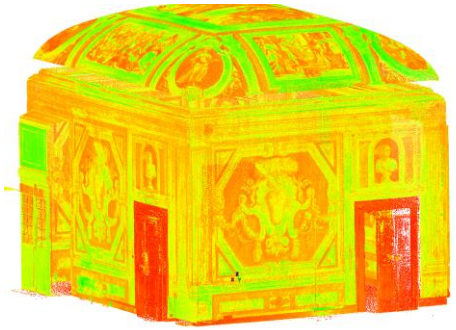
Dataset and evaluation

Why?

- An **open benchmark dataset** for the evaluation of 3D reconstruction algorithms is currently **not available** to the community.
- Most of the algorithms are tested on synthetic data and toy examples. Real data is used for qualitative evaluation only.
- **Accurate Ground Truth** from real data is often difficult to provide.

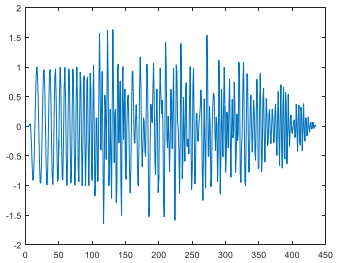
3DRoom dataset

The dataset provides:



3D laser scan of the room with high resolution point cloud acquistion using a Leica© C10 laser scanner: **accurate ground truth for planar reflectors positions.**

Accurate 3D position of sources and microphones using a Vicon MOCAP system with xx BONITA camera with localisation accuracy < 0.5 mm



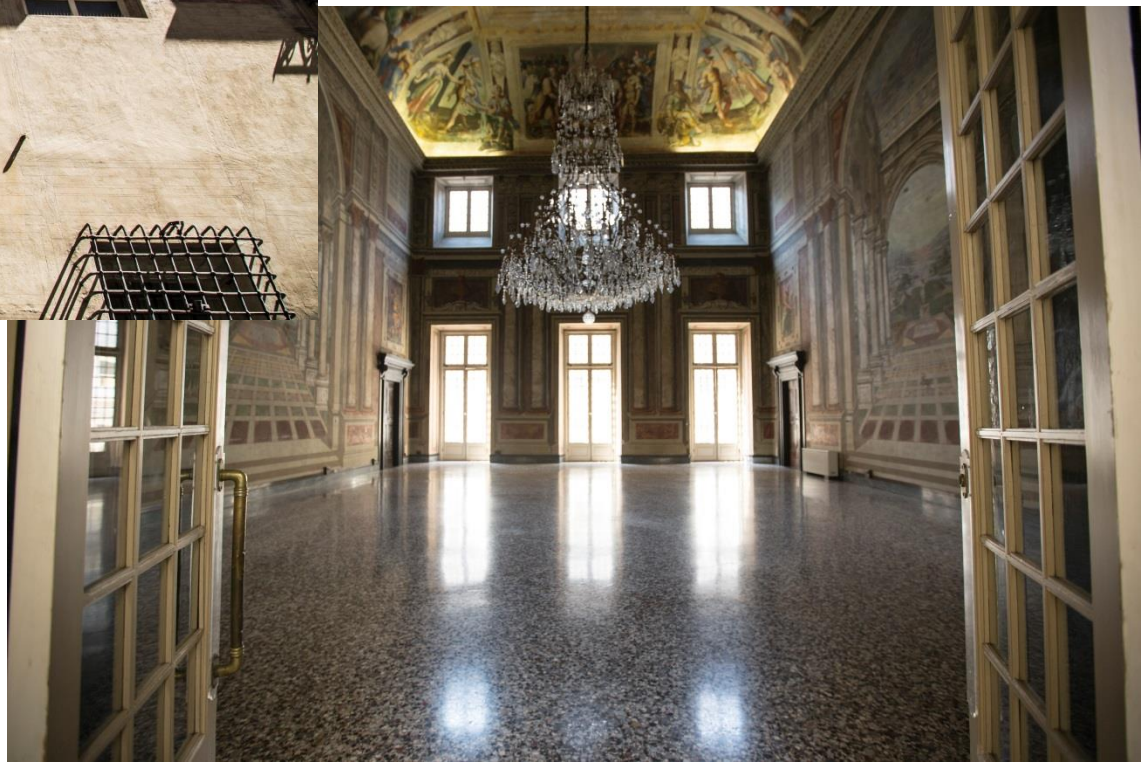
Large set of recorded signals.

