Improvisers in Life

An audiovisual project
Lilac Atassi
MFA - Composition
California Institute of the Arts
October 2018

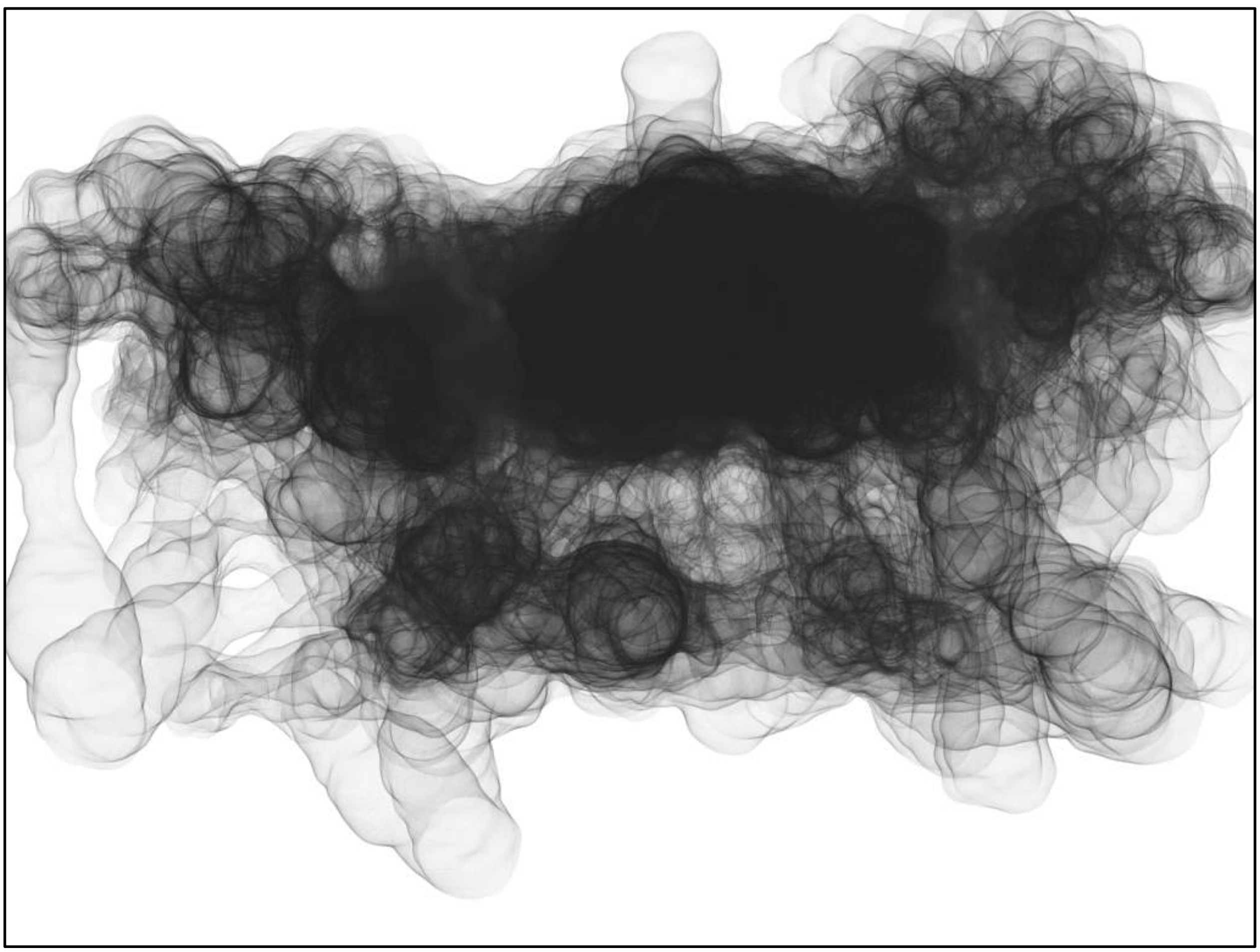

Figure 1: the developed visualization for this project, the tracked bodies paint on the canvas. The final image after a 10-minute piece performed by four performers.

Table of contents

Abstract

This report documents the artistic and technical aspects of my original audiovisual project *Improvisers in Life* (September 2018). Translating expressive qualities of body movement into sound expressions is the main motivation behind this project. Using a camera and a laptop, 17 body points are tracked. The position and velocity data at these points is then sent to two separate scripts to generate sound and visuals. In this report, the details of the implementation is discussed.

