Affine Tuning

Embodied Interaction and Dynamic Composition

Marc-André Weibezahn

AFFINE TUNING:
EMBODIED INTERACTION
AND DYNAMIC COMPOSITION

Marc-André Weibezahn

Submitted in partial fulfillment of the requirements for the M.A in Digital Media at the University of the Arts Bremen

Supervision
Prof. Dennis Paul
Prof. Kilian Schwoon

An online version of this documentation is available at https://affinetuning.com/thesis

PREFACE

This work is my third approach to develop interactive music that reacts to the body movements of the player. Since the second semester of my studies in the MA program in Digital Media, this topic has occupied me.

In my opinion, the medium “sound” in interaction with digital media—except for consumption—is much less appreciated than visual forms, except maybe for the field of video games. This is expressed by the fact that we look at screens and interact via visual interfaces every day, but only slowly are sound-based interfaces such as voice-controlled assistants entering the mass market. One anecdote to illustrate how these are still in their infancy: The voice assistant “Alexa” in Amazon’s “Echo” devices can be set to a “whisper” mode. The assistant confirms this activation - but at normal volume. Only after that it whispers.

A second observation is that bodily interaction seems to get more and more popular in digital interactions, be it direct manipulation like in computer games (often in virtual reality) or in a more mediated way, as performers for others on social video platforms (often using augmented reality applications, like filters etc.).

These two findings are not the main motivation for my work in this field. But they are undoubtedly a further indication that there is something valuable and worthwhile to be found here that I set out to explore.

I am by no means the first person who finds this topic fascinating and can’t quite get away from it. Two artists have told me that they have been occupied with this topic for years and that it has sometimes even been a burden to them. At times, I had similar feelings in the process of this work, and felt my enthusiasm for it almost like a curse. Inspired by this, I named one interactive composition *Nefrin*, which means “curse” in Farsi.

Marc-André Weibezahn, Leipzig 2021

Table of contents

List of images

List of tables

1. CONCEPT

This work is an experimental research project on the possible popularization of a form of interactive music that reacts to body movements. The core project is the development of a software that allows players to experience such music in their own homes (or any other place of their choice). The related tasks were research, concept, interaction design, composition, sound design, user interface design, programming and testing. The platform of the application is Apple iPhone. The software should be released via the Apple App Store in early 2021.

I have already worked on two projects with body-sensitive music. So far they have always been presented as installations, i.e. in an exhibition context. This had the advantage that I had the greatest possible control over the surrounding conditions, such as the design of the room, the installation, sound system, etc. When I was on site during the exhibition, it was very interesting to observe the reactions of the visitors and talk to them about it.

However, there were also some disadvantages in my opinion. Mostly, I had doubts about the exclusivity of the exhibition situation, where the installation is only accessible for a few days at a certain location. I also suspect that visitors behave differently in an exhibition than in a private situation. (I tried to cushion this effect by restricting the visit to one visitor at a time, unless otherwise desired).

For these reasons, I decided to create a new work that not only advances my research in this area, but shifts the experience from semi-public space to private. Ideally, this means a bigger audience with more freedom of experience.

In terms of content, this work should explore further possibilities of responding musically to body movements and forms. The decisive input parameter of my previous work was the intensity of the players' movements. The musical response to this was various rhythms, partly generated by algorithms. In this work, I want to focus primarily on the possibilities of tonal feedback.

This project shares one common ground with my earlier work: the feedback is exclusively auditory; there is no further visual feedback. The goal is to focus on the tonal experience in connection with body movement. (There is, however, the possibility to record a video of a performance). It is not meant to be a "new instrument" which can be perfectly controlled but should rather open a room of possibilities for interactions and highly subjective interpretation.

The Name

The title "Affine Tuning" is inspired by the concept of affinity, which has slightly different meanings in different fields of expertise, but generally refers to an inclination, similarity or desire for convergence. "Tuning" of course refers to the tuning of an instrument. But in German—my mother tongue—the direct translation "Stimmung" also refers to the current state of mind, or can describe an emotional quality of an artistic expression. In addition, one can “tune in” to something in order to behave in a similar or corresponding way. In the concrete case of interactive music, my ideal is that the composer, system and player tune in to open up a space for a sonic and physical experience that is ephemeral but leaves a lasting impression. And it is not by chance that the name is homophonic to “A Fine Tuning”, because this in turn describes quite well the balance between interactivity and artistic expression which I have to find in designing the system.

2. RESEARCH

2.1. Related works

The field of interactive digital systems that react to the body is not new, ranging from early works in the 1970s like Myron Krueger's *Videoplace,* [1] to 2020's *Super You* [2] by collective Universal Everything. Previous works which focus on music can be located in the following fields of application:

- **Installations**
- **Performances**
- **Games and Entertainment**
- **Therapy**

In the following I will introduce four quite different projects and products. Their specific interactions and effects can only be conveyed by text to a limited extent, so I include clickable links to online videos for each project.

2.1.1. Installation: *Very Nervous System* Video [3]

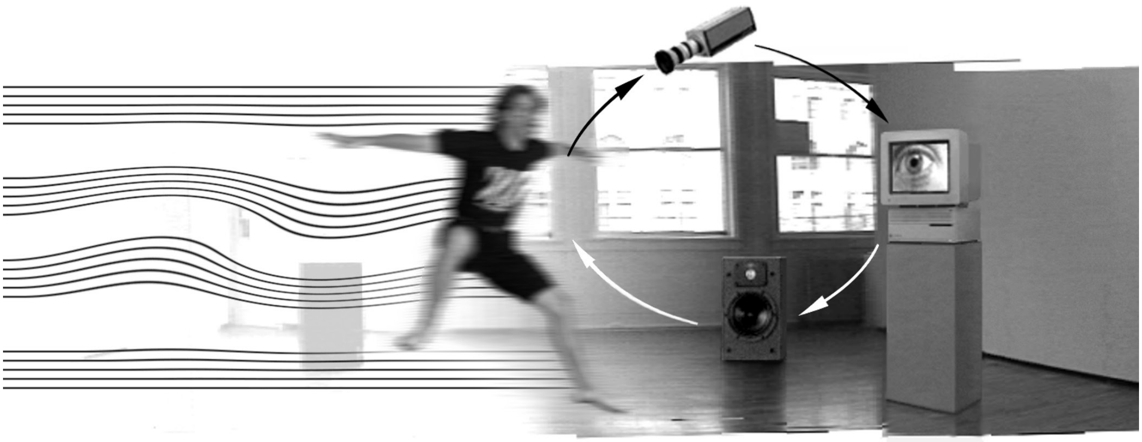


Figure 1. This montage illustrates the principle feedback loop of VNS: Action -> Observation -> Software interpretation -> Musical feedback -> Observation…

Image: Rokeby

The installation *Very Nervous System* by David Rokeby was presented at the 1986 Venice Biennale. It is a system of cameras that tracks the users, uses software to interpret their movements, and plays a synthetic orchestra as a response [4]. There is no visual feedback. Rokeby himself said that his main interest is the interaction between user and system, rather than the individual input and output phenomena:

> *“While the ‘sound’ of the system and the ‘dance’ of the person within the space are of interest to me, the central aspect of the work is neither the ‘sound’ nor the ‘dance’. It is the relationship that develops between the sounding installation and the dancing person that is the core of the work. […] The installation watches and sings; the person listens and dances. But the relationship that develops is not simply that of a dialogue between person and system. Dialogue in its back-and forthing implies a separation of the functions of perceiving and responding. But for the installation, perception and expression are virtually simultaneous. As a result the installation and participant form a feedback loop which is very tight, yet very complex.”* [5]

Rokeby relies on the visitor to give up direct control of the computer in favor of a synergy of human movement, and digital perception and reaction. There is interpretation on both sides: the machine has to process the low resolution images of the (initially self-crafted) cameras through preconceived rule-systems. The visitor, on their part, tries to relate the immediate musical feedback to their actions and can be inspired for further actions.

Rokeby began work on *Very Nervous System* in 1982, at a time when computers were still perceived primarily as “calculating machines” and were largely reserved for industry, science and technology enthusiasts. He continued to develop the system over 13 years, modernizing components such as video input and sound processing. The latter was later based on the Max/MSP graphical development environment, which is popular especially in the audio field. Until 2002 Rokeby also sold the system *softVNS*, a collection of Max objects that could be used for tracking and processing motion [6].

I had a brief contact with Rokeby by e-mail after I discovered his work during my research. I asked him directly if he considers my project obsolete. In his opinion this was not the case, for the following reasons: Everyone that he knew who had worked on this topic so far had found their own approach. In addition, the approach of bringing such a system into the private sphere should be quite interesting and worthwhile.

2.1.2. Video Game: *Just Dance* Video [7]

Figure 2. Screen from "Just Dance 2021" for Sony Playstation.

Image: Ubisoft

Just Dance is a representative of dance video games, which itself are a niche in the genre of music video games [8]. *Just Dance* was released on Sony Playstation, Microsoft XBox, Nintendo Wii, iPhone and for Android smartphones. The gameplay is based on players dancing choreographies to well-known pop hits. The chains of postures are strictly defined, and timing is of the essence. Depending on the version and platform, the movement and pose of the players is detected by various technologies, from camera input and peripherals such as hand-held motion-sensitive gaming controllers, to the smartphone, which registers the movements via acceleration sensors. The choreographies are displayed on-screen during the game, with successive steps scrolling through the image from right to left. These visual timing aids appear in many music games like *Singstar*, *Guitar Hero* or *Rock Band*, and before that had been used in karaoke videos, where a ball bounces from one syllable of the lyrics to the next. In addition to these timing hints, the rest of the interface of *Just Dance* is visually rich: for each track there is an exclusive music video with dynamic visual effects. In fact, the feedback to the players performance is purely visual as there is no change in the audio layer at all. The goal of the normal mode is to re-dance the choreographies as accurately as possible to achieve the highest possible score. The series lives strongly from the well-known songs and star performers and has been released in annual iterations with new songs since 2009.

