BeBeC-2020-D30

A LOW-LATENCY ACOUSTIC CAMERA FOR TRANSIENT NOISE SOURCE LOCALIZATION

Joëlle Fréchette-Viens[1], Nicolas Quaegebeur[1] and Noureddine Atalla[1,2]
[1]GAUS, Department of Mechanical Eng., Université de Sherbrooke
2500 Boul. de l'Université, J1K 2R1, Sherbrooke, Canada
[2]Mecanum Inc.

Abstract

In the last few years, acoustic cameras have become increasingly compact and performant. In order to compete with existing products, new contenders on the market must provide high-end features at low costs. This paper introduces a low-latency noise acoustic camera prototype for transient source localization, in partnership with the company Mecanum. It features a 10 channels I^2S MEMS microphone antenna coupled with time-domain localization techniques based on GCC-algorithm, an integrated depth-sensing camera and a GPU-based architecture on a mobile board (Nvidia Tegra X2). First, the choices of components are detailed. Then, the architecture and development details are provided. A few performance examples and preliminary experimental results are also shown. Finally, future challenges in the development of the product are addressed, as this project is a work in progress.

1 INTRODUCTION

Noise reduction is an important objective for modern societies. Deafness problems related to prolonged exposure to noise are not only a public health issue, but are increasingly becoming an economic issue. In fact, in 2018 in Quebec, ear, mastoid or hearing disorders represented 70.2% of occupational disease cases opened and accepted for compensation at the provincial occupational health and safety office [5]. But people may also be exposed to noise in their everyday life whenever they are in the presence of means of transportation or household, resulting in a deterioration of quality of life. Many consumers will therefore make noise a selection criterion when buying products, especially in the automotive field [6]. In any case, in order to be able to effectively reduce environmental noise, it is necessary to be able to diagnose its causes, or more precisely determine and reduce the impact of problematic noise sources.

A very common method in the field of sound source localization is the use of microphone arrays [1]. Many variations have been developed over the past decade to adapt the method to the requirements of various environments, whether in terms of microphone number, type and location, antenna size, or the type of electronic platform used for data acquisition and processing. However, commercial microphone arrays are often expansive, bulky and require to be handled by trained acoustic technicians, since care must be taken to ensure proper microphone placement and calibration, and specialized processing software often needs to be used to obtain a valid imaging result. In addition, technical knowledge is very often required to understand the results properly.

More and more microphone arrays are now being combined with cameras to form "acoustic cameras". These offer the advantage of making diagnosis easier for untrained people by superimposing the results of acoustic imaging on the image filmed by a camera. However, the user still has to specify the distance of the acoustic source from the camera, which is a technical calibration step. The goal of this project is to facilitate access to acoustic cameras for small and medium-sized enterprises by designing a compact, inexpensive and easy-to-use product. More specifically, in partnership with the acoustic company Mecanum, we aim to create an acoustic camera that can perform real-time measurements by automatically detecting the depth of the imaging points grid. The camera could thus be used to localize transient noise sources. This paper presents the first developments of this project. Section 2 details the conception requirements based on previous research, section 3 describes the integration of the different components and the current challenges we are facing, and section 4 shows a few preliminary results with our prototype.

2 CAMERA CONCEPTION

Some attempts have been made previously in the development of acoustic cameras with depth detection. For example, Iyama et al. [3] have used a Microsoft Kinect to determine clusters corresponding to objects identified by the camera. The user can instruct the camera to focus on one or more specific clusters and perform imaging only around the centroid of each cluster. This method is well suited when used for surveillance cameras, for instance, but is less appropriate for locating sound sources in a complex environment because of its poor space resolution.

Some studies have also demonstrated the use of an artificial depth-sensing acoustic camera. Ding et al. [2] performed imaging on several 2D planes parallel to their camera plane and determined the depth of the source from the maximum value found. Thus, they do not use depth detection as such. Moreover, the camera still requires input from the user, who must specify the depth interval to be scanned and the distance increment between planes.