1. Introduction

Improvisation involves processing and reacting to inputs that we receive from the environment by continuously producing outputs in real-time. The inputs could be visual, auditory, kinesthetic, gustatory, olfactory, psychological and emotional. As long we are breathing we are improvising with our environment consciously or subconsciously. If we regard life, in general, as a story, each living creature has a sub-story to live and improvise during their lifetime.

In *Improvisers in Lifer,* the camera's field of view represents life. Once a person enters in the camera's field of view (FOV), their body movements produce sound. Figure 2, depicts the FOV. The performer and the sonification method form a single system. The two interact with each other, and the performer can decide to what degree to lead the system. This system can be expanded to produce unique musical sounds for each of a group of performers. Therefore, each performer makes their own story, or own piece of music, by improvising in front of the camera.

The feedback loop in this system is a small scale representation of the universe. In this system, sound reacts to the performer's movements. This reaction could be determined stochastically. Concurrently, the performer perceives the stochastically generated sound and reacts to it. This bidirectional influence is apparent between nature and humans.

There are several previous projects on translating bodily movement into sound generation. Winkler in [1] provides a concise review of several classic works. One novel aspect in my project, *Improvisors in Life*, is that the performer produces art through multiple media. The performer strives to:
a) Choreograph the movements.
b) Compose the audible performance.
c) Design the pictorial result painted on the canvas.

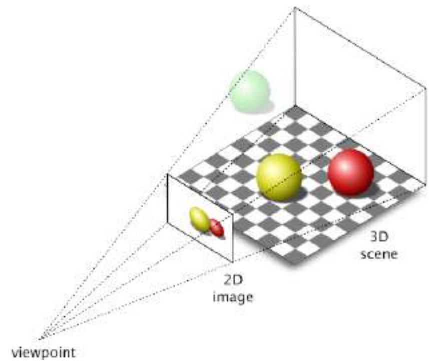

Figure 2: the field of view (FOV) of camera.