2.1.3. Interactive Entertainment: *Nagual Dance* Video [9]

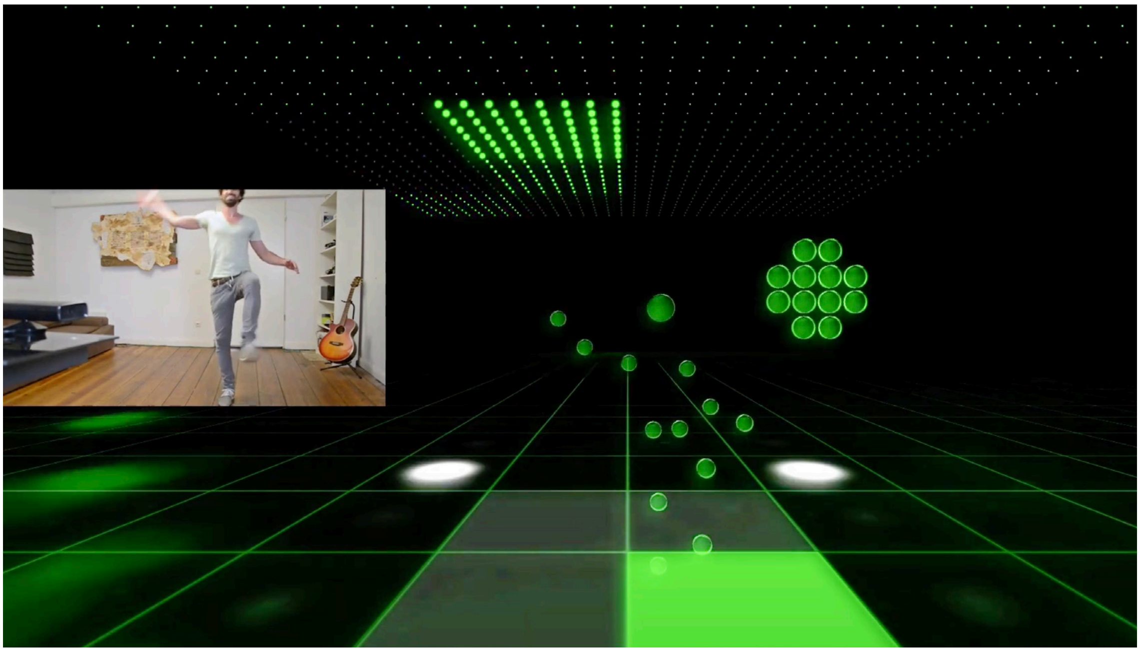


Figure 3. Interface of a Nagual Dance Session with an overlay of the video stream on the left.
Image: Nagual Dance

Nagual Dance was developed by the Berlin startup "Nagual Sounds" [10] from 2011 to 2015. I discovered it quite late in the process, event though back then it got a relatively broad media response, won several prizes and was financed by investors. The project used a first-generation Microsoft Kinect camera for body tracking and applied a unified interaction model that took into account both the position of the limbs and the position of the players in space. Players move freely on a 2 × 2 grid, and, depending on which field they are in, different levels of the song are played. By moving their hands and feet they activate rhythm, harmony and melody, which show variations according to the joint positions. *Nagual Dance* can also be played by two players simultaneously. One person then controls the tonal elements and the other person the percussion. The music selection consists mainly of pop and dance tracks without vocals. Abstract avatars of the players standing in the playfield are displayed on the screen.

I was able to try different iterations of *Nagual Dance* during a meeting with co-founder Matthias Strobel. Strangely enough, the whole thing felt somewhat stiff. Technically, the software worked well most of the time, but I missed an emotional spark. I can't put my finger exactly on the cause. Maybe it was also a different taste in music, but I remember constantly looking at the screen in front of me to rationally establish a connection between my action and the musical events, merely *executing* a function in the system. This can simply be my individual perception.

Nagual Dance was discontinued in 2015 after investors backed out. The company filed for bankruptcy in 2019.

2.1.4. Therapy and Performance: *Motion Composer* Video [11]

Figure 4. Image of the Motion Composer device, taken from the marketing PDF.

Image: Motion Composer

Another area in which motion-sensitive music can be used is therapeutic applications. Sonification of body movements for this is an active area of research [12]. One product in this category is *Motion Composer*. The project was founded in 2002 by choreographer Robert Wechsler and consists of a team of twelve people, including six composers. *Motion Composer* explicitly addresses people with limited movement possibilities and therefore wants to be especially accessible. The company is technically supported by the Chemnitz-based company "Fusion Systems", which offers image analysis solutions. In the artistic field Motion Composer works together with the "Palindrome Dance Company". It has also been used for purely artistic on-stage dance performances but the project seems to be strongly focused on therapeutic use cases now.

According to the offer flyer for "Motion Composer 3.0", an all-in-one solution is sold for 12.450 €. This includes a device with stereo cameras, "more than 50 music worlds", speakers, an introductory workshop and a book with activity suggestions [13].

2.2. Platform and technology

The first and in my opinion obvious idea was to publish a web-based software. This would mean the widest possible distribution and accessibility and also promises absolute freedom to me as an author, because I would not be bound to the rules of proprietary app distribution platforms, whose vendors have the last word on what can be published. I already had experience with the web-based body tracking software *Posenet* [14], which — like all modern algorithms in this field — is based on machine learning. It was developed at Google by Dan Oved and provides 2D coordinates for several people simultaneously. Back then I was disappointed with the precision of the tracking and was sure that visible progress could only be a matter of time. Although *Posenet 2* has been released in the meantime, the results are still not satisfactory for my application, which made me consider the following alternatives.

PC application with Microsoft Kinect 2

The Microsoft Kinect 2 can calculate 3D positions of up to six people through the combination of a 2D camera image and a projected grid of infrared light points. The tracking has a very low latency, and the PC as a platform is widely used. With this properties, the device is popular for media installations and such. The clear disadvantage is that not only additional hardware is required, but that this hardware is also no longer officially distributed by Microsoft. There is a successor in the form of the "Azure Kinect", but it is explicitly not aimed at consumers but at the industry [15].

PC application which uses advanced camera image-based tracking

There are now several pose estimation algorithms that calculate 3D position data in real-time from a 2D image alone. One of the most popular ones is *OpenPose* by Gines Hidalgo et. al. which calculates 2D data for several people at once, or 3D data for one person. A more recent and impressive model is *XNect* by Mehta et. al., which can calculate 3D data for several people [16]. All current algorithms in this area are based on Neural Networks and require a powerful (i.e. expensive) graphics card to provide fast tracking performance.

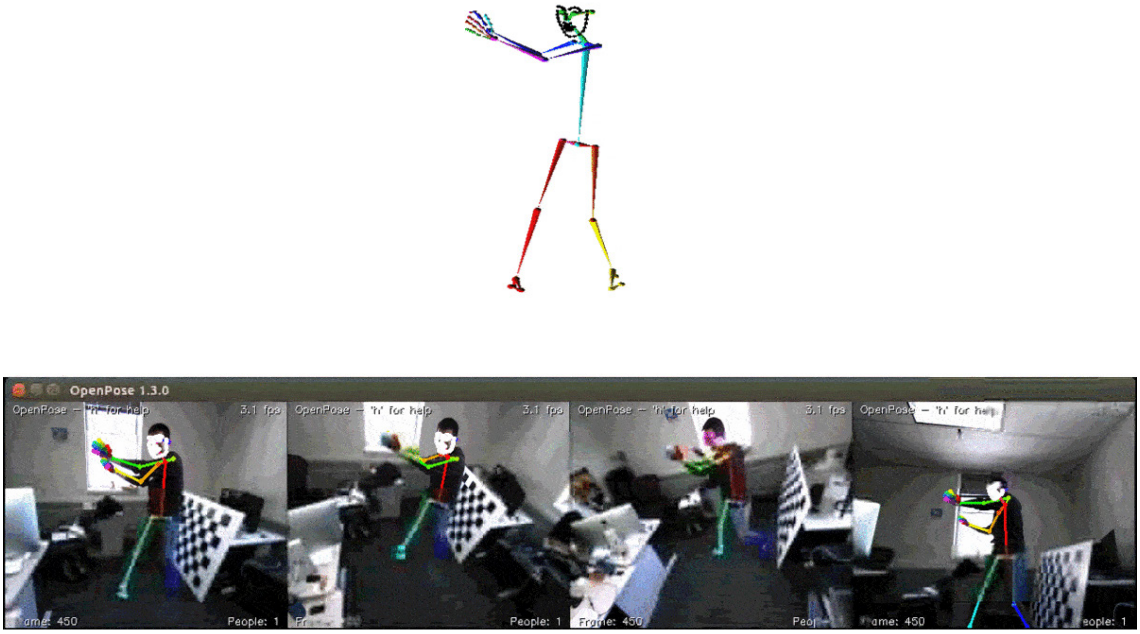


Figure 5. Demonstration of 3D body estimation from a 2D image with OpenPose.

Image: OpenPose, upper part color-inverted by the author to make it more printer friendly

iOS-App using the ***Motion Capture*** feature

Motion Capture for iOS was introduced in version 3 of Apple's proprietary augmented reality framework *ARKit* in 2019. It delivers three-dimensional position data of virtual joints of a person at a rate of about 45 frames per second. The great advantage of such a device is its mobility: it can be used in a wide variety of situations and locations. A disadvantage is the relatively high price of the supported devices.

After some consideration, I finally decided to use iOS as a platform to reach the easiest possible setup for users and the largest possible audience.

2.3. Tracking

2.3.1. API and input image

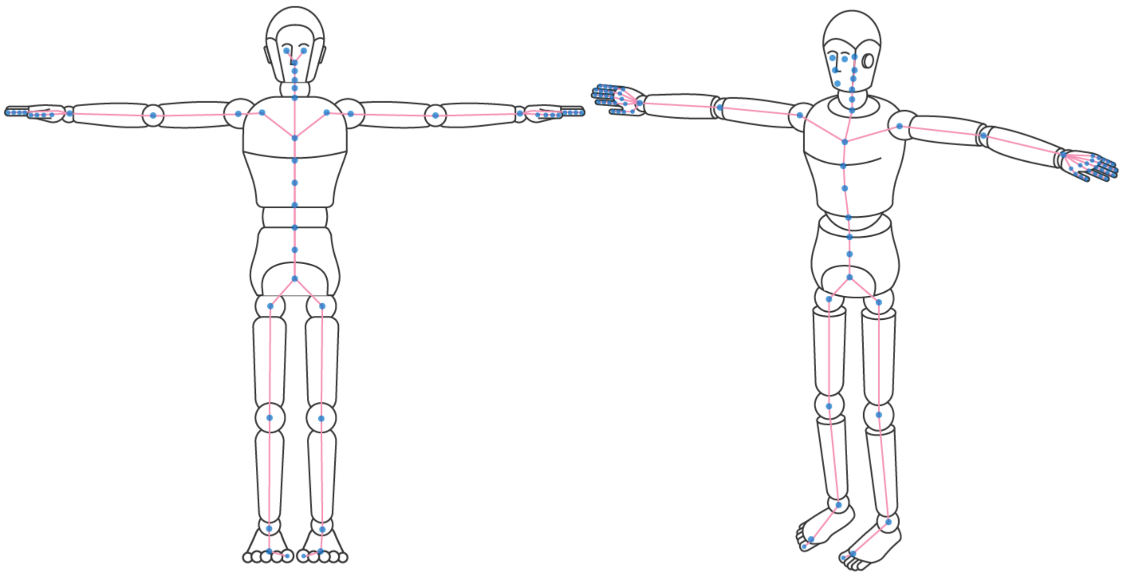


Figure 6. Illustration of the joint structure provided by "Motion Capture". This Image is taken from an article called "Rigging a Model for Motion Capture". This title emphasizes that the feature is originally made for "virtual puppeteering".

Image: Apple

The Application Programming Interface (API) for Motion Capture provides data for 91 joints, but only 26 of them are actively tracked [17]:

- **Feet**
- **Knees**
- **Hips**
- **7 spine points**
- **4 neck points**
- **Head**
- **Shoulders (between shoulder blade and collar bone)**
- **Humeral heads**
- **Elbows**
- **Hands**

The positions of the joints that are not actively tracked are added algorithmically to obtain a standardized and detailed skeletal structure to which a humanoid 3D model can be attached. The system's unit of measurement is meters.

The API requires devices with at least an "A12" chip, which was the first one to include a dedicated machine learning processor, meaning that all iPhone models released after 2018 can use this feature. Although the devices have cameras on the front and back, the feature only works with the rear camera. The input image has a resolution of 1920 × 1440 pixels. This 4:3 aspect ratio makes it possible to use the functionality in both portrait and landscape mode. In tests, a distance of 3 meters was sufficient to capture the entire body of a person in portrait or landscape format.

