# SONIFICATION OF 3D POINT CLOUDS FOR SUBSTITUTION OF VISION BY AUDITION FOR BLIND USERS

*Louis Commère*
Université de Sherbrooke
NECOTIS GEGI, CIRMMT
Sherbrooke, Canada
louis.commere@usherbrooke.ca

*Florian Grond*
McGill University
CIRMMT
Montréal, Canada

*François Côté*
1296 Avenue Maguire
Québec, Canada
G1T1Z3

*Jean Rouat*
Université de Sherbrooke
NECOTIS GEGI, CIRMMT
Sherbrooke, Canada
jean.rouat@usherbrooke.ca

## ABSTRACT

A low level sonification prototype of 3D point clouds for the sensory substitution of vision by audition for the visually impaired is investigated. The aim of this work is to study which point cloud features can be understood through the sonification of raw 3D data without the extraction of high-level features through algorithms. Preliminary results show the possibility for the user to localize objects and estimate their sizes but not to understand shapes of objects.

## 1. INTRODUCTION

Sensory substitution refers to the use of one sensory modality to access information usually available from another sensory modality. Sensory substitution devices are designed to help people with sensory loss to compensate with other remaining sensory modalities. This work focus on systems which substitute vision by audition for blind people. The first such device is the vOICe [1] which was originally developed by P. Meijer in 1992. The vOICe sonifies grayscale images by mapping the position and value of pixels to sound properties of short sine tones. Many other systems have subsequently been conceived for various purposes like text reading, locomotion, color recognition, obstacles detection, etc.

Systems like the ones by Zhong *et al.* [2] or some commercial applications [3, 4, 5] use sophisticated image recognition algorithms to give the user an easily understandable access to visual information. However, this strategy can produce false detections, often requires an internet connection for cloud computations and provides abstract information, which can lead to a lack of generalization if the system is used in a new environment. On the other hand, systems like the vOICe sonify raw data without the use of recognition or detection algorithms. This strategy has the advantage of providing direct information to the user. The work of Kim and Zatorre [6] shows that blindfolded sighted participants can learn to perceive tactile shapes with the vOICe. Studies have also shown that thanks to the great plasticity, the brain can use visual areas to process auditory stimuli [7]. The brain can interpret new sound stimuli to represent the visual space, and it does not seem necessary to use complex recognition algorithms.



We investigate which features of basic 3D point clouds can be perceived by sonifying the raw data. For this purpose, we propose a raw 3D point cloud data sonification prototype inspired by the work of coauthor F. Grond [8], where each point acts as a sound source. We use 3D data as inputs to the sonification prototype. Depth information is beneficial to plan actions and get a good understanding of the surroundings. Instead of using point clouds captured from depth sensors, in this work we generate artificial 3D points to ease the evaluation with reproducible stimuli.

The closest work we found related to ours [9] sonifies the interior contour of 3D points clouds captured by depth sensors to allow the user to localize and recognize the 3D real objects. Compared to this work we do not use any algorithm to extract contours or shapes. We sonify each point as describe above to study how far the brain can interpret them without complex processing.

## 2. INPUT DATA

We use four 3D artificial point clouds as illustrated in figure 1 to simulate virtual objects. Each object is initially composed of 5000 points. Points that are non-visible from the user's point of view are then deleted to mimic occlusion. The use of artificial points instead of real ones coming from 3D sensors gives us a complete control on the experimental conditions. Distances in the artificial space are arbitrary. To simulate pseudo-scenes (comprising one or two objects for now), point clouds are translated, rotated, rescaled and combined. Issues faced with real scenes are not addressed in this abstract.

## 3. SONIFICATION STRATEGY

### 3.1. Parameter mapping

As stated before we choose to sonify the raw data without preprocessing to study which features of a pseudo-scene, the subject can perceive. Each point acts as a short spatialized sound source. The emitted sound is a sinusoidal signal of 100 ms duration multiplied by a percussive envelope. The sinusoidal characteristics depend on the position of the point in the virtual space. The mapping between the sound characteristics and the point position is designed to be as natural as possible for the human perception to make the sound easier to interpret.