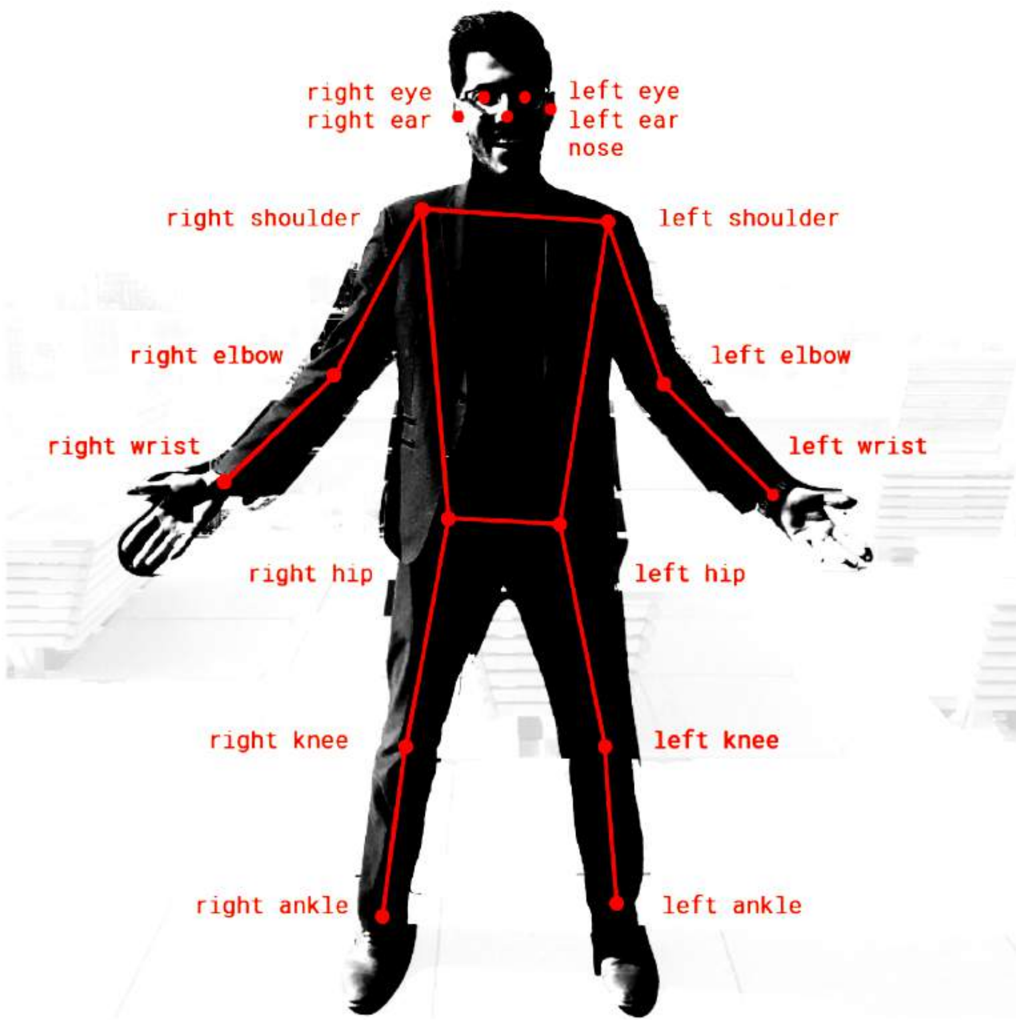

Figure 3: the 17 body points that the deep-learning-based body-pose estimator can detect and track. For each of these points a set of properties is calculated.

2. Implementation

With current state-of-the-art image processing computer algorithms, it is possible to track 17 body points using a laptop and a webcam, as illustrated in Figure 3. This low cost solution can replace the other body-pose tracking solutions that are prohibitively costly for most students, researchers and hobbyists.

The main deliverable of this project is an open and free framework that tracks one or multiple performers in front of a webcam, and via Open Sound Control (OSC), sends the tracked body points to specialized applications to produce sounds and images. Figure 4 depicts the high-level structure of the framework. Multiple example scripts and patches for Max, Processing/P5.js, and Supercollider are being developed to demonstrate the system.

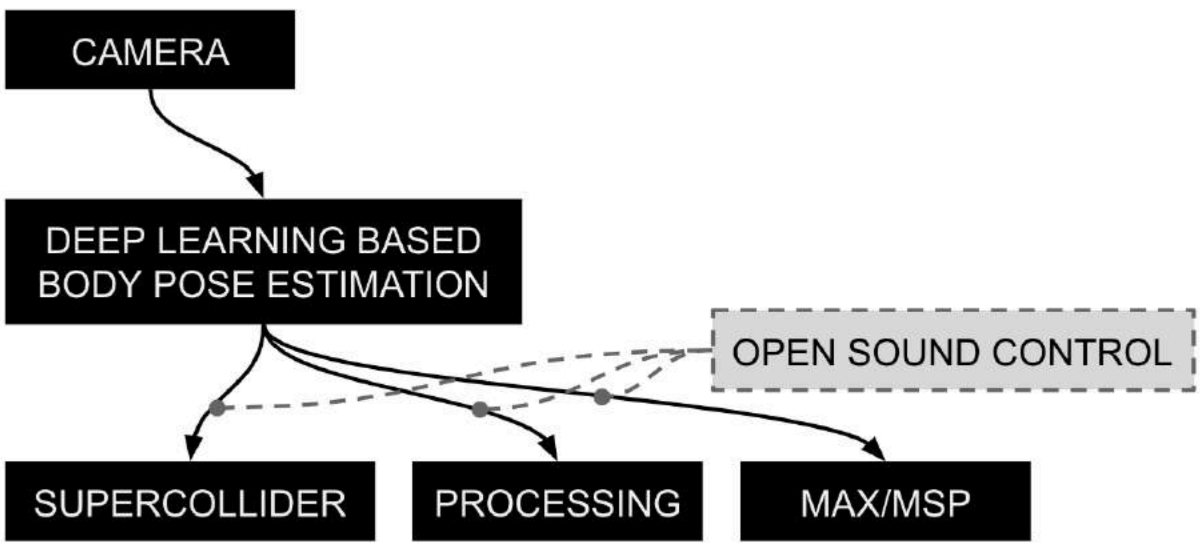

Figure 4: the framework’s structure: the deep-learning-based body-pose estimation is implemented in Python and uses Google’s TensorFlow library.

The computer vision algorithm is based on PoseNet [2], an artificial neural network that detects the 17 body points and returns their coordinates in the image. From the point coordinates, the velocity and relative velocity to the other points are calculated. This set of positions and velocities are generic enough for most bodily movement to sound mappings.

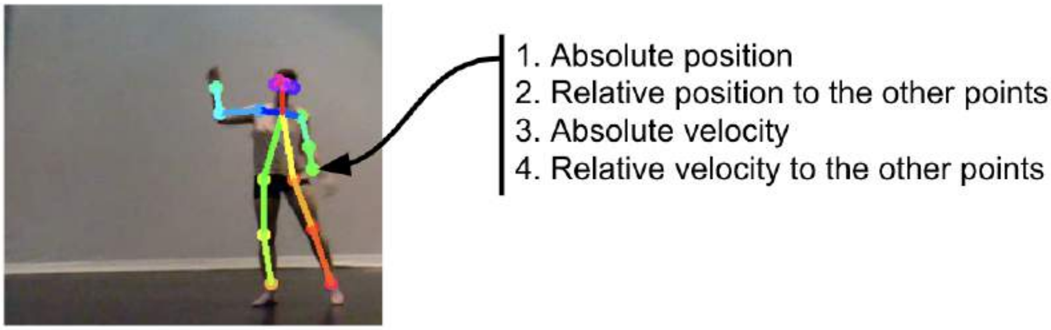

Figure 5: the set of features for each detected body point.