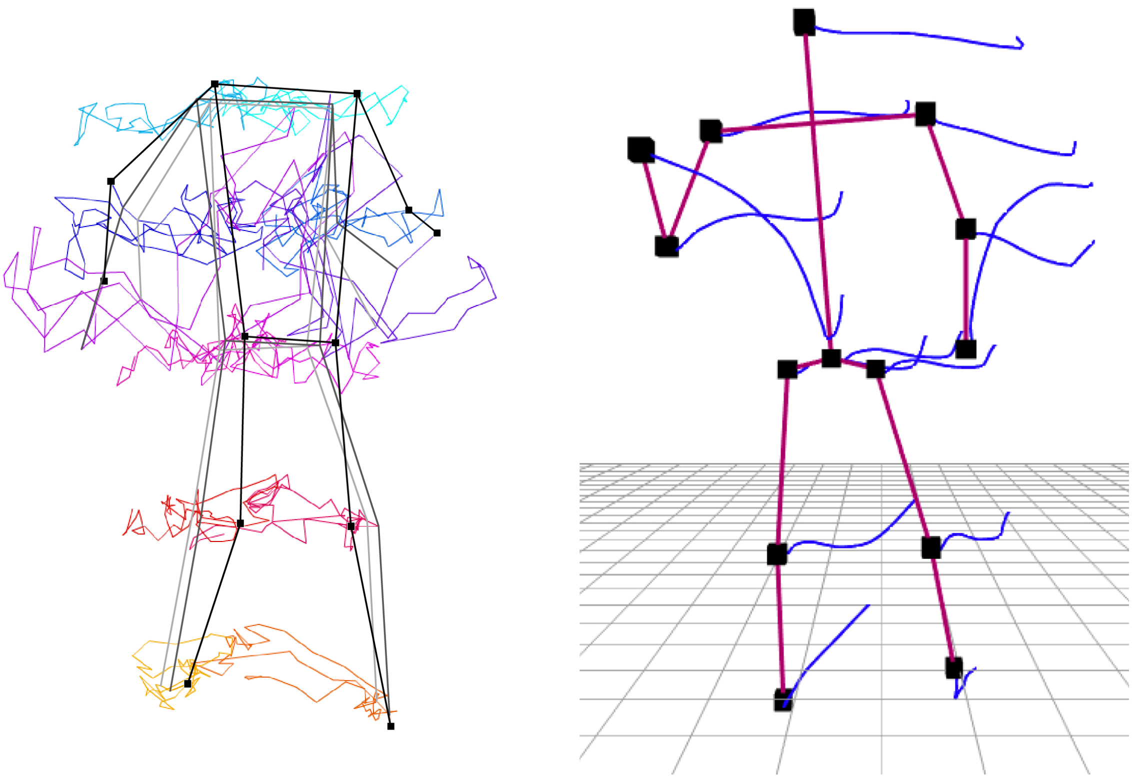


Figure 7. Left - Joint path visualisation from an older project of mine using Posenet. Right - Joint path visualisation using Motion Capture on iOS.

2.3.3. Tracking quality

Noise

Despite the name “Motion Capture” one should not expect the precision of professional motion capture systems. The accuracy of the data varies for different joints and postures, and is also influenced by lighting conditions. In one test I measured the noise of the most important joints in a static scene over a period of 10 seconds. The test person stands absolutely still in a neutral pose, and the camera’s view is unobstructed. The average size of the bounding box of the measured noise is 2.091 × 0.982 × 3.123 cm. For 12 of 14 joints, noise dominates in the z-dimension. Also the rotation of the pose is not estimated correctly: Compared to the actual scene, the virtual skeleton is rotated about 30 degrees counterclockwise. I have documented this and the following findings in video [18].

Joint ID	Noise bounding box size in cm (x, y, z)	Joint ID	Noise bounding box size in cm (x, y, z)
Root	1.840 0.862 2.731	Head	2.453 1.084 4.393
Right hip	1.790 0.968 2.540	Right shoulder	2.030 1.154 3.375
Right knee	2.045 1.092 2.171	Right elbow	1.811 1.223 2.971
Right foot	2.317 0.728 2.158	Right hand	2.020 1.238 3.102
Left hip	1.914 0.882 2.908	Left shoulder	2.370 0.876 4.516
Left knee	1.864 1.008 2.462	Left elbow	2.515 0.844 4.237
Left foot	2.277 0.765 1.782	Left hand	2.022 1.021 4.380

Table 1. Noise bounding boxes of the detected joints of a person standing still for ten seconds

Mis-detections leading to “Artificial Motion”

In another test, a person initially stands still in a static scene, with her hands resting on the height of her hips. She then raises her right hand in a circular fashion over her head, then lowers it again until it hangs by her hip. The rest of her body does not move significantly at all. The tracking data, however, suggests that not only the right arm is moving, but also all other joints. Especially the left arm and hand seem to move backward and forward, upward and downward significantly, as if this arm is swinging back and forth. This happens both during the upward and the downward movement of the right arm. Over the course of this test, the movement bounding box for the left hand is 11.351 × 3.049 × 10.301 cm in size, when it should be much closer to the size of the one recorded when standing perfectly still.

My speculation is that these errors are a result of the model trying to fit the data into its learned patterns. I faced this problem before with Posenet to a much higher extend. Unfortunately, unlike other software in this field, the API does not provide a "confidence" factor along with the estimation, so errors can not be inferred easily by the developer.

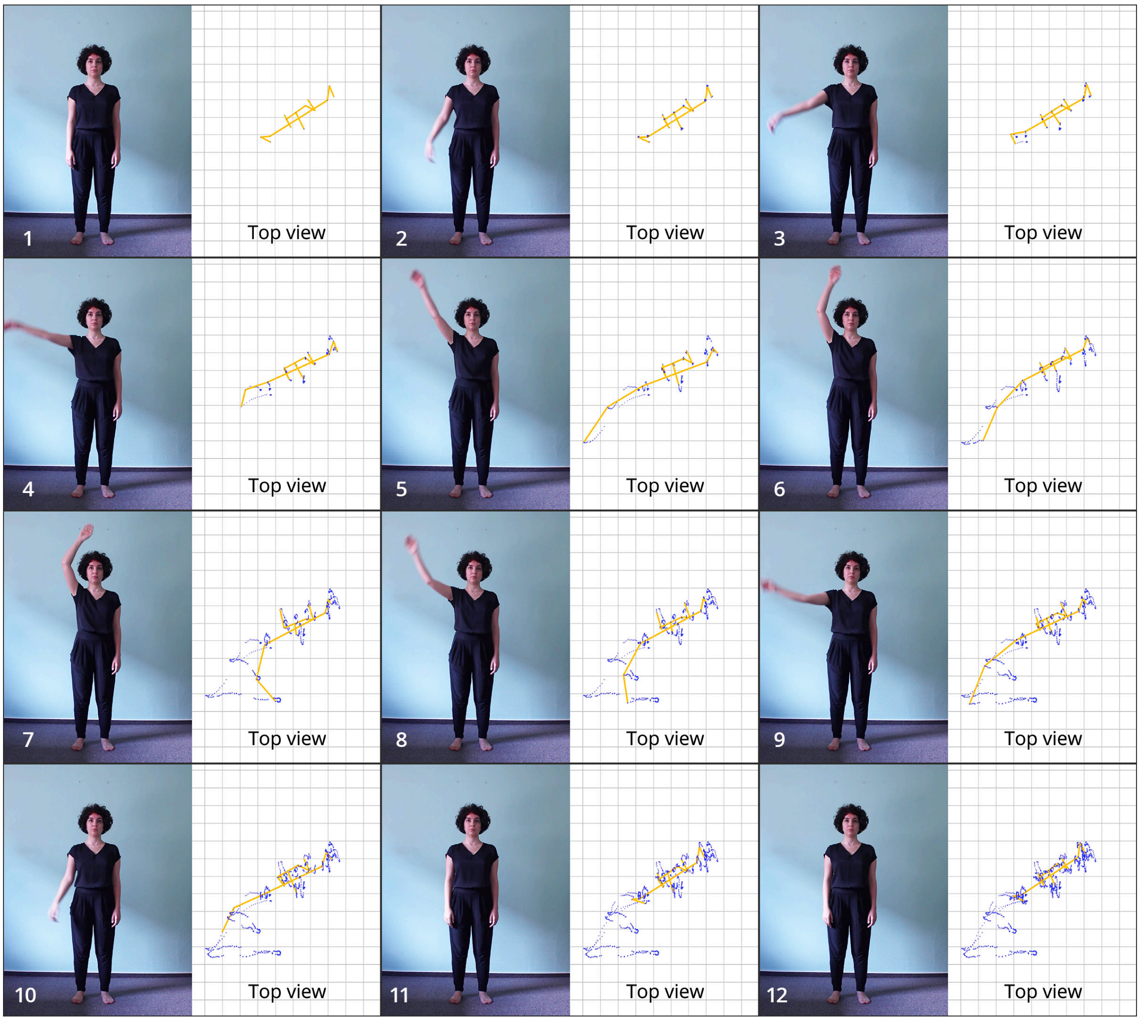


Figure 8. The sequence shows the misinterpretations of the model, here in the case of a unilateral arm movement.

2.4. Interpretation of motion data

Laban Movement Analysis

This system was conceived by Rudolf von Laban and further developed by Irmgard Bartenieff as a vocabulary for human movement in several dimensions [19]. Laban Movement Analysis (LMA) describes different spatiotemporal components of such motion, not only in dance. To do so, movement is analysed and expressed in four main categories: "Body" describes the organisation and relationships of body parts in the so called Kinesphere, which is different from the general space in which the body moves as a whole and is the concern of the "Space" category. "Effort" then describes the quality of the movement, for example the directness of a gesture (pointing somewhere VS waiving randomly) or how much kinematic force one puts behind it (punching VS nudging). Later, the system was extended by Bartenieff to include the categories "Phrasing" (temporal structure of the first four categories) and "Relationship" (describing the relationship to other subjects). LMA is referenced in a lot of research carried out in the field of human movement analysis, from robotics [20] to [21] art. I do not use the terms from LMA in my work but it is useful as a general guidance when thinking about the body.

Quantisation of body movement

In "The Dancer in the Eye: Towards a Multi-Layered Computational Framework of Qualities in Movement", Camurri et al. describe a conceptual framework for the analysis of expressive qualities of movement [22]. The research deals not only with camera input but also with other possible measurement methods, such as body mounted accelerometers, pressure sensitive floor plates, breath sounds and more. The authors divide the processing of these data into four layers with increasing degrees of abstraction. I use some aspects from the first two layers (bold):

Layer 1
"Physical Signals, Virtual Sensors" – **Position data,** boundary forms, silhouette, breathing sounds, non-verbal sounds, body function sensors (EEG etc.), weight.

Layer 2
"Low Level Features" – **Speed**, acceleration, jerk, gravity, kinetic energy of point-volume, symmetry, **contraction**, motion uniformity, "dynamic balance", "tension" of posture.

The other two layers have a high degree of abstraction and take into account more complex structures and longer time periods:

Layer 3
"Mid Level Features" – e.g. equilibrium, repetitions, simultaneity, pauses, rotation patterns.

Layer 4
"Qualities Communication" – e.g. predictability, resistance, "groove", conveyed emotion.

Visual analysis

To illustrate the tracking, I developed a testing and visualization software for the PC, as well as a rudimentary tracking app for iPhone, based on an open source project [23]. The iPhone app sends the tracking data using the Open Sound Protocol (OSC) to the PC via WiFi. Among other things, it contains various visualizations, like joint positions and motion paths. Furthermore, the bounding boxes of these paths can be displayed, as well as their average direction of movement and the bounding box of the virtual skeleton. The software already includes a module called "MotionProcessor", which contains the algorithms for processing the tracking data: Smoothing by using averaged values, calculating directions of movement, path volumes, detecting rotations. In the course of the project I tested many ideas and approaches here. The software can also store capturing sessions, which is especially useful for repeatable tests. The software can be seen in action in a video [24].