In terms of real-time solutions, Vanwynsberghe et al. [7] have shown that it is possible to use a GPU to perform very fast acoustic imaging. In our case, the imaging algorithm used is defined in the time-domain and the Generalized Cross-Correlation with Phase-Transform (GCC-PHAT) solution is retained [4]. This algorithm requires the definition of a measurement grid and the computation of the distance between each point of the measurement grid and each microphone of the antenna. Since we aim at using depth detection to determine the points on the measurement grid, and since the coordinates of these points fluctuate with each video frame, we have to be able to recalculate the distances in less than 50 ms in order to achieve a target of 20 frames per second (fps). An architecture based on the use of a GPU is promising to achieve

this goal, as it is a simple calculation that must nevertheless be repeated for each pixel and for each microphone pair; the ideal type of calculation for a GPU.

For this reason, it made perfect sense to choose a depth sensing camera running on the GPU of a given platform. This is specifically what the Stereolabs ZED camera does. This camera was designed to be used in virtual reality devices. It has two "eyes" that can recreate the 3D point cloud of an environment using the difference between the images provided by each camera. The calculations are performed on the GPU of a compatible platform to which the camera is connected via a USB 3.0 bus. The point cloud can then be retrieved by the user via a Python library developed by the company. Its maximal frame rate depends on the resolution chosen by the user; for instance, a 720p resolution comes with a maximal frame rate of 100 fps, which is five times greater than our target.

The platform selected to manage the camera is NVIDIA's "computer-on-a-chip" Jetson TX2. This supercomputer has a 256-core GPU architecture, 8 GB of RAM and runs on an adapted version of Ubuntu. It can also accommodate up to 6 I^2S inputs, but for now, as we are currently working with the development version of the board, only two I^2S channels are easily accessible via GPIO headers. In these inputs, we are using chains of up to 16 ICS-52000 microphones from Invensense. These are digital MEMS microphones that can be used in TDM mode that theoretically allows to daisy-chain up to 16 microphones over a single channel. However, we have designed a modular PCB design that restricts us to a maximum of 11 microphones per channel at first, because the clocks become too noisy when the microphones are too far apart.

3 CAMERA DEVELOPMENT AND CHALLENGES

Figure 1 represents the desired steps and relations between the camera components to obtain an image. The duration of each step of the acquisition is detailed in Table 1. These numbers were obtained with one channel of 10 microphones recording 1024 samples at a time. The imaging is done on a grid of 672×376 pixels. The total computation time for a frame is 336 ms, which is visibly greater than the target of 50 ms, and would be even greater if more microphones were used. However, there are a few solutions proposed to reduce the computation time. First, some specific calculation steps (denoted by a * in Table 1) could be performed on the board's GPU rather than on the CPU in future developments. Also, the acquisition of the microphone/camera data, as well as the GCC calculations are done sequentially for now. These operations could eventually be performed in separate processes using multi-threading, as they are intrinsically independent. This solution would also allow for a better synchronization of the RGB image and the acoustic image.

Another challenge of this project is the synchronization of the different microphone channels. Digital MEMS require a fast clock (SCLK) and a slow clock (WS). The fast clock indicates when to write each bit of data, while the slow clock indicates which microphone in the channel should be writing to the data bus. These clocks are generated by the board, but are not the same from channel to channel. A "master" channel must then be designated, and its clocks have to be physically linked to those of all the other channels, which then become its "slaves". It is also necessary to perform a whole gymnastics of modification of the pinmuxes at the hardware level internally and to route the different sound cards appropriately to recombine the channels together at the software level. For now, the prototype used to obtain the results in the next section uses only one channel of 10 microphones.

