Read Your Voice

A Playful Interactive Sound Encoder/Decoder

Hugo Pauget Ballesteros
IRIT, UMR CNRS 5505
Toulouse, France
hugopauget@gmail.com

Gilles Azzaro
Independent Artist
Marssac-sur-Tarn, France
azzarogilles@gmail.com

Jean Mélou
IRIT, UMR CNRS 5505
Toulouse, France
Jean.Melou@irit.fr

Yvain Quéau
GREYC, UMR CNRS 6072
Caen, France
Yvain.Queau@ensicaen.fr

Jean-Denis Durou
IRIT, UMR CNRS 5505
Toulouse, France
durou@irit.fr

ABSTRACT

Read Your Voice is a playful interactive multimedia system that allows the user to record a sound, encode it as an image, and then play it back using his smartphone, while controlling the speed and direction of playback.

KEYWORDS

Voice, interactive, sound encoder/decoder.

ACM Reference Format:
Hugo Pauget Ballesteros, Gilles Azzaro, Jean Mélou, Yvain Quéau, and Jean-Denis Durou. 2022. Read Your Voice: A Playful Interactive Sound Encoder/Decoder. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22), October 10–14, 2022, Lisboa, Portugal.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3503161.3549974

INTRODUCTION

Recording sounds: this dream was realized for the first time around 1860 by a Frenchman named Edouard-Léon Scott de Martinville. His device, named *phonautograph*, made possible the first "material transcription" of the voice, seventeen years before Thomas Edison. Nevertheless, if this early invention is not well known, one should not be mistaken: it is indeed Thomas Edison's *phonograph* which first allowed to record a sound *and to be able to play it back*. The recording of a sound and its restitution present very different levels of difficulty. It was only in 2007 that four American researchers from Berkeley succeeded in replaying several *phonautograms* [3], 150 years after their recording!

French artist Gilles Azzaro follows a similar approach to Scott de Martinville. But while sound is a 1D-signal, which can be represented as a 2D-curve, its time-frequency representation in the form of a *sonagram* provides a 3D-model: one dimension for time, another one for frequency, and the last one for power. This well-known 3D-representation of a sound signal is thus particularly suited to 3D-printing. Thanks to some modifications, Gilles Azzaro knows how to transform it into a real artwork (see Figure 1).

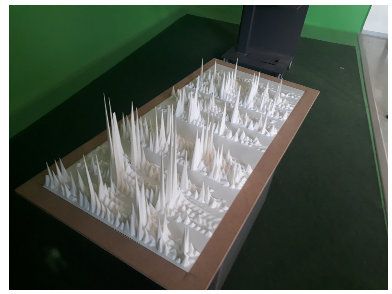

Figure 1: 3D-printing of a sound by Gilles Azzaro [1].

3D-SONAGRAMS

Under perfectly controlled acquisition conditions, the sound restitution of such a 3D-representation can be performed. In a previous edition of the ACM Multimedia Art Exhibition [2], we showed how to read a 3D-sculpture using a depth sensor (see Figure 2). Now, is this still feasible if the reading is done by a visitor using his smartphone? In such much less controlled acquisition conditions, we revert to a 2D-representation of the sonagram, in order to avoid problems of hidden parts and to limit parallax effects.

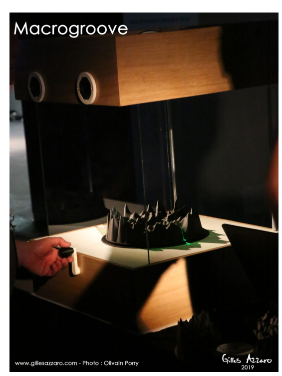

Figure 2: Reading of a 3D-sonagram with a depth sensor [2].

MM '22, October 10–14, 2022, Lisboa, Portugal

ACM ISBN 978-1-4503-9203-7/22/10.
https://doi.org/10.1145/3503161.3549974

2D-SONAGRAMS

Figure 3 shows another example of 3D-sonagram, but in the form of a *lithophane*: the 3D-printing faces the light placed at the back and is seen by transparency. It turns out that the image produced is the same as if it had been printed on a plane. Parallax effects are then reduced to a transformation called *homography*, in the case where the shooting is not fronto-parallel, which is easy to correct using "markers": in the example of Figure 3, this explains the presence of two series of parallel black lines, above and below the sonagram. Such lithophanes are particularly aesthetic, but do not lend themselves to a playful interactive application such as *Read Your Voice*, since 3D-printing is time-consuming and costly. This is why we decided to rather use *real* 2D-sonagrams.