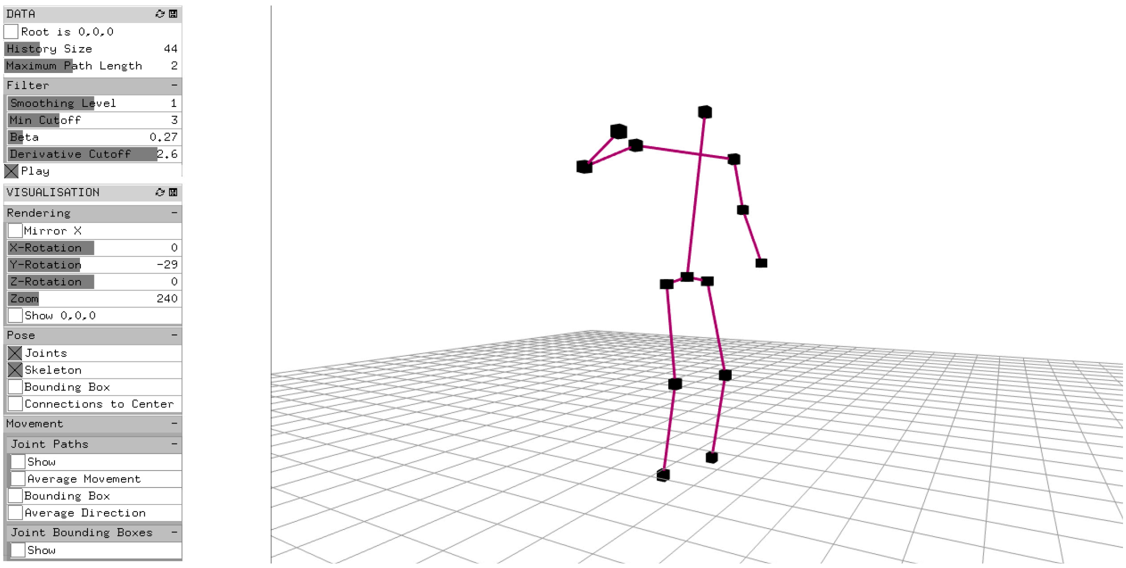


Figure 9. The visualization software in use. On the left an interface for visualisation options and calculation parameters. All these screen captures are shown with inverted colors in this document.

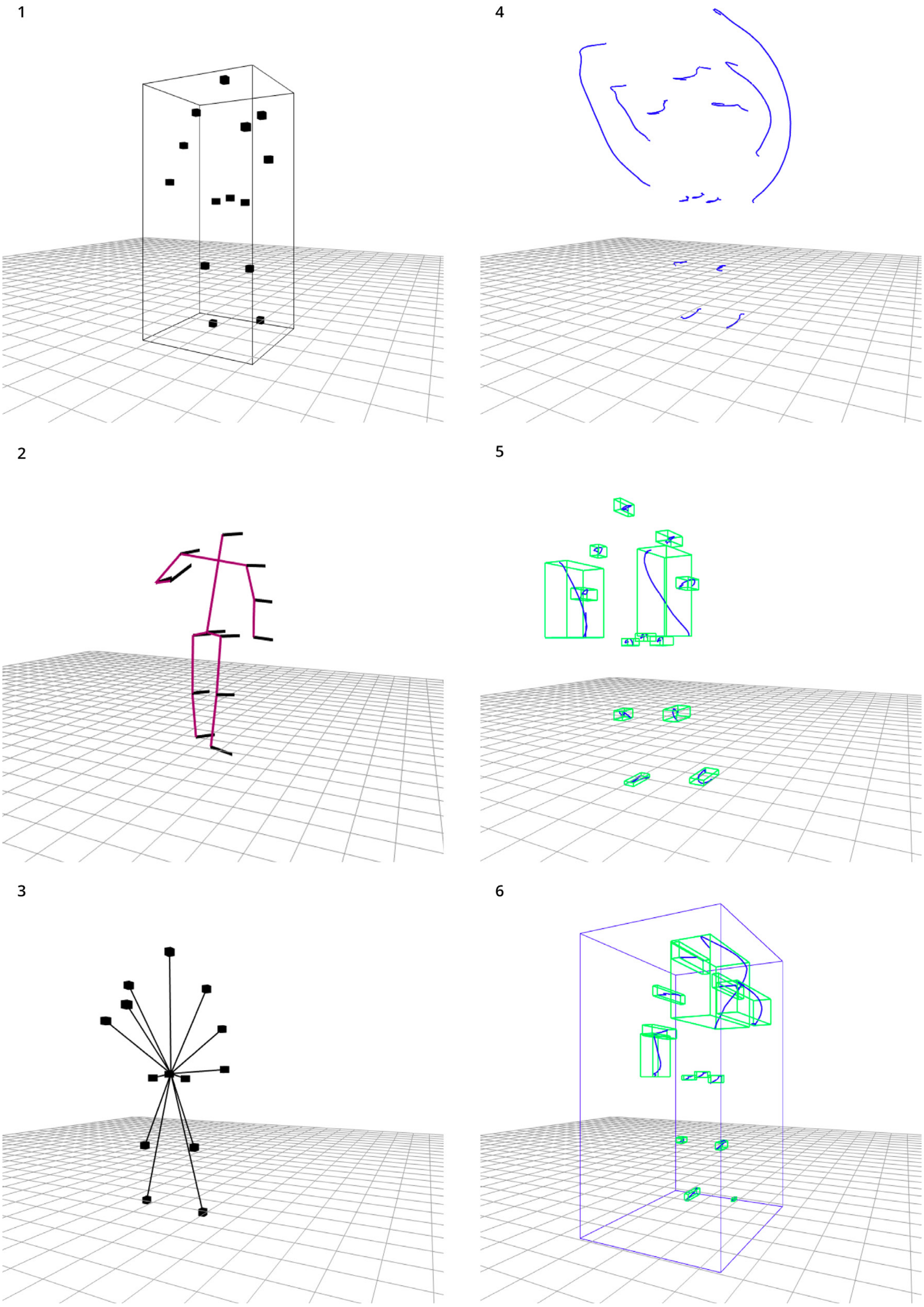


Figure 10. Montage of visualisations for different kinds of data: 1 - Joint positions, pose bounding box. 2 - Virtual skeleton, average joint directions. 3 - Joint positions, connections to center (average position data). 4 - Joint motion paths. 5 - Joint motion paths, paths bounding boxes. 6 - Motion paths, path bounding boxes, overall paths bounding box.

2.5. Electroacoustics and audio software development

I developed my first audio experiments on the PC using data received from the phone. For this I used the popular C++ framework *openFrameworks* [25], with which I also had build the visualization. Among other things I was able to create and combine oscillators from scratch and program step sequencers for sample loops. But at some point these abilities were exhausted and I would have had to go deep into audio programming myself. Since I only had a basic knowledge of electro-acoustics, this would have been beyond the time frame of this thesis. In addition, I had problems with the performance of openFrameworks on the iPhone regarding audio.

In my search for a professional iOS framework for audio, I found *Audiokit* [26]. It offers numerous features, from live synthesis, samplers, filters, effects, dynamics and much more. *Audiokit* can simulate analog devices and offers well-known techniques of electroacoustics, such as FM synthesis. There was still a steep learning curve: Compared to graphics programming (where most of my experience lies), the functionalities and performance profiles of many components were not very accessible to me. Also the phenomenon of unwanted sound behaviours like clicks and droning, especially during live synthesis with many oscillators end effects, caused problems that were unpredictable for me. Since I had previously mainly developed systems with visual feedback, debugging by ear was challenging, because I could not easily make out the connection of the sonic phenomena and their root cause. So a big part of my research in this area was learning to use the software.

3. INTERACTION

3.1. “Players” instead of “Users”

What do I call those who interact with this music? For me, this question is more than a semantic sophistry, because it expresses my view of the relationship between the work and the recipient, who is also an active contributor here. The term “listener” is therefore directly excluded because of its passive nature. One can also listen “actively”, but this means rather a mental activity. There are several other possibilities:

“Dancers”
Many types of concerted movement patterns can be called dance, especially when they happen in conjunction with music (but also without it). Therefore the term “dancer” seems to be quite appropriate for someone who interacts with my music. The problem here is my presumed effect of this term. In the understanding of many people, “dance” has a narrower meaning and may be connected to specific situations, like dancing as a performance for others, dancing in a club etc. I am concerned that this word already evokes such ideas and contexts.

“Musicians”
Since using the software means triggering and influencing sounds, they could be simply called “musicians”. There are several arguments against this. For one thing, there is the problem of perception again. The term “musician” would certainly lead many to expect that a certain amount of practical experience and theoretical knowledge is necessary. On the other hand, actual musicians would most likely miss further possibilities and feel limited by the tool. Another argument is that making music is a controlling activity; but direct control is not the focus of this project.

“Users”
“Users” use a product or tool, mostly to solve a specific problem. In the case of music, one can think of “utility music”, such as Muzak or “elevator music”, which serves a very specific purpose (pleasant acoustic background, animation for consumption, etc.). Users have an idea of sovereignty over the tool. The relationship is asymmetrical and focused on an end result.

“Players”
The “player” is known in the digital world mainly from the field of “games”. Players interact with rule systems with the aim of making certain (especially emotional) experiences. The nature of this experience is made possible by the authors and producers of the games. Players are accustomed to engage with these virtual environments with a certain degree of openness, as long as they have the feeling that they participate in the success of the experience through relevant (inter)actions.

My decision: as long as someone navigates in the app’s interface, they are traditional “users”. When they enter the performance space, they are “players”.

3.2. Dynamic design parameters

3.2.1. Sound

“Sound” first of all describes everything that takes place acoustically “in the moment”. Music is then the artistic organization of sound. When I speak of sound as a creative parameter, I do not yet mean patterns of sounds in the form of a melody for instance, but rather the acoustic phenomenon in a relatively short time frame, in the range of milliseconds to a maximum of maybe a few seconds. With electronic sound generation, a general distinction must be made between real-time synthesis and sampling. The former offers more instantaneous design freedom, but this often requires more computing performance; after all, in the case of 16 bit stereo sound, the waves are computed 88.200 times per second. The sampling of already created or recorded sounds can be used here to create more complex sound designs. A third possibility is live input via microphones (e.g. vocals) or digital instruments. However, this input source is not used in this thesis.

In real-time synthesis, the parameters are first of all wave frequency and amplitude, as well as various properties that are responsible for the “timbre”, such as pitch, waveform, noise factor, harmonic structure, and their development over time. The “Volume” parameter is equally suitable for real-time synthesis as it is for samples. If only certain ranges are modulated in this way, it is called a filter: High-pass, low-pass and band filters. In this project I work with the following dynamic parameters:

- **Pitch**
- **Amplitude**
- **Single oscillator waveforms**
- **Noise**

3.2.2. Time / Rhythm

Another essential component of music is time. This does not mean the trivial fact that sound physically cannot take place without time, but time as a dimension of design with a significant structuring function. This takes place on numerous levels, from ADSR envelopes that modulate the amplitude/volume, through successive tones that can be perceived as melodies and rhythmic patterns, to the length of individual parts or pieces:

> *"Rhythm unfolds on multiple timescales, from the micro-rhythms of the waveform, to the undulations within a note, to the higher layers of phrases and macroform." (Roads, 2015, p. 138)* [27]

Effects such as distortion or reverb are also based on the interaction of sound over time. Another factor that should not be underestimated is silence, the empty space between sound events.