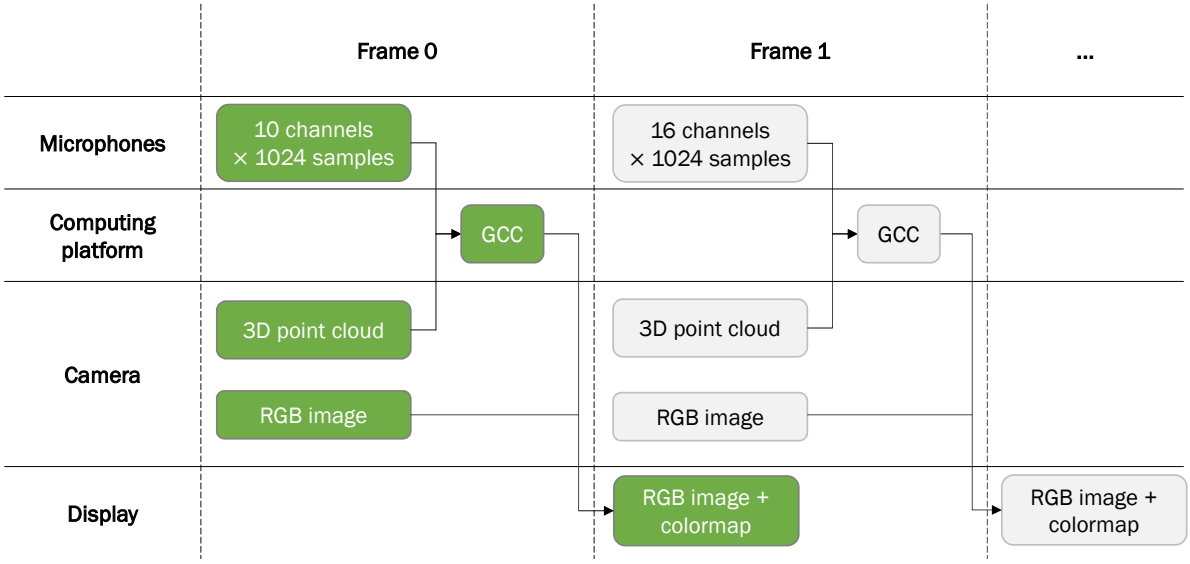


Figure 1: Acquisition of an image. Each frame must be completed in less than 50 ms to achieve a frame rate of at least 20 fps.

*Table 1: Computation time for each step of an image acquisition. A * indicates that this step would take less time to complete if performed on the GPU.*

Step	**Duration (ms)**
Microphone data acquisition	21
Camera data acquisition	5
*Calculation of the distance between each point of the measuring grid and each microphone	105
*Calculation of correlations between each single pair of microphones	150
*Conversion of the results to colors and addition to the RGB image	55
TOTAL	336

4 MEASUREMENT RESULTS

Some preliminary measurements were made with the prototype. The microphones were mounted on a flat screen and positioned as shown in Fig. 2. The measurements were made in an anechoic room using two loudspeakers located at two different distances and driven by two uncorrelated white noise. The noise sources were calibrated to produce a Sound Pressure Level (SPL) of 64 dB at the center of the camera.

Figure 3 represents the depth map of the room as seen by the ZED camera. The speakers are detected at the correct distance from the camera with a very small error (below 0.01 m). The

depth of some pixels (white contours in Fig. 3b) could not be measured by the camera due to irregularities, and normally occurs close to the contours of objects. For these grid points, the GCC calculations are simply not performed.

