features for sound description (similarity and classification) in the CUIDADO project (CUIDADO IST Project report), 2004.

[27] E. G. Williams, Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography, Academic Press, New York, 1999.
DOI: 10.1016/B978-0-12-753960-7.X5000-1

[28] W. Aigner, S. Miksch, H. Schumann, C. Tominski, Visualization of Time-Oriented Data, Human-Computer Interaction Series, Springer-Verlag, London, 2011, ISBN: 978-0-85729-079-3.

[29] N. Zioulis, A. Karakottas, D. Zarpalas, P. Daras, Omnidepth: Dense depth estimation for indoors spherical panoramas, Proc. of

the European Conference on Computer Vision ECCV, Munich, Germany, 2018, pp. 448-465.
DOI: 10.1007/978-3-030-01231-1_28

[30] A. Pamart, D. Lo Buglio, L. De Luca, Morphological analysis of shape semantics from curvature-based signatures, Digital Heritage Conference, Granada, Spain, 2015, pp. 105-108.
DOI: 10.1109/DigitalHeritage.2015.7419463

[31] J. Y. Blaise, I. Dudek, Identifying and Visualizing Universal Features for Architectural Mouldings, IJCISIM, vol. 4, 2012, ISSN 2150-7988, pp. 130-143.

[32] M. A. Cohen, Conclusion: Ten Principles for the Study of Proportional Systems in the History of Architecture, Architectural Histories, vol. 2, no. 1, 2014.
DOI: 10.5334/ah.bw

[33] E. Grilli, F. Menna, F. Remondino, A review of point clouds segmentation and classification algorithms, ISPRS Archives Volume XLII-2/W3, 2017, pp. 339-344.
DOI: 10.5194/isprs-archives-XLII-2-W3-339-2017

[34] 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, Paris, France, 6-7 Sep 2019, pp. 67-72.
DOI: 10.25836/sasp.2019.15

[35] L. Aspöck, M. Vorländer, Room geometry acquisition and processing methods for geometrical acoustics simulation models, Proc. of the EuroRegio 2016, Porto, Portugal, 13–15 June 2016.