Considering "beat", one could naively come up with the approach to change the overall tempo of a piece, according to the speed of movement. I actually tried this out some time ago and it was very clear that this simple method does not work well because the player would get "out of step" frequently. One explanation for this is that the perception of "beat" does not only deal with the past and the current moment but is also a very old, intuitive way to look into the future, according to research in neurosciene:

> *"The manner in which people synchronize to the beat reveals that musical beat perception is a predictive process." (Patel, Iversen, 2014)* [28]

Here, I open up the following design means to the input of the player:

- **Distances within tone sequences of a fixed tempo**
- **Parameters of effects (e.g. reverb, bitcrusher or virtual guitar amp)**
- **Frequency of amplitude modulation through another oscillator**

3.2.3. Larger compositional structures

Besides these micro- and meso-parameters there are even larger structures. These range from the arrangement of smaller parts like phrases to the overall structure of the whole piece, considering prelude, choruses, breaks etc. As far as I can see, none of the works presented in 2.1. operates on this level. *Very Nervous System* seems to come closest, but Rokeby emphasizes that the virtual observer-musicians have quite a lot of autonomy and the actual influence of the player is limited.

> *"Each instrument is basically a behavior, an electronically constructed personality. It's watching you. It's looking out of the video camera at your body, and taking playing cues from your movement. These behaviors are just algorithmic definitions - computer subroutines. I construct them to suggest whether this instrument, for instance, tends to play on offbeats, or perhaps plays on offbeats but doubles its rhythm if you move faster. […] Think of a jazz band: different players, each with his or her own style. In the case of Very Nervous System, these are the "behaviors," defined by the software. Now give good jazz players some input - say, a chord chart or an old standard - and each player will improvise within his or her own style." (Rokeby in an interview with WIRED magazine, 1995)* [29]

The interactive design of larger structures is also not part of this project. I think that the main challenge here is the following: These larger structures have a longer time frame, which means that it is not possible to react directly to momentary actions of the player, but that behavior has to be analyzed as a bigger picture. Using the research of Camurri et al, this would happen in levels three and four. Since music is a medium that has a strong emotional effect, this would mean translating the player's movement expression into emotional qualities, and then countering this with musical expression. This seems to be a very difficult undertaking, which would probably have to work with rough generalizations.

I do not believe that good results are impossible here, and would like to try it, for example, in collaboration with experienced composers. However, such an undertaking goes beyond the scope of this project.

3.3. Processing data for musical feedback

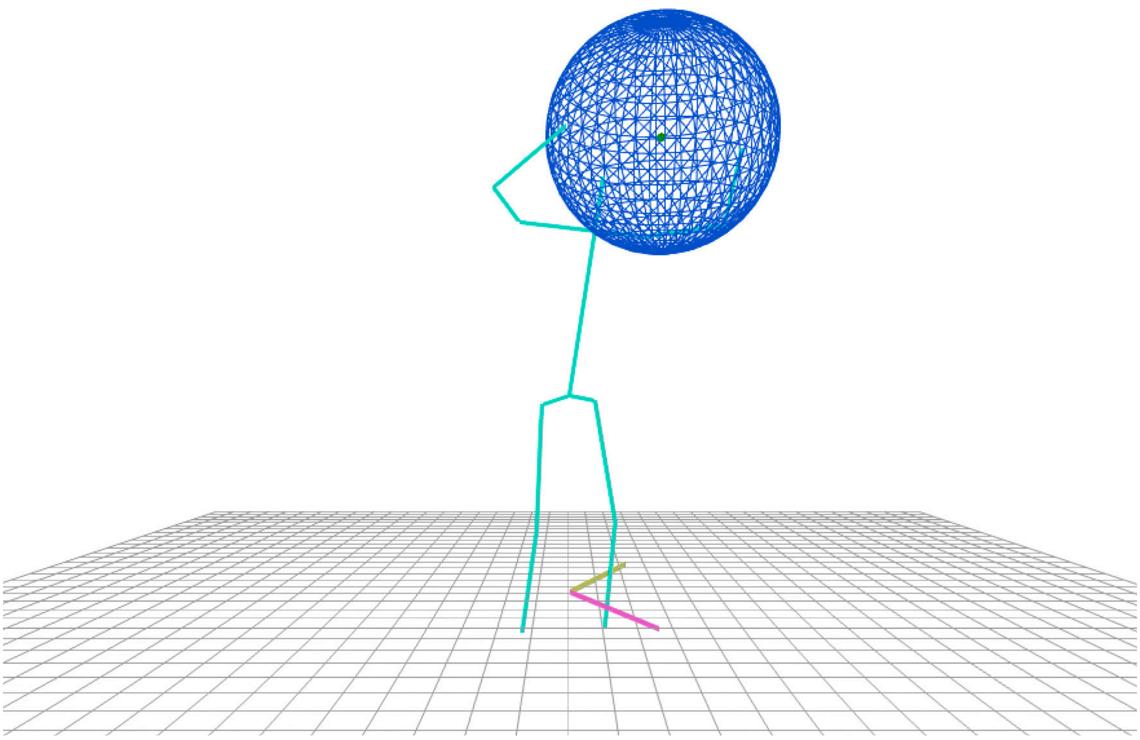


Figure 11. Early in the project I also experimented with virtual objects. The center of this virtual orb emits a pulsating sound which is manipulated according to the size and position of the sphere. Video [40]

3.3.1. Mappings

A relatively simple and basic method of translating information of various kinds into sound is to use mappings. This can also be used well for sonification. Here, an input parameter — this can be a color, a position, a speed and much more — is mapped to an output range that in some way influences the sound generation. A very simple variation would be to map the speed of a hand, i.e. the Euclidean distance from two successive positions, to the volume of the sound. Through the combination of several mappings, somewhat complex and interesting sound structures can already be created. In one experiment I multiplied the outputs of two oscillators, one generating the basic frequency (the pitch), the other oscillating at a much lower rate. The height of the basic frequency is based on the Y-rotation of the hip. The frequency of the low frequency oscillator is coupled to the movement of all limbs. A video is available online [30].

These mappings are also useful for influencing the playback of recordings. In one experiment I coupled the playback speed of the song “It’s Your Move”, in Diana Ross’ interpretation, to the intensity of the player’s movement. The song selection was not random: The cover version “リサフランク420 / 現代のコンピュー” (“Lisa Frank 420 / Modern Computing”), by MacintoshPlus [31], which is a “classic” of the “Vaporwave” genre, is mainly based on sampling this song at reduced speed (including a lower pitch). Here is a video.[32]

In principle, mappings are suitable for all effects and manipulations that are controlled by continuous parameters. I have further explored how players can intuitively and playfully “shape” the sound of an existing recording in real time through various body articulations. I use the term “shaping” because changing both the playback speed and the global pitch change make the sound appear as some kind of object with an invisible morphology. The source for this can be any recording as a flat digital file. I concentrated on vocal tracks of pop songs, especially to test the sonic effect of such a “shaping” on human singing. I tried to make the different input channels independent from each other so that they only interfere when the player wants to use more than one effect at a time. This avoidance of interdependencies should make the system flexible and open. The table below lists the parameters and the associated effects. I have documented this system in a video [33] showing the interaction with the song “Halo” by the R’n’B and soul interpreter Beyonce.

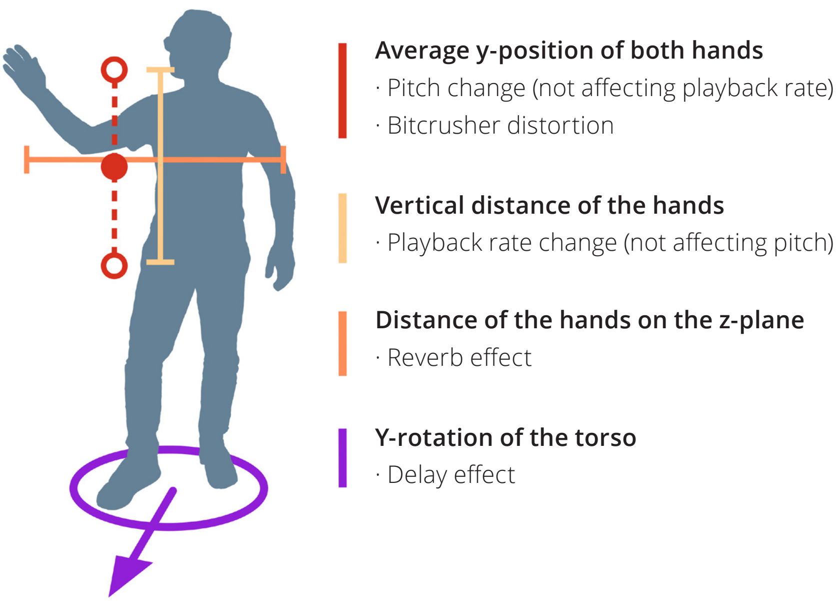


Figure 12. Effect mappings.

The audio playback and manipulations are connected in this order one after the other:

Original Playback → Time & Pitch → Delay → Distortion → Reverb

This is based on the following consideration:

- **For technical reasons, the pitch and tempo change must be applied directly to the original source**
- **The delay should be applied as directly as possible to the voice to make it sound like a "weird choir"**
- **The distortion should affect the entire sound before playback in this virtual space**
- **The reverberation should affect the entire output to create a virtual room whose size can be manipulated by the player**

Input parameter	Effect/manipulation	Input range	Output range
Average y-pos. of both hands	Pitch change (not affecting playback rate)	0.00 m … 0.60 m	-450 % … 0 %
Vertical distance of the hands	Playback rate change (not affecting pitch)	0.20 m … 0.70 m	100 % … 10 %
Distance of the hands on the z-plane	Reverb	0.50 m … 1.25 m	0 … 1
Average y-pos. of both hands	Bitcrusher Distortion (bit depth 4, sample rate 4000)	1.20 m … 2.00 m	0 … 1
Torso y-rotation	Delay mix (1 second)	0°* … 180° … 360° *facing camera	0 … 1 … 0

Table 2. Parameter mapping for articulations

3.3.2. Triggers

Where mappings have a continuous effect, triggers work discretely. An analogy from the world of physical instruments would be the striking of a string or the beating of a drum. Where this physical interaction is very direct, similar things in the virtual world require a high degree of mediation. The string to be struck is replaced by one or more threshold values, the transgression of which triggers the event, for example a boundary line in virtual space, or a deliberately set verlocity value. Of course, in this case the triggered virtual object does not give any physical feedback like the bouncing off a drumhead.

Triggers are especially useful with position and speed data, as this can be understood by the player relatively easily. My goal is a playful interaction with as little initial cognitive load as possible; i.e. I want to avoid explanations. Therefore, in my opinion, reasonably direct triggers are most effective.

From a technical point of view, triggering can be a challenge for slow movements, because from a certain point, noise and error detection lead to false positives. According to research by Luke Dahl, "jerk" is the best value to trigger onsets [34], but I could not achieve reliable results due to the quality of the tracking. I use an "indirect" velocity value instead: if the position of a joint has shifted by a minimum distance after 5 frames, a triggering motion is recognized. This approach should reduce noise in the detection. The chosen number of 5 frames is a compromise between latency and noise reduction, which I have determined by subjective tests.