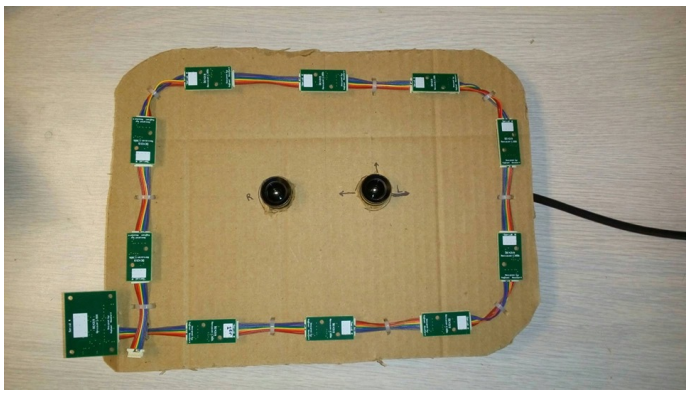
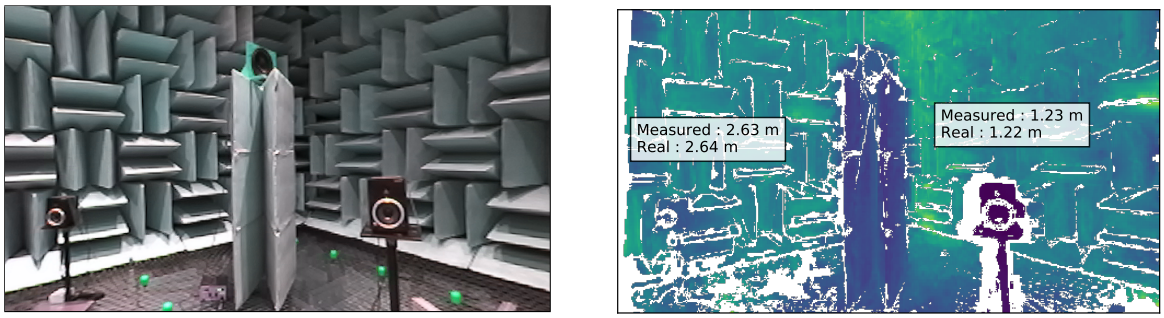


Figure 2: Prototype of the camera. The 10 microphones PCBs are visible in green. The PCB in the bottom left links the microphones to the Jetson board (not shown in photo). The two eyes of the camera are visible in black in the center of the picture.



(a) *(b)*

Figure 3: (a) Experimental setup. Only the left and right speakers were used. (b) Depth map of the setup as rendered by the ZED camera. Distances to the center of the speakers as measured by the camera are indicated; real values were measured with a tape.

Figure 4 shows the source localization results when different combinations of speakers are active. It can be seen that the camera detects accurately the presence of each source, even when the two speakers are driven at the same time. In this case, a 0.1 dB discrepancy between both sources is measured, confirming the sensitivity of the imaging. This shows that the depth camera allows the user to detect the presence of two sound sources at different distances without their location having to be specified to the algorithm.

In the following steps of the project, the distance between each source and the microphone array will be compensated, allowing the estimation of the source level at its origin. Moreover,