- Synthetic: chirp, white noise.
- Natural: knock, scratch.
- Speech

Location: Genova, Italy

Via Garibaldi, Palazzo Spinola

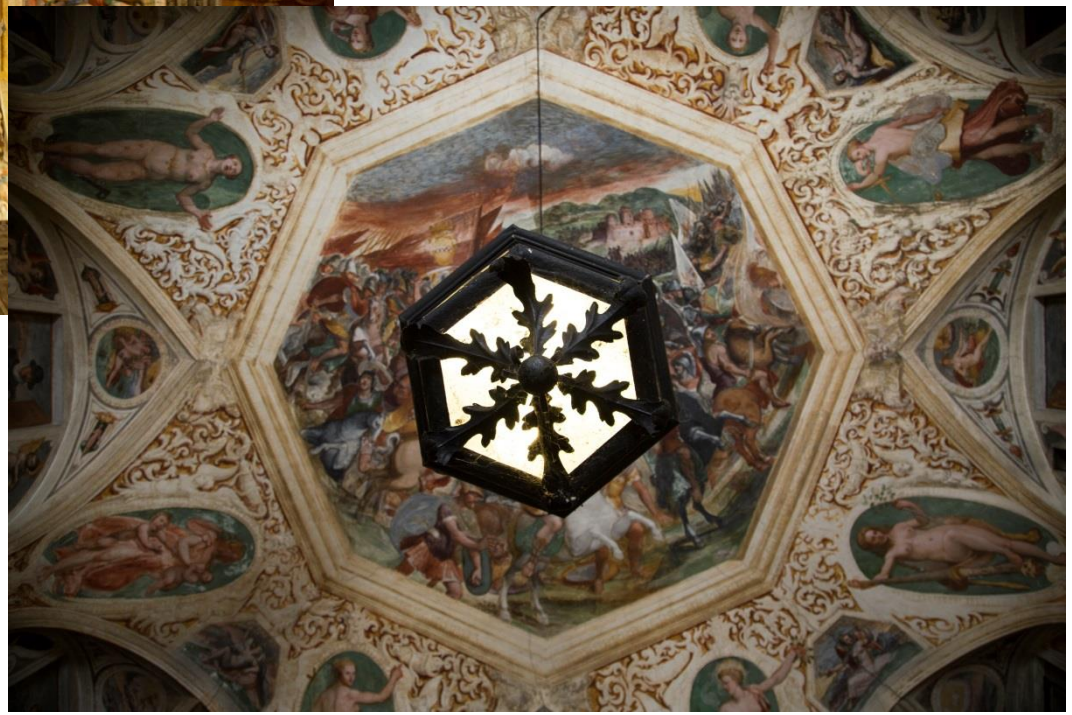
Palazzo Spinola, Via Garibaldi 5
UNESCO heritage.



XVI century palace, mansion of one of the leading families of genovese oligarchy.

Rooms

Outstanding frescos from XVI century Genovese school.



Audio Setup - Hardware

- **Rectangular room** of about 7 x 7 x 6 m with a vaulted ceiling and a wooden floor.

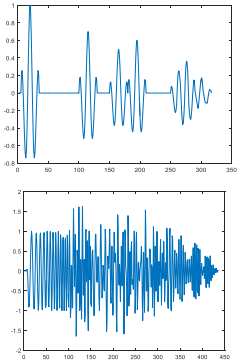
- **12 RF microphones** displaced in the central part of the room (about 3 m x 3 m square) **in 2 different spatial configurations**. Each microphone hangs from a tripod with different heights (from 20 cm to 2 m).