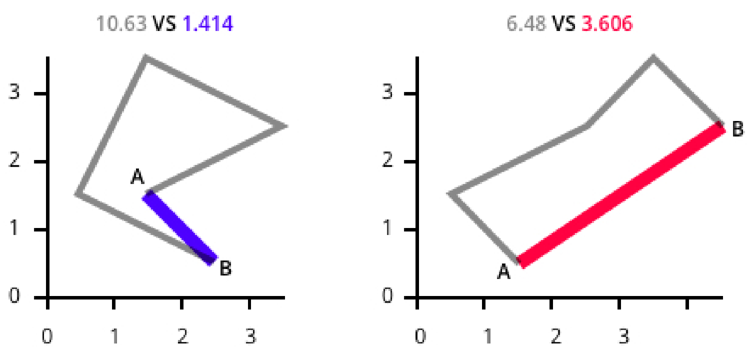


Figure 13. Illustration of detection logic. The combined length of the paths on the left is greater than on the right, but the path has less overall direction, which is inferred from the distance of the start and end points.

3.3.3. Gestures and complex data collections

Gestures have discrete start and end points, and provide additional continuous data during execution. They can be used as spatial metaphors. On touch screens, the spreading of two fingers is a conventional gesture for enlarging the displayed content, such as a map section. In this project, one gesture could be a raising of both hands. Depending on the complexity of the movement pattern, gestures require a certain degree of learning and concentration on the users part.

Gesture detection is achieved through the interpretation of data bundles beyond single positions or simple relations. They often also include temporal information. The goal of the design of such data bundles is to filter information that says something about the behavior. There is a virtually unlimited number of them, and they can also take place on different levels of abstraction. Here are some that I tried out:

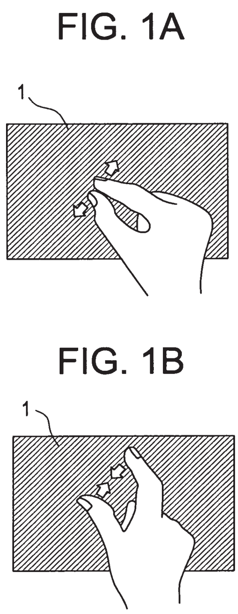


Figure 14. Drawing for a patent filing of the "Pinch to zoom" gesture.

Image: Apple

Rotation gesture

The Y-rotation of the player is calculated as the mean of the upper and lower body rotation, and the rotational speed is calculated from frame to frame. If speed above a certain threshold is sustained for a number of frames, a "rotation" gesture is detected.

Path length

The accumulated distance that a joint has travelled over time. This length is suitable for determining speed over a longer period of time, for example to ensure that the intensity of movement remains at a certain level. Path lengths of several joints can also be added together to obtain an overview and identify dominant limbs.

Complexity

The ratio of Euclidean distance of start and end point of a joints path and its total length. 1 means lowest complexity; with more complexity, this value apporaches 0.

Contraction

The total distance of the joints to the barycentre of the body. A lower value means a more contracted posture. As a barycenter needs weight values, I used old anthropometric research data from the US Aerospace Medical Research Laboratory. [35]

Bounding Boxes

The surrounding box of multiple positions, in this case the historical positions of one or more joints, representing a point cloud. This box is a rough approximation of the volume. (A more accurate approximation would be a three-dimensional convex hull.) Besides the size of such boxes, their shape can be used as input. For individual joints, bounding boxes can only be created for historical position data. From at least three joints on, a momentary bounding box can be calculated. Here too, rough conclusions can be drawn from the ratio of the side lengths for the body articulation, the most simplest being how much space the player occupies. Another use case: if a joint moves primarily vertically, this can be determined by the aspect ratio of the box.

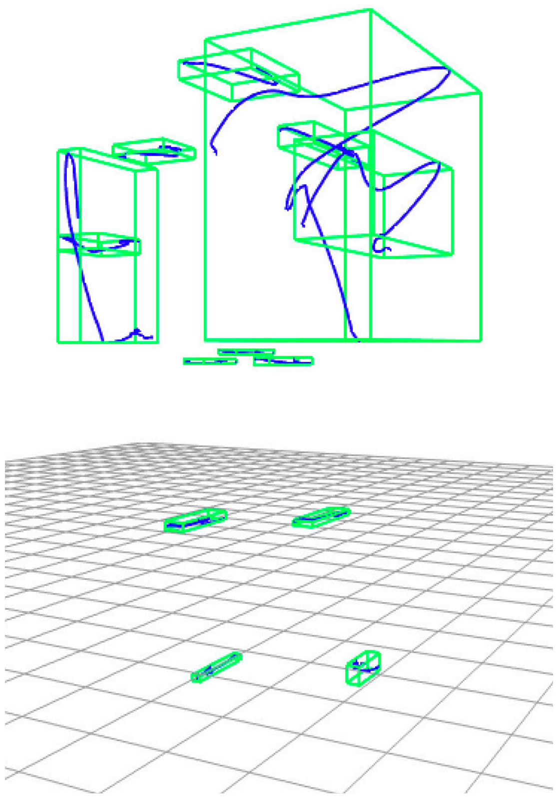


Figure 15. Bounding boxes of joint paths

Movement eccentric

This is an approach to find out how much players use different parts of their body, in different configurations. The idea is that if many motion paths intersect, the movement can be considered more "eccentric". A rough approximation of this can be achieved by counting the intersections of the movement paths bounding boxes. This is a very indirect and rough parameter, but might be useful to alter a sonic texture or trigger probability based events (for example, in a rhythm-generating algorithm).

Hand dominance

People usually have a dominant hand that they use for activities like throwing etc. To "automatically" find the dominant hand of the player, I ask them at the beginning to raise *one* hand, but do not specify which one. I assume that they will raise their dominant hand. Later, I use this parameter for the mapping of harmony and melody.

This approach is inspired by computer games: In some first person shooters, players are asked to "look" up, down, left and right before they do anything else. This is used to determine if the camera control should be inverted.

3.4. Used interaction models

3.4.1. Joint-to-Pitch Method

This is one of the essential interaction models. The basic idea behind this method is to translate the position of a joint to a pitch. It is thus a mapping whose basic design parameters are the lower and upper limits of the input and output ranges. I use linear interpolation for that, but others are possible. Input and output can also be tit in a grid, for example, twelve semitones, or the notes of a scale for the output.

We are used to the relationship between pitch and physical space, either because of actual tonal effects (such as the length of a string) or practical, but ultimately arbitrary spatial arrangements, like the key arrangement on a piano. I use the Y-axis as a dimension for the input, since here the metaphor of "high" and "low" translates well to pitch.

Based on an already existing chord progression, I use the Joint to Pitch method to vary the chords and to generate melodies as follows.

Chord

The vertical position of a hand in a certain range (e.g. 0.5 m - 1.7 m) is mapped to a range of integers which denote octaves. When the chord is played, the notes are transposed accordingly by the selected octave ("octavated").

Melody Note

First, a chord is selected as described above. Then the y-position of the hand is mapped inside the range for the resulting chord to an integer value between 0 to 2, representing the first note ("fundamental"), second note ("third") and third note ("fifth") of the chord.

I think that this basic method is suitable for further varying implementations. Thus one could perhaps use inversions instead of octavations of the whole chord to create a smooth transition. But I had no success with this so far judging from hearing the result; maybe this idea is too simplistic and thought from a technical perspective. I have presented a paper about this method, and the effects mapping method described in 3.3.1. as well, at the International Conference on Multimodal Interaction in October 2020. The paper has been published in the companion proceedings of the conference. [36]

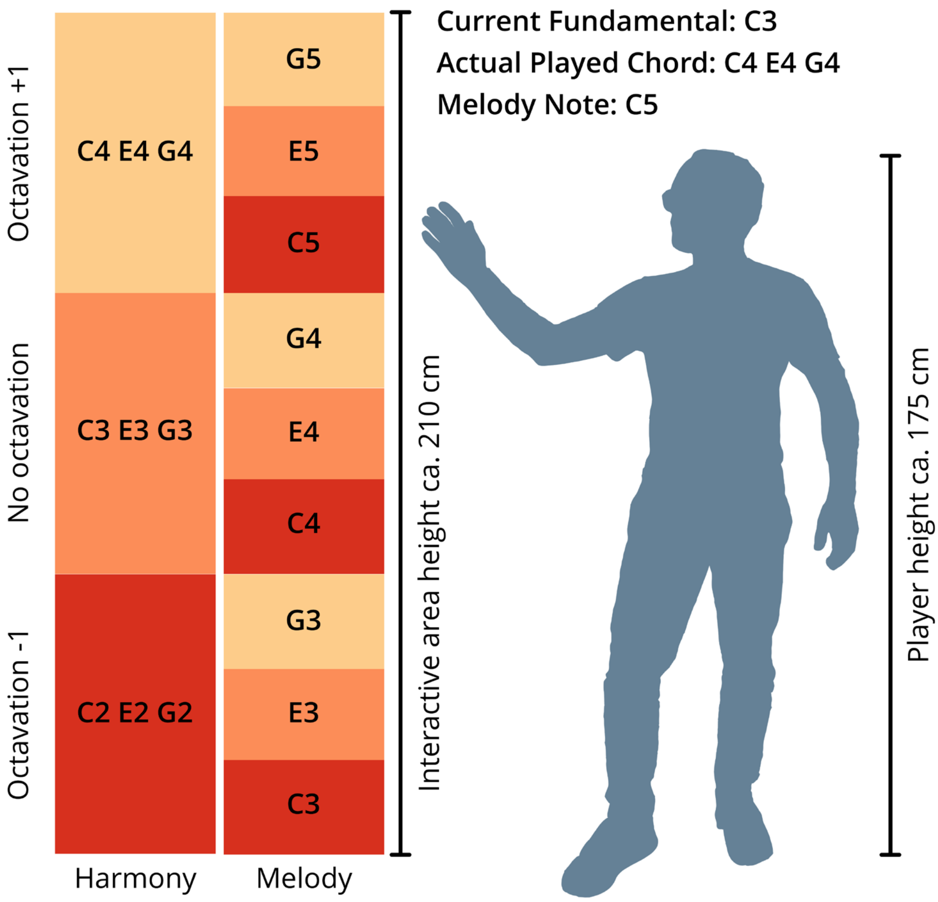


Figure 16. Example of the note selection in the Joint to Pitch method. In this case, the right hand controls the harmony as well as the melody. Additionally, the melody notes are transposed one octave further to diversify them from the harmony.

3.4.2. Others

Activity Level

Besides the positions of the player's limbs, the intensity of the movement is another important factor. To infer an "activity level", I use two parameters:

- **The speed of the joints**
- **The expansion in space**

To evaluate the speed, I use the Euclidean distance between the current position of a joint and the position 20 frames prior, which is about half a second. I explicitly do not use the path length directly (the cummulated distances from frame to frame), because this value would be more distorted by the noise (see "Tracking Quality"). The expansion in space is calculated by the volume of the bounding box of the body. These values are combined to obtain an "activity level" and are used in several ways:

- **The activity level of individual joints regulates the volume of instruments.**
- **The overall activity level can control additional effects and instrument layers, e.g. a distortion filter or the selection of drum patterns of different densities and speeds.**

Whole body characteristics

This means the appearance of the body as a whole in the space, in shape and orientation. I do not use a structured model for this kind of input but try to provide somewhat meaningful reactions, e.g. shifting the waveform of an oscillator from a sine to a square for an unstable pose. Another expression could be a stretching of both arms towards the ceiling, or being very small (contracted). I use the data bundles mentioned in 3.3.3 in an experimental way for various feedback types.