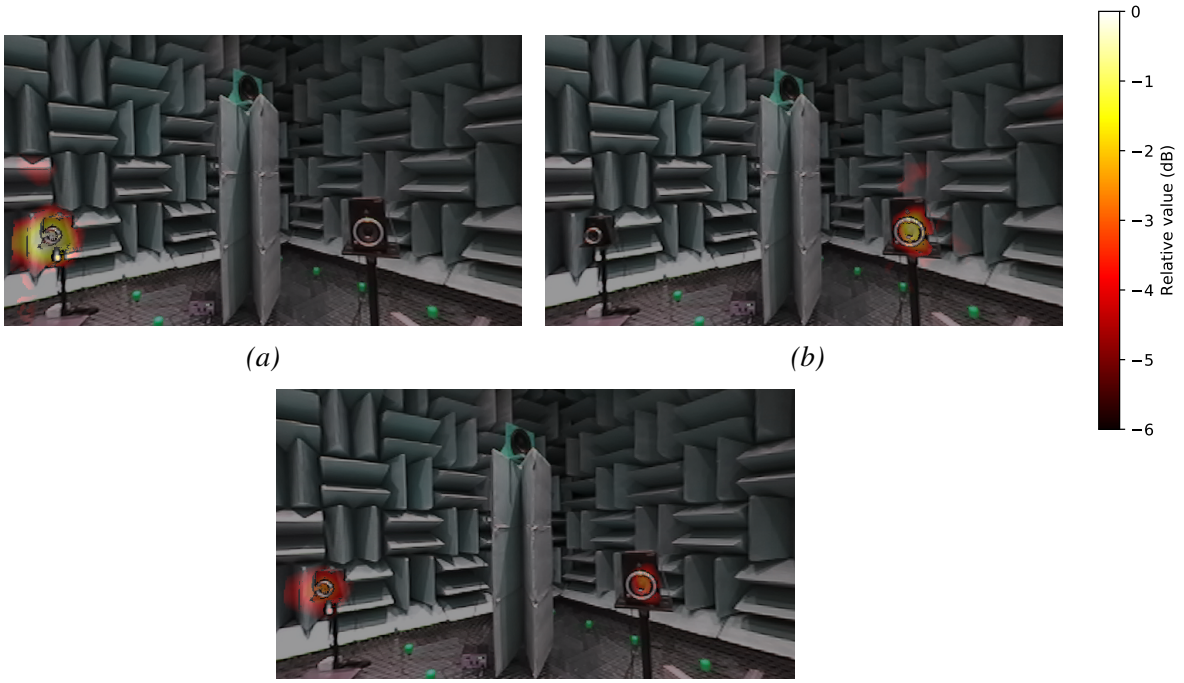


Figure 4: Image obtained by the acoustic camera when (a) only the left speaker is driven (b) only the right speaker is driven (c) both speakers are driven.

the correlation between source level and 3D image plot may enable new possibilities in terms of 3D mapping of enclosed spaces or for acoustic characterization of complex noisy environments.

5 SUMMARY

The developed acoustic camera prototype is compact, cost effective and works autonomously. It can measure about three frames per second and identify the presence of noise sources without the need for the user to enter any parameters. Many challenges will have to be solved for the rest of this project, but these first steps in the development suggest that it should be possible to reach the targeted fps by configuring multiple I^2S channels in order to increase the number of microphones, performing the calculations in separate processes and properly exploiting the board's GPU in order to increase the number of frames per second. When this is achieved, transient noise sources could be localized effectively as the camera would be able to continuously spot them using its depth map.

6 ACKNOWLEDGEMENTS

This research was supported by the Fonds de recherche Nature et technologies (FRQNT) and the Natural Sciences and Engineering Research Council of Canada (NSERC). Many thanks to

Raphaël Bouchard for the conception and assembling of the microphone PCBs, and to Lucas Carneiro for the help with the measurements.

REFERENCES

[1] P. Chiariotti, M. Martarelli, and P. Castellini. "Acoustic beamforming for noise source localization – reviews, methodology and applications." *Mechanical Systems and Signal Processing*, 120, 422–448, 2019. ISSN 08883270. doi:10.1016/j.ymssp.2018.09.019.

[2] H. Ding, Y. Bao, Q. Huang, C. Li, and G. Chai. "Three-dimensional localization of point acoustic sources using a planar microphone array combined with beamforming." *Royal Society Open Science*, 5(12), 181407, 2018. ISSN 2054-5703, 2054-5703. doi:10.1098/rsos.181407. URL https://royalsocietypublishing.org/doi/10.1098/rsos.181407.

[3] T. Iyama, O. Sugiyama, T. Otsuka, K. Itoyama, and H. G. Okuno. "Visualization of auditory awareness based on sound source positions estimated by depth sensor and microphone array." In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2014. URL https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6942814.

[4] C. Knapp and G. Carter. "The generalized correlation method for estimation of time delay." *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 24(4), 320–327, 1976. ISSN 0096-3518. doi:10.1109/TASSP.1976.1162830.

[5] D. Lamarche and J. Aubin. "Statistiques annuelles 2018." Technical report, Commission des normes, de l'équité, de la santé et de la sécurité du travail du Québec, 2019.

[6] M. J. M. Nor, M. H. Fouladi, H. Nahvi, and A. K. Ariffin. "Index for vehicle acoustical comfort inside a passenger car." *Applied Acoustics*, 69(4), 343–353, 2018. ISSN 0003682X. doi:10.1016/j.apacoust.2006.11.001.

[7] C. Vanwynsberghe, R. Marchiano, F. Ollivier, P. Challande, H. Moingeon, and J. Marchal. "Design and implementation of a multi-octave-band audio camera for realtime diagnosis." *Applied Acoustics*, 89, 281–287, 2015. doi:10.1016/j.apacoust.2014.10.009.