

Pict2Audio : Sound Generation by Hand-Drawn Images Analysis using Convolutional Neural Networks

Scan and get more information !



Joséphine Calandra, Pierre Hanna, Pierrick Legrand, Myriam Desainte-Catherine

SCRIME, Bordeaux University, France

The SCRIME

This project has been developed during Josephine Calandra's end-of-study internship as a student at ENSEIRB-MATMECA, which occurred at the SCRIME, the Studio of creation and research in computer science and experimental musics.

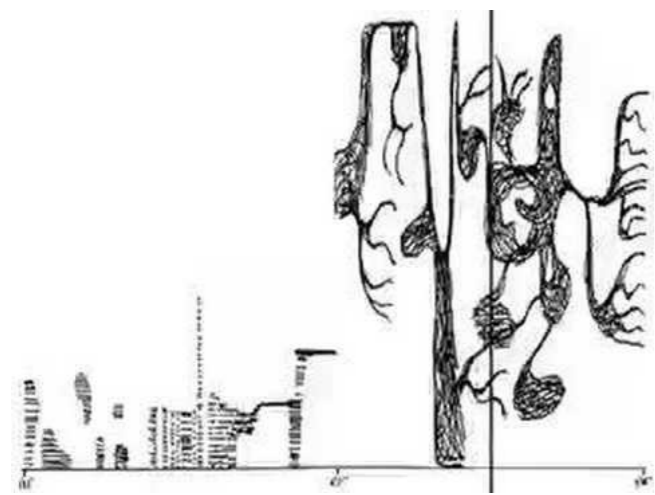


Figure: Extract from the music sheet Mycenae Alpha, Iannis Xenakis.

From electroacoustic music to new means of expression...

Electroacoustic music composition invites people to think about the instrumental composition in other ways. The use of transformed, synthetic or artificial sounds created by a computer as a tool leads to new reflections about the process of music creation, creation of new sounds and interaction between the tool and the composer. Since the XXth century, new tools, languages and edition software has been developed by research centers and companies.