We linearly map the characteristics of the spatialized sound to the point position in the geometric space as follows :

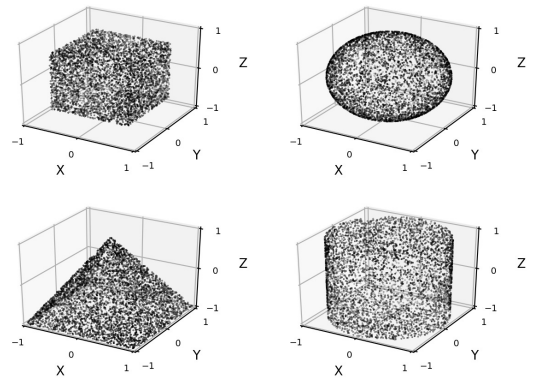


Figure 1: Illustration of the four point clouds used for the sonification: a cube, a sphere, a pyramid and a cylinder. Units in the figure are arbitrary. We choose for our preliminary test (section 4) that one arbitrary unit is approximately 10 cm.

$$\begin{cases} Freq\ [200Hz\ ;\ 2000Hz] \Leftrightarrow El[-70^\circ\ ;\ 60^\circ] \\ Amp\ [0dB\ ;\ -20dB] \Leftrightarrow Dist[0.5\ ;\ 5] \\ Delay\ [0.5s\ ;\ 5s] \Leftrightarrow Dist[0.5\ ;\ 5] \\ Sound\ Az\ [-62^\circ\ ;\ 62^\circ] \Leftrightarrow Point\ Az\ [-62^\circ\ ;\ 62^\circ] \end{cases} \quad (1)$$

where $Freq$, $Amp$, $El$, $Az$ and $Dist$ are respectively the sound's frequency, amplitude, the point's elevation, the sound's and point's azimuth, and the distance between the user and the point. The distance has no units as we work with artificial data. The limit of human horizontal binocular vision is used to define the mapping between the sound azimuthal position and the point azimuthal position. The limit of the human vertical field of view and the frequency sensitivity of the human hearing are used to define the mapping between the sound's frequency and the point's elevation position. We spatialize the sound only on the azimuthal plane with binaural first order ambisonic techniques [10], as the elevation clue is not accurate without an individual tuning of HRTF filters.

### 3.2. Insights about the sonification

The sonification process resembles granular synthesis where each point produces a grain which features depend on the position of the point. The result [1] is a whistling sound with the following features:

- Larger point clouds give longer sounds with wider frequency and amplitude coverage;
- Farther point clouds give weaker and delayed sounds;
- Elevation of point clouds is encoded into the pitch of sounds;
- Azimuthal position of the point clouds are encoded with sounds with the same azimuth;
- Denser point clouds generate louder sounds.

Density and distance of point clouds both impact the loudness of sounds. Although the distance is also encoded within the delay, this cross dependence can be confusing for the user. To address this issue, we currently work on a method to reduce the number of points to be sonified. We do not discuss in this abstract the problem of perceptual limit when two sounds produced by two close points cannot be differentiated by the human hearing system.

### 3.3. Technical details

We use python with the libraries PyQT [11], Numpy and Transformation [12] to generate and manipulate the point clouds shown in figure 1. We then use Supercollider to read the point cloud data (stored as csv files) and generate sounds. We use the ambisonic toolkit [13] available in Supercollider (first order ambisonic) to spatialize the sound on the azimuthal plane. As there are thousands of microsounds to generate, we use non real-time synthesis techniques in Supercollider.

## 4. PRELIMINARY TESTS

We have so far tested our sonification strategy with blind coauthor François Coté (F.C.). The goal was to see which features of pseudo-scenes the subject could understand with the 3D raw data sonification prototype. Examples of pseudo-scenes presented to the subject are illustrated in figure 2. F.C. had no prior knowledge of the sonification strategy we used.

We first tested the object localization potential of the sonification strategy. For this purpose, F.C. had to explore by hand few pseudo-scenes comprising real objects. At the same time, we were playing the corresponding sonified artificial point clouds. After this short training time (2 to 4 minutes), we played the sound corresponding to a pseudo-scene F.C. didn't know and asked him to point towards the object. F.C. was able to perfectly point towards the object almost always on the first trial. We then put two objects in the pseudo-scene and F.C. was also able to perfectly point to them at the first or second trial while listening to the sound.

We then tested the potential of the sonification to evaluate the size of objects. This time, F.C. had to sort different objects by size. During training, we place in the pseudo-scene two real objects having different sizes. F.C. touched them while we were playing the corresponding sounds. F.C. then had to sort four objects of different sizes based on the sonified point clouds corresponding to the objects. F.C. was able to perform this task perfectly after listening one by one to the four objects.