4. DEVELOPMENT AND PRODUCTION

The main software applications have been developed by me. Apple's "home user DAW" *Garage-band* was used to create samples, and the open source software *Audacity* was used for analysis and further processing and editing.

4.1. iPhone App

4.1.1. Motion Processor

The “Motion Processor” is responsible for receiving the tracking data and processing it. For each joint it stores a running histogram of 60 frames, i.e. a little more than one second. The following data is calculated:

- **Movement directions of the joints**
- **Current speed**
- **Boundary box of the current virtual skeleton**
- **Joints path length**
- **Boundary boxes of the individual joint paths**
- **Complexity of the joint paths**
- **Rotation of the hips and shoulders, and their mean value**
- **Y-rotation speed**
- **Eccentric**
- **Contraction**

There are also special classes for the detection of trigger movements like joint velocity or rotational speed. They can observe one or more joints at a time. If they are connected to multiple joints the observers can be configured to use either the average or the maximum values for their logic.

4.1.2. Music Engine

The heart of the audio part is the Music Engine. An essential quality feature should be the responsiveness of the music. Ideally, each note could be influenced by the player’s input. Therefore, an approach of merely varying the volume of a continuous playback (be it “flat” or several layers) is out of the question, because it would sound not very organic if, for example, a cymbal sound is just “cut off” when stopping the rhythm section. Much better is a format in which each note is actually triggered individually with certain parameters. The proven MIDI format offers exactly these capabilities and is therefore used here. I use it to operate virtual samplers, as well as to modulate real-time synthesis and effects.

A piece consists of one or more sections. Each section can contain one or more looping tracks which encapsulate the interactive sound generation. Tracks have at least one audio output; effect tracks also have an audio input for a source signal to work on. Theoretically, each track can contain any number of any of instruments or effects. In practice I usually use only one per track for clarity and modularization.

The timing of all tracks is handled by a step sequencer. To respond to input, tracks can access continuous data from the body tracking model, or in response to MIDI notes. They also use the observer objects mentioned in 4.1.1.

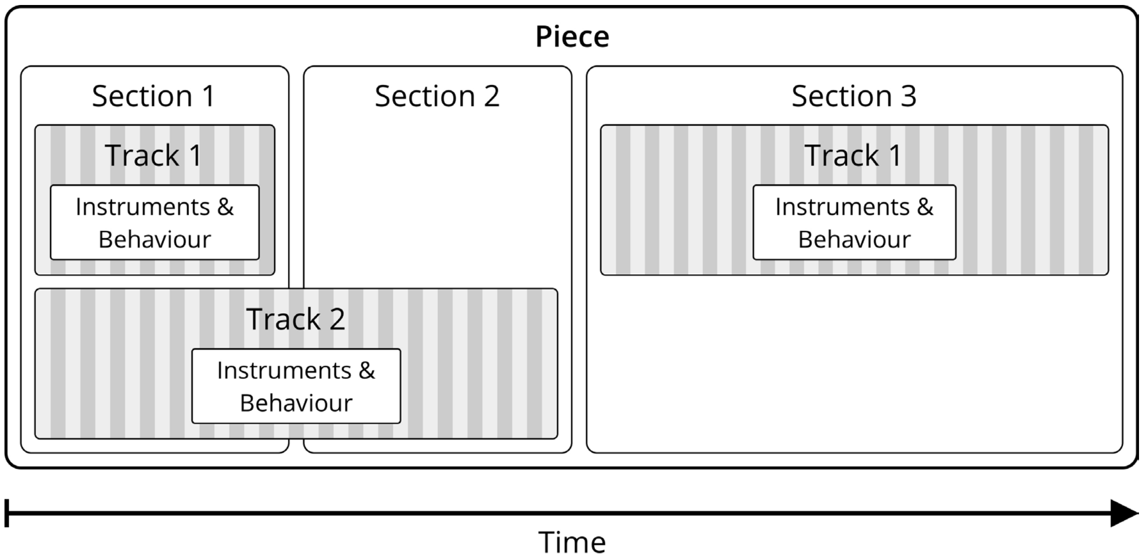


Figure 17. Technical structure of a piece.

4.1.3. Instruments

I use a number of different instruments, which I list here. Furthermore, I use a number of effects which act on the outputs of the instruments. Following is a list of basic instruments which are used in the track classes explained later.

Notes sampler
Sampler which is triggered by MIDI note and selects the appropriate sample file by note name, e.g. „piano_C#2.wav“ for note nr. 37. It also uses velocity from the MIDI data.

Continuous OSC Bank
A bank of multiple oscillators (OSC). It has a configurable harmonic structure, expressed in tuples of frequency offsets and amplitude multipliers. Each OSC can be configured to have a waveform, which is represented by a floating value between 0 and 3. It can interpolate smoothly between triangle (0), square (1), sine (2) and sawtooth (3). The main overall configurations are main frequency and main amplitude.

Triggered OSC Bank
The bank properties are identical to the continuous OSC bank. Additionally, this bank can be triggered and has configurable timings for for attack-decay-sustain-release (ADSR) envelope.

Triggered Noise Bank
Three layers with white-, pink- and brownian noise with configurable specific volumes and one ADSR envelope. Can be gated using configurable high pass and low pass filters.

Triggered String
Imitating the plucking of a string using a the Karplus-Strong-algorithm (built in in *AudioKit*), is triggered with a note as the fundamental for the "string".

Arbitrary sounds sampler
Playback of a selection of sounds, triggered, at certain intervals, or looped. Can be used for sound scapes (especially combined with effects), background layer, voice track or more.

When realizing a concrete composition, I often subclass the instruments to add custom behaviour.

4.1.3. Track Types

I designed several track classes as interfaces between body capture input and musical reaction. They implement the aforementioned interaction methods and can mostly use either samples or live synthesis. Following is an overview of the basic track types.

Variable volume track

This is the most basic interface between motion input and sound output. It simply maps the current velocity to the volume of a continuously played sound, be it a recording or live synthesis.

Input:	Configuration:
• **Joint velocity**	• **Velocity-volume mapping**

Triggered note sequence

A sequence of notes and timings. The current note in the sequence is only played if the joint velocity passes a certain threshold. The velocity of the playback is mapped to the input velocity. An exemplary mapping would be from [0.05, 0.25] m/s to [32, 128] MIDI velocity.

Input:	Configuration:
• **Joint velocity**	• **Note sequence**
	• **Velocity mapping**

Triggered transposed note sequence

This track inherits the playback behaviour of the "Triggered note sequence" and additionally transposes the notes, using the "Joint to Pitch" method.

Input:	Configuration:
• **Joint velocity**	• **Note sequence**
• **Joint y-coordinate**	• **Velocity mapping**
	• **Transposition mapping**

Triggered transposed chord sequence

This track inherits the behaviour of the "Triggered transposed note sequence" but uses chords instead of single notes. There is one velocity mapping for each note of the chord. Those mappings are "staggered" from the fundamental upward. This means that the fundamental is more easily triggered than the second note which is still "more responsive" than the third note (and so on, if more than three notes are used). At maximum input velocity, all notes are played with the same maximum velocity.

Input:	Configuration:
• **Joint velocity**	• **Chord progression**
• **Joint y-coordinate**	• **Transposition mapping**
	• **Velocities mapping**

Dynamic melody

This track inherits the behaviour from the "Triggered transposed note sequence" but, like the "Triggered transposed chord sequence", uses a chord progression as a base. At playback timing, the ad-hoc-note selection picks the current note using the "Joint to Pitch" method.

Input:	Configuration:
• **Joint velocity**	• **Chord progression**
• **Joint y-coordinate**	• **Transposition mapping**
	• **Velocities mapping**

Elastic Dynamic Melody

This track inherist the behaviour from the "Dynamic melody" and adds the parameter of "cadence". This means that the current note might be triggered after the base timing and fill up the space until the next timing, at 8th or 16th notes. The actual cadence is based on a mapping to the joint velocity and is determined at the base trigger timing.

Input:	Configuration:
• **Joint velocity**	• **Chord progression**
• **Joint y-coordinate**	• **Transposition mapping**
	• **Velocities mapping**
	• **Cadence mapping**

Dynamic drum machine

This drum machine is basically a combination of instrument tracks with respective individual patterns. The basic interaction method is to trigger the tracks based on a threshold value. Additionally, each sub-track can have several layers, corresponding to different levels of movement intensity. Other, custom interaction methods can be built in, e.g. the selection of the current instrument set based on the raise of the players hands, or body rotation etc.

Input (per track):

- **Joint velocity**

Configuration (per track):

- **Note pattern**
- **Trigger velocity**
- **Layer selection mapping**

Effect and filter tracks

These tracks do have an audio input as well as an output. Apart from that, there is no basic implementation because the configuration and behaviour of these tracks is very specific with each use case, as the mapping of motion- and effect parameters varies a lot with each effect. What they do have in common is that the parameter mapping is applied continuously. An example would be a reverb track with a virtual room size linked to the volume of the posture bounding box. I will list some concrete examples in section 4.2.

Non-interactive tracks

Static loop

Static loops are just that: a looping sample which starts with a fixed volume. It can fade out and in.

Sample sequence

A looping sequence of arbitrary sample files. The selection of sample files can be randomized and the master volume can be configured at initialization.

When realizing a concrete composition, subclasses of those tracks are often extended with additional behaviours, by adding conditions which are only useful for a specific part of that composition.

The next page shows simplified examples in pseudo code of an instrument track and an effect track connected to it.

```
MELODY_TRACK

// DATA:
instrument
motionDetectorLeftHand

// PROCEDURES:
onTiming(chord, velocity) {
	if motionDetectorLeftHand.detectsMotion {
		var note = selectNote(chord)
		velocity = selectVelocity(velocity)
		instrument.playNote(note, velocity)
	}
}

selectNote(chord) {
	if BodyTracking.handLeft.y > BodyTracking.height * 0.8 {
		return chord.topNote
	} else if BodyTracking.handLeft.y > BodyTracking.height * 0.5 {
		return chord.middleNote
	} else {
		return chord.fundamental
	}
}

selectVelocity(velocity) {
	if motionDetectorLeftHand.velocity < 0.1 {
		return velocity * 0.5
	} else {
		return velocity
	}
}

REVERB_TRACK

// DATA:
input
reverbEffect

// PROCEDURES
udpateReverb() {
	reverbEffect.dryWetMix = BodyTracking.playerVolume
}
```

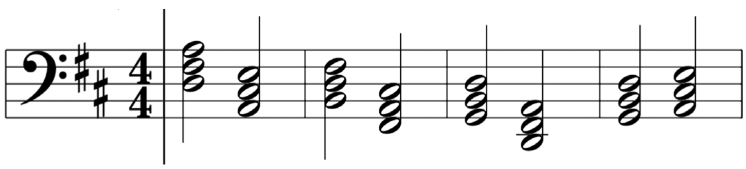
4.2. Compositions

Introduction: “Pachelbel & You”

To introduce players to the concept carefully, I offer an introductory piece. This approach is inspired by computer game tutorials. Modern games often start at a training level, where players are familiarized with basic movements and (inter)actions. The introduction has three main goals:

- **Familiarize the player with the Joint to Pitch mechanics, even if only on a sub-conscious level**
- **Convey that intensity has relevance**
- **Evoke a certain playfulness by assuring that there are no rules to follow and no “penalty” system**

The piece is based on the chord progression of the well-known piece “Canon in Gigue in D-Major” by Johann Pachelbel (“Pachelbels-Canon”): D A h f# G D G A



The piece has only one part so as not to ask too much of the player at once. The idea is to gather a few interactive instruments and let the player try out this little “orchestra”.