...To the necessity to create a dedicated support

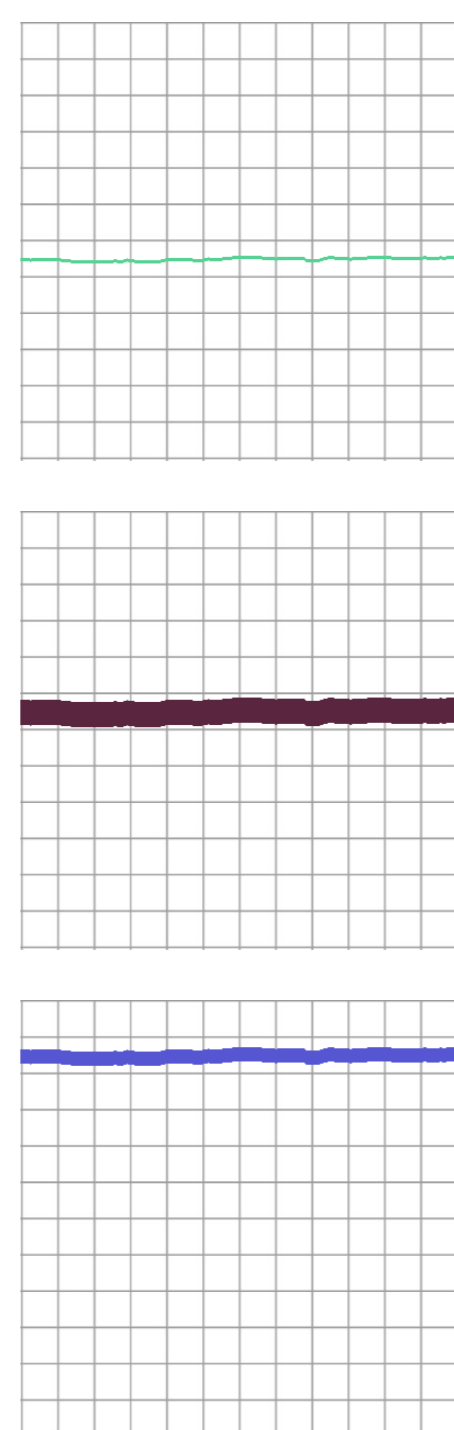
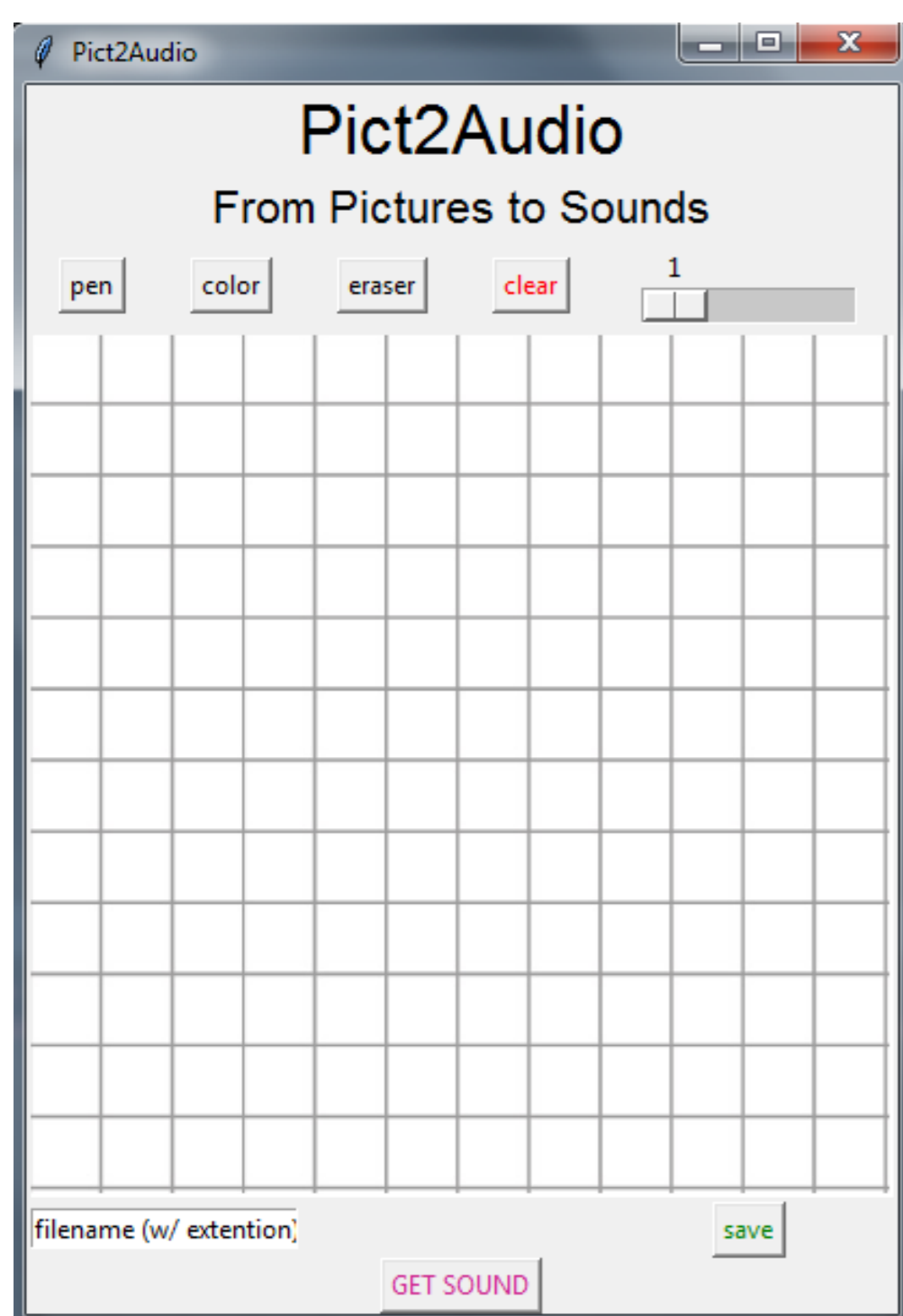
Composers from the XXth and XXIst century develop their own tools and languages to compose. Nevertheless these tools may not be universal but application-specific, and they could be barely intuitive nor customizable. Sounds in edition software are fixed, the software being more a way to edit than to express the sound. A problem has emerged : how to create a universal tool that could facilitate the musical creation ? We would like to create an intuitive tool that leads to a natural expression of the composer who could create her/his own language but also manipulate it easily.

Pict2Audio : A tool for composition

- ▶ Matches drawing and sounds.
- ▶ Creates a customizable graphic language.
- ▶ Uses sound databases belonging to the composer.
- ▶ Provides a drawing interface *via* graphic tablet augmented with a stylus.