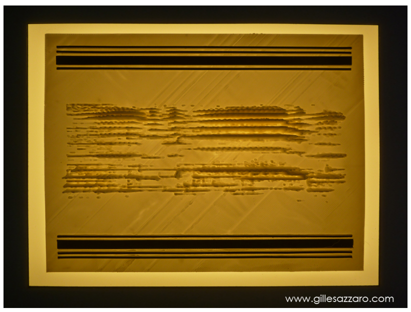

Figure 3: *Miracle* (2016), 3D-lithophane by Gilles Azzaro [1].

The basic operator to produce a sonagram is the *Short Term Fourier Transform* (STFT). However, a number of modifications are made to the STFT of the signal, in order to make the image acquisition more robust and to concentrate the relevant information:

- Only the positive frequencies are kept.
- The ordinate corresponds to the logarithm of the frequency, in order to "spread out" the low frequencies.
- The graylevels correspond to the square of the complex modulus, expressed in *dB*, and then *quantized* on eight values.

An example of such a sonagram is shown in Figure 4.

PHASE RETRIEVAL

As already said, reading a sonagram is much trickier than creating it. Amongst the sonagram modifications listed above, several are non-invertible. Nevertheless, the main obstacle to reading a sonagram such as the one in Figure 4 comes from the fact that the phase of the STFT is lost. This results in distorsions in the reconstructed sound. A video by Gilles Azzaro (https://www.youtube.com/watch?v=kXANN0Nfpeo), where the 3D-lithophane of Figure 3 is read using the PhonoPaper application [5], highlights these distorsions.

To overcome this obstacle and make the reconstructed sound audible, we use the *phase reconstruction* method proposed by Griffin and Lim in 1984 [4]. This method was initially designed to reconstruct the phase in the case where the sonagram of the original signal has been modified, for instance, to improve sound quality. In such a case, the original phase is no more coherent with the modified sonagram. Our case is different, since the phase is lost and must be retrieved before playing the restored sound back.

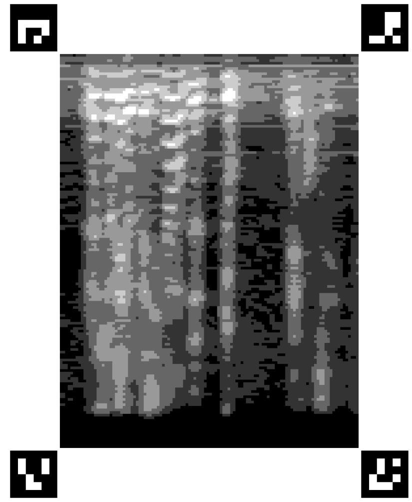

Figure 4: Example of a sonagram surrounded by markers: time on the abscissa, frequency on the ordinate (oriented downwards). The power in *dB* is quantized on eight values.

READ YOUR VOICE

To better evaluate the relevance of the proposed multimedia system, called *Read Your Voice*, an example accompanies this paper (https://bit.ly/3PIzYPc). Any sound signal could be treated in the same way. In order to make the workshop fun, each visitor will be asked to briefly introduce himself. Therefore, a collection of sonagrams can be printed or projected on a white wall, and then read by other visitors belonging a smartphone.

An important peculiarity of the proposed multimedia system is to make it interactive and playful. In the version which will be presented at the exhibit, the user will have the control on the playback, giving the pace of lecture and also deciding on its direction (forward or backward), by simply moving his smartphone.

PERSPECTIVES

A first perspective of this work consists in trying to reconstitute the phase "on the fly", which would make it possible to read a sonagram printed on a sheet of paper, then to crumple the sheet and to note the effect produced on the restored audio signal, as this is done in a video presentation of PhonoPaper (https://www.youtube.com/watch?v=lzoVnqLy29U&t=0s).

Finally, another perspective is to print some famous sentences in 3D, and to constitute an exhibition of lithophanes such as that of Figure 3, beside the recreational workshop.

REFERENCES

[1] G. Azzaro. 2022. Homepage. http://www.gillesazzaro.com.

[2] P. Chable, G. Azzaro, J. Mélou, Y. Quéau, A. Carlier, and J.-D. Durou. 2019. Macrogroove: A Sound 3D-sculpture Interactive Player. In *Proceedings of the 27th ACM International Conference on Multimedia.*

[3] P. Feaster. 2017. Playback Methods for Phonogram Images on Paper. In *Proceedings of the 2016 Joint Technical Symposium.*

[4] D. Griffin and J. Lim. 1984. Signal estimation from modified short-time Fourier transform. *IEEE Trans. on Acoustics, Speech, and Signal Processing* (1984), 236–243.

[5] A. Zolotov. Latest release - 1.6 (2018-03-10). PhonoPaper. http://www.warmplace.ru/soft/phonopaper/.