- Synchronous acquisition and ADC conversion by a National Instrument PXI ©.

- One small speaker, nearly omnidirectional VEHO360© (about 1.5 cm radius tweeter) moved in **20 different locations** of the room.

Audio Setup - TX signals

- Large set of different signals:
  - Synthetic: chirp, white noise
  - Natural: knock, scratch
  - Speech



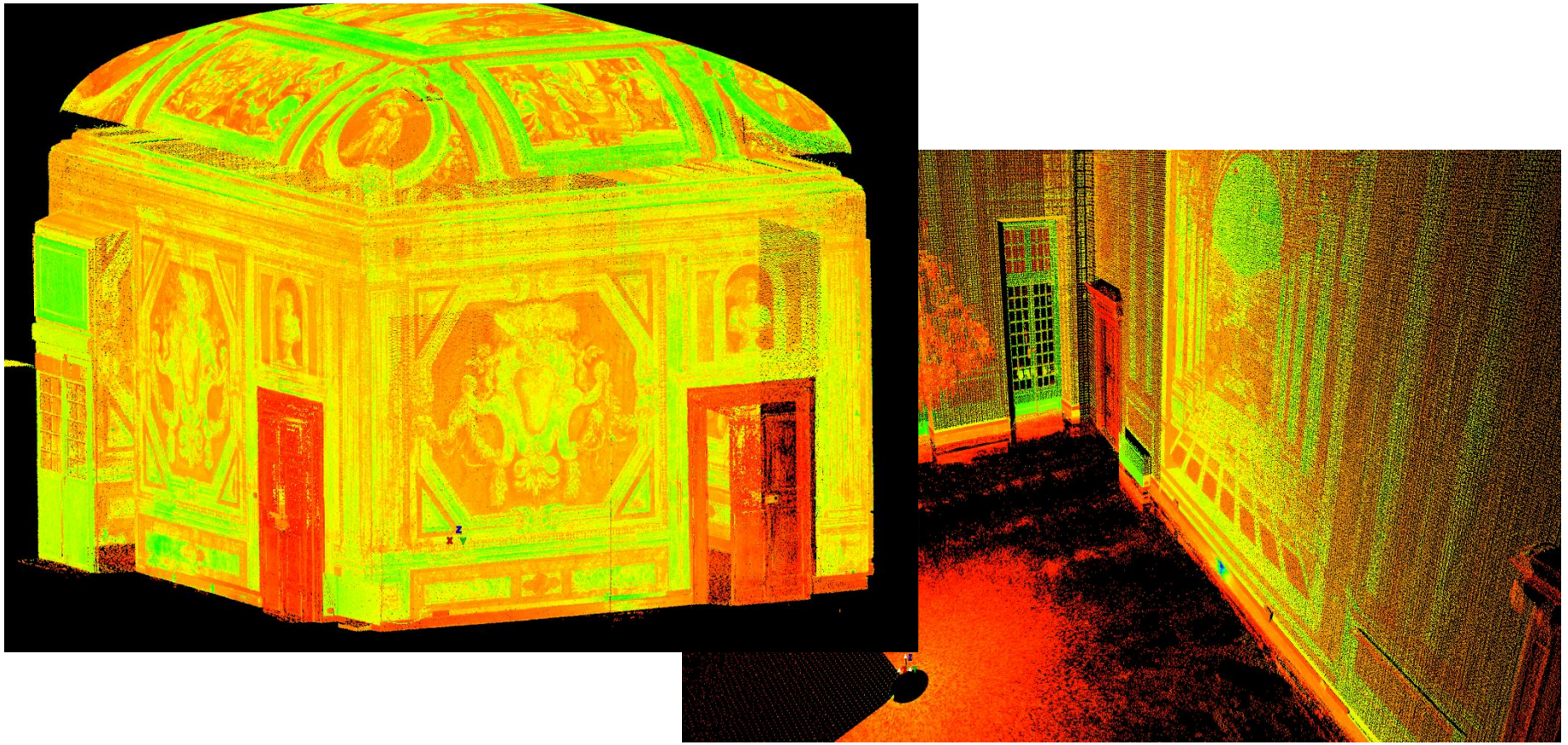
- For each of the 20 source locations a different TX signal is transmitted.
- For coded signals (chirps) a matched filter is also provided based on the chirp signal emitted by the loudspeaker.