Technologies used

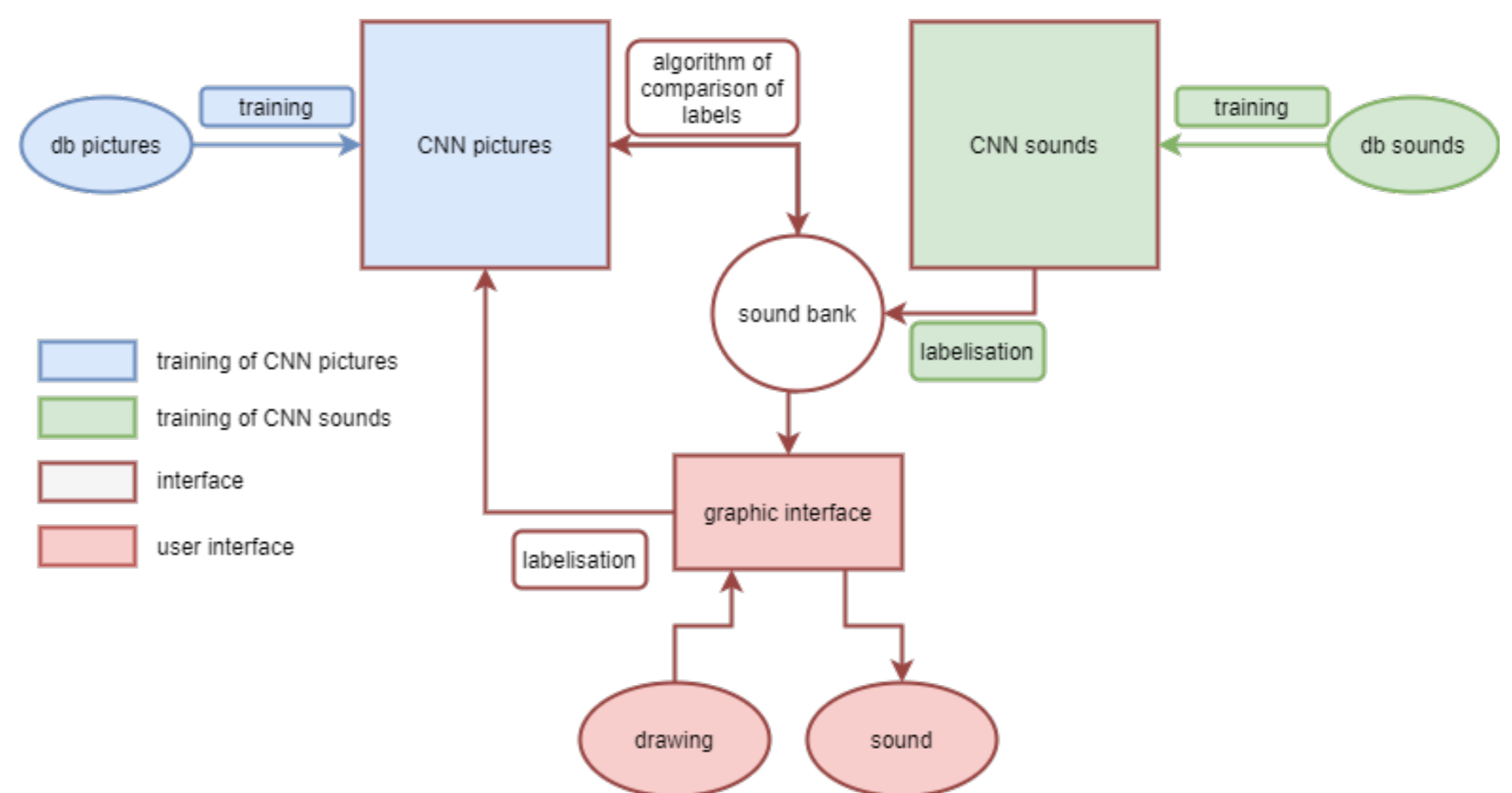
- ▶ Neural Networks developed in Python with the libraries Keras and Tensorflow.
- ▶ Use of Google Colab environment, that runs neural networks on Google servers.
- ▶ Graphic Interface developed in python using the library tkinter.



The system of Pic2Audio

The system is divided into two parts :

- 1) **The training of the system** : The drawer gives sounds to the system. These sounds are used to train a group of neural networks according to characteristics associated with the sound. Then, the neural networks associated with pictures are trained in the following way : the composer draws on the interface pictures such as visual characteristics correspond to audible characteristics.
- 2) **The use of the system** : once the neural networks are trained, the composer can draw anything that corresponds to the audible characteristic desired. The neural networks will analyse the picture, and the system will return the associated sound thanks to a dedicated matching algorithm.



Why do we use neural networks ?

Neural Networks are specifically efficient for classification of pictures and sounds. Moreover, it does not need to know the algorithms nor the architecture of the systems that generate the signals at the entry of the neural networks. This enables a possible abstraction of music theory and sound analysis. Then, the neural network detects the characteristics of the pictures, which leads to a personalisation of the system.

Specific framework and improvement

In this context we limited our researches to the specific training of the neural network with a predetermined language where three visual characteristics are associated with three audible characteristics. The drawings are lines where the colors correspond to a tone, the height of the line corresponds to the height of the note and the thickness of the line corresponds to the volume of the sound. The sound databases are augmented databases of NSynth, the database proposed by Magenta, the music generation research project of the Google AI team.

Limits

To go towards a usability and universality of the system, it still has to be tested with various and complex drawings. Moreover, the creation of the databases can be laborious for the composer, so the databases should be automatically augmented. Also, there could be a bad training of the neural networks, so there is a need to create an assistance for the composer.

Perspectives

The future developments of this system could be the implementation of the polyphony, the management of time in the system, and the drawing and emission of sound in real-time.

And why not...

We could go further by imagining the system in 3 dimensions, coupled with an augmented-reality system. Moreover, even if Pict2Audio is aimed at helping composers, this could also be a tool used for pedagogy, to help learning music.