Finally, we tested the shape comprehension potential of the sonification. F.C. had to figure out whether the shape was a cylinder, a sphere, a cube or a pyramid. This time F.C. was unable to succeed with this task. F.C. told us that he did not find a natural link between the sound and the shape he was touching.

Although we still have to conduct quantitative experiments with more subjects to confirm these results, this test gives us an indication of what we can accomplish with this 3D raw data sonification strategy.

[1] sound examples available under:
https://www.gel.usherbrooke.ca/rouat/soundsICAD2018.zip

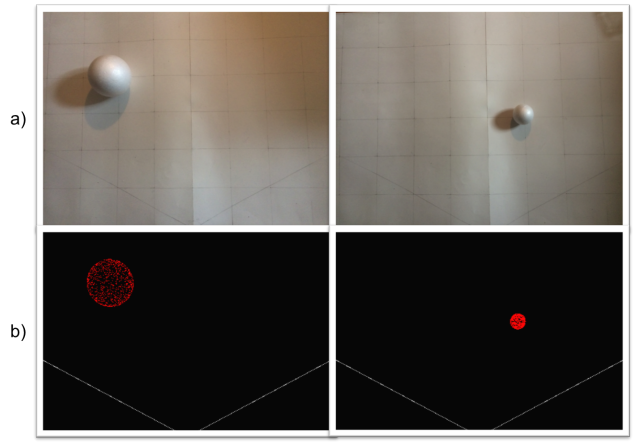


Figure 2: a) example of two different pseudo-scenes seen from above b) corresponding artificial point clouds (visualized with python and PyQT) that were used to generate the sounds.

## 5. CONCLUSION

This extended abstract presents an ongoing work to design a 3D point cloud sonification prototype which is still in development. The initial tests lead us to believe that we are going in the right direction to encode the position and the size of objects with the sound. For the shape recognition task, a pre-processing step using more elaborated algorithms seems to be needed. Feedback from the first study has provided avenues we are currently exploring to better encode shapes through sounds. Further work also includes the development of an interactive near real-time system. This poses the challenge of point cloud data reduction without losing too much information in order to have a smaller number of points to sonify.

## 6. ACKNOWLEDGMENT

We thank the financial support from Québec FRQNT. We also thank the CIRMMT for the infrastructure support, Franco Lepore for the stimulating discussions and Etienne Richan and Luca Celotti for a partial proofreading. We thank the 2 anonymous reviewers for their constructive comments.

L.C. and J.R. conceived the prototype and conducted the experiments; L.C. made the programming, built the prototype and wrote the manuscript; J.R. and F.G. proofread the abstract; F.G. contributed to the sonification based on his PhD thesis and his artistic installation: *the Haptophone*; F.C was the testing subject.

## 7. REFERENCES

[1] P. B. L. Meijer, "An experimental system for auditory image representations," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 2, pp. 112–121, Feb 1992.

[2] Y. Zhong, P. J. Garrigues, and J. P. Bigham, "Real time object scanning using a mobile phone and cloud-based visual search engine," in *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 2013, p. 20.

[3] http://taptapseeapp.com.

[4] https://camfindapp.com.

[5] https://www.aipoly.com.

[6] J.-K. Kim and R. J. Zatorre, "Can you hear shapes you touch?" *Experimental Brain Research*, vol. 202, no. 4, pp. 747–754, May 2010. [Online]. Available: https://doi.org/10.1007/s00221-010-2178-6

[7] H. Duffau and C. Duchatelet, *L'erreur de Broca: exploration d'un cerveau éveillé*. Michel Lafon, 2016.

[8] F. Grond, "Listening-Mode-Centered Sonification Design for Data Exploration," Ph.D. dissertation, 2013, Bielefeld University.

[9] H. Pourghaemi, T. Gholamalizadeh, A. Mhaish, G. İnce, and D. J. Duff, "Real-time shape-based sensory substitution for object localization and recognition," in *Proceedings of the Eleventh International Conference on Advances in Computer-Human Interactions*, 2018, pp. 45–50.

[10] M. Noisternig, T. Musil, A. Sontacchi, and R. Holdrich, "3D binaural sound reproduction using a virtual ambisonic approach," in *Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2003. VECIMS '03. 2003 IEEE International Symposium on*, July 2003, pp. 174–178.

[11] http://www.pyqtgraph.org.

[12] https://github.com/davheld/tf/blob/master/src/tf/transformations.py.

[13] J. Anderson, "Introducing... the ambisonic toolkit," in *Ambiosnics Symposium 2009*, 2009.