2.1. Psychoacoustics

An arbitrary subset of the movement-related quantities controls the sound parameters. For instance, standing at the left side of the FOV lowers the amplitude, and standing at the right side has the opposite effect. During a practice session while testing this system, an intriguing phenomenon became apparent. The speed of a body point, for instance the right wrist, was mapped to the sound amplitude. Even though the mapping was done with a linear function, the loudness was not appropriately proportional to the amount of movement effort.

This observation recalls the non-linearity of human sensation as observed in psychoacoustics. Fechner's law generalizes this relationship and states that human perception is proportional to the logarithm of stimuli measured with physical instruments. Non-linear mappings, including logarithmic mapping in Fechner's law, have been traced to the neural structures in the brain [3], providing scientific support for this psychophysical law. This also suggests that the brain will find non-logarithmic mappings between perception and physical quantities unnatural. This is important in order for the performers to feel the generated sound follows their movements.

My early experiments suggest that exponential and power functions result in a more intuitive mapping between the speed of movement and sound loudness. In these experiments, the felt sound loudness was more proportionate to the speed of movement over a wide range. This requires a more detailed study and experimentation. Another closely related concept that could benefit the perceived quality of mapping movement to sound is just-noticeable difference (JND). That is, the JND of the input, for instance the speed of arm movement, should correspond to the JND of the output, such as the sound loudness. If this is not the case, the mapping will feel inconsistent.

3. Demonstration

In this video[1], a set of four distinct sonification algorithms have been developed in Supercollider that permit dancers to artfully explore the sound space. To generate visuals, I used P5.js. On the screen, a translucent circle is drawn at every frame. The location of the circle on the screen corresponds to the location of the performer in the camera's field of view. at every frame, Perlin noise is added to the radius in the polar coordinate system before converting to Cartesian coordinates. As the performers move in front of the camera, more noise circles are drawn, occluding the screen.

Figure 6: demonstration video on YouTube.

4. Code

The computer code for tracking body points using a camera is available on my github profile at `https://github.com/lilacatassi`. The main script is `run_webcam.py`.

```
body_parts = [("/right_wrist", 4), ("/left_wrist", 7),
              ("/right_elbow", 3), ("/left_elbow", 6),
              ("/right_shoulder", 2), ("/left_shoulder", 5),
              ("/right_knee", 9), ("/left_knee", 12),
              ("/right_ankle", 10), ("/left_ankle", 13),
              ("/right_hip", 8), ("/left_hip", 11)]
```

[1] https://youtu.be/5A5OnXPJetg

The OSC messages are mapped to the body point numbers that the computer vision algorithm returns via the first field of the message. The horizontal and vertical position and the speed are appended to the message in the code.

The IP address and port number of the OSC server can be passed to run_webcam.py at launch time. The script acts as a client broadcasting OSC messages. The receiving application needs to be in the server mode. Supercollider, by default, has an integrated OSC server when running. In Processing and P5.js, the OSC server needs to be coded.

On the OSC client side, we can filter the messages—for instance the messages for the left wrist—and use the data in the message to control the sound parameters. The following snippet shows the Supercollider code for that purpose.

```
OSCFunc({ arg msg, time, addr, recvPort;
    // msg is an array sent from the body tracking application
    },
 '/left_wrist');
```

5. Directions for Further Exploration

There are multiple directions in which to expand the project. Having more than one performer in the camera's field of view can create interaction between two or more people and the sonification system.

In its current state, the sonification system provides to the performers a small set of mapping from certain movement qualities to audible features. Expanding this set gives the performer more freedom in expression. Particularly I would like to explore more non-instantaneous options, for instance movement repetition controlling some sound aspect.

A psychophysical framework for defining the mappings from movement to sound will give a better sense of control to the performers. This would likely also improve the aesthetic effect for the audience.

6. References

[1] T. Winkler. Making Motion Musical: *Gestural Mapping Strategies for Interactive Computer Music.* In *Proc. of the 1995 International Computer Music Conference*. San Francisco, Calif.: International Computer Music Association, pages 261–264, 1995.

[2] Alex Kendall, Matthew Grimes, and Roberto Cipolla. 2015. PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*.

[3] Dehaene, S. The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, 2003, 7(4), 145-147.