Details on the 8 TX signal sets

- Synthetic:
  - chirp 1 (0.1 s @44 kHz);
  - chirp 2 (1 s   @44 kHz);
  - chirp 3 (5 s @16 kHz);
  - white noise1 (0.1 s @44 kHz);
  - white noise2 (1 s 44 kHz);
- Real
  - Knock: two staplers crunching together (about 5s @16 kHz);
  - scratch: plastic bag motion (about 5 s @ 16 kHz).
- Speech:
  - Male speech (about 5 s @ 16 kHz).

Ground Truth

GT room geometry



High resolution point cloud acquistion of the environment geometry by a Leica© C10 laser scanner: **accurate ground truth for planar reflectors positions.**

Microphones and source GT positions

Vicon© Motion Capture System: 8 BONITA© cameras with a 4 x 4 x 2 meters recording area.

Ground truth 3D position of microphone and sources acquired by the motion capture system: markers placed on each microphone and on the loudspeaker position. Reference system registered with the 3D scan.

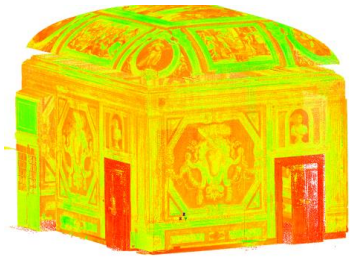
Evaluation of room geometry estimation

- Evaluate the impact of real rooms in the estimation of the 3D geometry (walls, mic/sources positions);
- Evaluate different real signals and how complex is the inference of TOAs;
- Find discrepancies in the image source model and possible improvements from the ideal case.

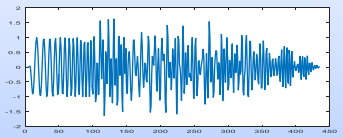
Still missing:

- Non-convex rooms: in this case we cannot use direct path to calibrate mics/sources;
- Empty rooms, evaluate the effect of noise given by furniture, people in the environment

Conclusions

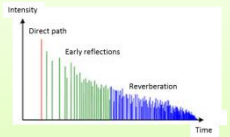


$$y_m^n(t) = h_m^n(t) * s^n(t) + \nu_m(t)$$



Signal acquisition

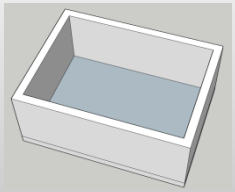
$\{\tau_e^n\}\{\tau_m^o\}\{\tau_{m0}^n\}$



Time of Flight (TOF) estimation

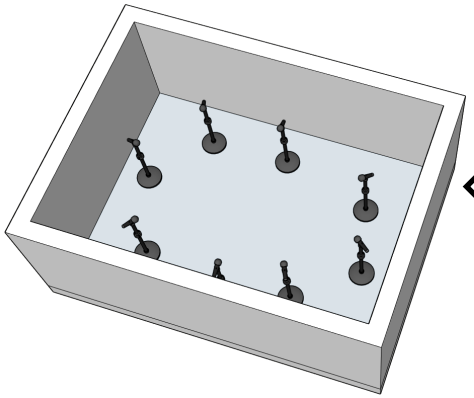
Mic/Sources calibration

$\{\mathbf{s}_m\}\{\mathbf{b}^n\}$



Room estimation

$\{\mathbf{r}_k\}$



For the next...

We showed that room reconstruction is feasible but for still for “simple rooms”. We can do better in:

- Including TOAs from higher order reflections, especially for validation.
  Tervo & Korhonen, “Estimation of reflective surfaces from continuous signals.” ICASSP 2010 .
- Dealing with missing data i.e. undetected TOAs.
  Dokmanić et al., "Acoustic echoes reveal room shape.“, PNAS 2013
  M. Crocco et al. “A Bilinear Approach to the Position Self-Calibration of Multiple Sensors,” IEEE TSP, 2012.
- Non-convex rooms – this assumption is at the basis of many methods.
  Dokmanić et al., “Source Localization and Tracking in Non-convex Rooms.”, ICASSP 2014 .
- Still most of the methods are based on the knowledge of the number of reflectors!

- Global optimization: more efficient and robust approaches to refine all the variables jointly.
  Crocco et al., “Towards Fully Uncalibrated Room Reconstruction with Sound”, EUSIPCO 2014

@ICASSP 2016

Tuesday, March 22

16:00 - 18:00

AASP-L1: Room Acoustics and Geometry Estimation

AASP-L1.1: **ROOM GEOMETRY ESTIMATION FROM ACOUSTIC ECHOES USING GRAPH-BASED ECHO LABELING**

Ingmar Jager; *Delft University of Technology*

Richard Heusdens; *Delft University of Technology*

Nikolay D. Gaubitch; *Delft University of Technology*

AASP-L1.2: **ACOUSTIC SIMULTANEOUS LOCALIZATION AND MAPPING (A-SLAM) OF A MOVING MICROPHONE ARRAY AND ITS SURROUNDING SPEAKERS**

Christine Evers; *Imperial College London*

Alastair H. Moore; *Imperial College London*

Patrick A. Naylor; *Imperial College London*

AASP-L1.3: **ECHOSLAM: SIMULTANEOUS LOCALIZATION AND MAPPING WITH ACOUSTIC ECHOES**

Miranda Krekovic; *École Polytechnique Fédérale de Lausanne*

Ivan Dokmanić; *École Polytechnique Fédérale de Lausanne*

Martin Vetterli; *École Polytechnique Fédérale de Lausanne*

AASP-L1.4: **MULTICHANNEL IDENTIFICATION OF ROOM ACOUSTIC SYSTEMS WITH ADAPTIVE FILTERS BASED ON ORTHONORMAL BASIS FUNCTIONS**

Giacomo Vairetti; *Katholieke Universiteit Leuven*

Enzo De Sena; *Katholieke Universiteit Leuven*

Michael Catrysse; *Televic N.V.*

Søren Holdt Jensen; *Aalborg University*

Marc Moonen; *Katholieke Universiteit Leuven*

Toon van Waterschoot; *Katholieke Universiteit Leuven*

AASP-L1.5: **FIRST ORDER ECHO BASED ROOM SHAPE RECOVERY USING A SINGLE MOBILE DEVICE**

Tiexing Wang; *Syracuse University*

Fangrong Peng; *Syracuse University*

Biao Chen; *Syracuse University*

AASP-L1.6: **ESTIMATING DIRECT-TO-REVERBERANT RATIO MAPPED FROM POWER SPECTRAL DENSITY USING DEEP NEURAL NETWORK**

Yusuke Hioka; *University of Auckland*

Kenta Niwa; *NTT Corporation*

Tutorial material & dataset

Alessio Del Bue and Marco Crocco

Visual Geometry and Modelling Lab

Istituto Italiano di Tecnologia (IIT)

Genova, Italy

Check: www.iit.it/research/lines/visual-geometry-and-modelling
for updated material and instructions on how to download the dataset

Email: alessio.delbue@iit.it