This orchestra is comprised of five tracks. The main roles are played by dynamic harmony and melody tracks. They are supported by two lower and simpler note sequences which are inspired by the “basso continuo” of the original composition. To accent the use of space, I added some drums which are triggered by rotational movement. With this ensemble, it is quite possible to already create a dramaturgy, as I show in my interpretation in this video [37].

I deliberately chose a very well-known piece so that the dynamics are easily noticeable while playing it. It is also very harmonic and melodious. I named this piece “Pachelbel & You” to hint at the role which the player plays in performing the piece. The piece has a length of 32 bars with a 4:4 beat; the tempo is 80 BPM.

Following is an overview of the parts and the tracks used.

Triggered transposed chord sequence

Joint.. Dominant hand

Instrument .... Notes sampler: Piano

Chord sequence .... Main progression

Octave transposition .... -1 to +2

Elastic dynamic melody

Joint.. Secondary hand

Instrument .... Notes sampler: Harp

Chord sequence .... Main progression

Octave transposition .... -1 to +2

Cadence .... 4th note to 16th note

Triggered transposed note sequence

Joint.. Left foot

Instrument .... Triggered OSC bank. Amplitudes: (0, 0.4) , Offsets: (0.5, 0.25)

.... Wave: 2. ADSR: (0.1, 0.1, 0.5, 2)

Note sequence................................. d d h h g g g g

Octave transposition .... 0 - 1

Triggered transposed note sequence

Joint.. Right foot

Instrument .... Triggered OSC bank. Amplitudes: (0, 0.4) , Offsets: (0.5, 0.25)

.... Wave: Sine. ADSR: (0.1, 0.1, 0.5, 2)

Note sequence................................. d d h h g g g g

Octave transposition .... 1 - 2

Dynamic drum machine

Global trigger .... Rotation gesture

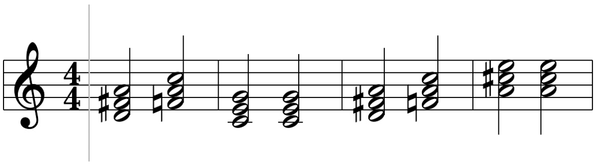
Layers

Instrument	**Trigger/Condition**	**Pattern (1/4 bar)**
Low orchestral drum	Hands are not raised	· · · x
Low orchestral drum (2)	Hands are not raised	x x · x
Orchestral drum	Both hands raised over chest	· · · x
Orchestral drum (2)	Both hands raised over chest	x x · x

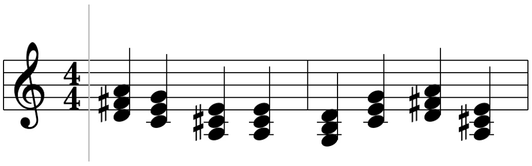
“Nefrin”

This is currently the only completed piece which I have composed myself. While “Pachelbel & You” only consists of one part, Nefrin is subdivided roughly into five parts. The compositional goal is to create a musical narrative whose different moods are interpreted by the player through their movement. This narrative itself describes a kind of journey, during which the protagonist moves from a simple, friendly and uncluttered environment to an increasingly complex, sometimes uncomfortable and challenging world.
There are four chord progressions used in this piece to which I will refer to as “Progression 1, 2, 3, 4”. The note lengths are only their base timing; depending on the track type they are divided into 2, 4 or 8 shorter notes.

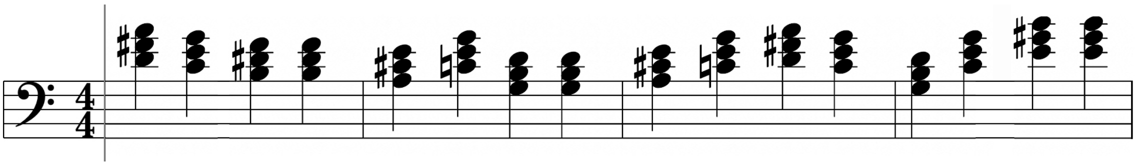
Progression 1: D3 F3 C3 C3 D3 F3 A3 A3



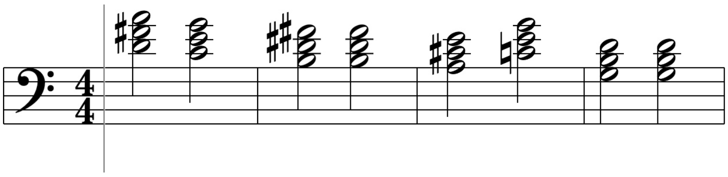
Progression 2: D3 C3 A2 A2 G2 C3 D3 A2



Progression 3: D3 C3 B2 B2 A2 C3 G2 G2 A2 C3 D3 C3 G2 C3 E3 E3



Progression 4 (shorter version of Progression 3): D3 C3 B2 B2 A2 C3 G2 G2 A2



The piece has a length of 104 bars with a 4:4 beat; the tempo is 90 BPM (simplified, technically there is actually a small section in 180 BPM in the intro). Following is an overview of the parts and the tracks used.

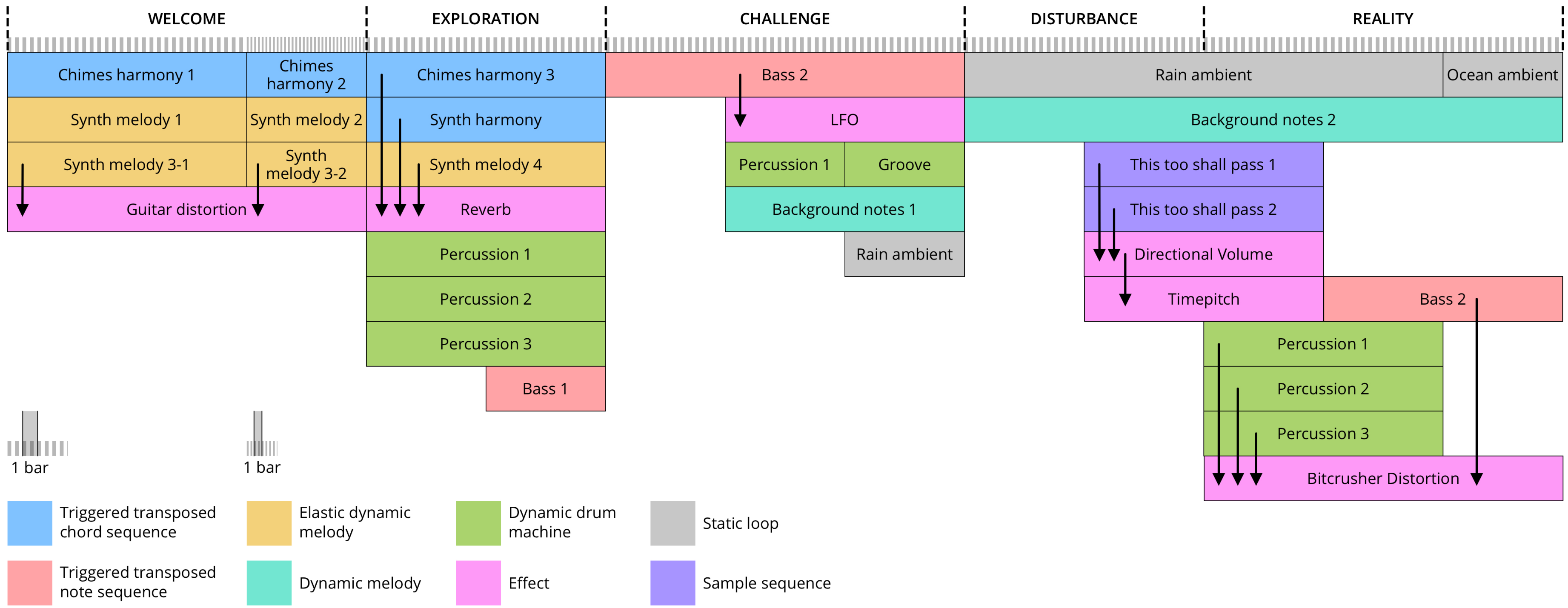


Figure 18. Overview of the piece. The parts are explained on the next pages.

Triggered transposed chord sequence

Name	Trigger joint	Chord sequence	Octavation transposition range	Instrument	Details
Chimes Harmony 1	Dominant hand	Progression 1	-1 … 2	Notes sampler: Chime samples	- - -
Chimes Harmony 2		Progression 2			
Chimes Harmony 3					
Synth Harmony	Secondary hand		-1 … 3	Triggered single OSC	OSC configuration: ADSR (0.25, 0, 0, 2), waveform index 2.05

Elastic dynamic melody

Name	Trigger joint	Chord sequence	Octavation transposition range	Cadence range	Instrument	OSCs configuration
Synth Melody 1	Secondary hand	Progression 1	-1 … 2	1 … 3	Triggered OSC bank with 2 OSCs	ADSR (0, 0, 0, 4), amplitudes (1, 0.5), frequency offsets (0, 1), waveform index 2
Synth Melody 2		Progression 2				
Synth Melody 3-1	Both feet (either or)	Progression 1	-1 … 4			ADSR (0.5, 0.25, 0.5, 2), frequency offsets (0, 1), waveform index 2.05
Synth Melody 3-2						
Synth Melody 4		Progression 2				

Dynamic melody

Name	Trigger joint	Chord sequence	Octavation transposition range	Instrument	OCSs configuration
Background Notes 1	Both feet (either or)	Progression 3	0 … 2	Continuous OSC bank with 2 OCSs	Amplitudes (1, 0.5), frequency offsets (0, 1), waveform index 2.05, ramp duration 0.2 sec
Background Notes 2		Progression 4		Continuous OSC bank with 3 OCSs	Amplitudes (1, 0.5, 0.25), frequency offsets (0, 0.5, 0.75), waveform index 1.97, ramp duration 0.5 sec

Triggered transposed note sequence

Name	Trigger joint	Note sequence	Octavation transposition range	Instrument	Details
Bass 1	Both hands (either or)	Fundamentals of progression 2	-1 … 3	Continuous OSC bank with 3 OSCs	3 OSCs, configuration: amplitudes (1, 0.5, 0.5), frequency multipliers (0, -0.5, -0.25), initial waveform index 2, ramp duration 0.025 sec. The waveform index is constantly mapped to the motion path length of the secondary hand along these ranges: 0.25m … 1.5m, 2 … 2.3
Bass 2		Fundamentals of progression 3			

Static loop

Name	Sound
Rain Ambient	Heavy rain soundscape
Ocean Ambient	Ocean waves soundscape

Sample sequence

Name	Instrument	Details
This too shall pass 1	Arbitrary sounds sampler	Sounds are randomly selected from four voice samples. They are varieties of the phrase „Auch das geht vorbei" (German for „This too shall pass")
This too shall pass 2		Sounds are randomly selected from four voice samples. They are varieties of the phrase „īn nīz bogzarad" (Persian for „This too shall pass")

Effect

Name	Description
Guitar Distortion	This „Rhino" guitar distortion gives a little „sting" to the input tracks. The distortion amount is mapped to the horizontal distance of the hands along these ranges: 0m … 1m, 0 … 0.5
Reverb	This dry-wet-mix of this simple reverb effect is mapped to the total volume of the motion paths from the hands and feet, with a range of 0 … 1 for both input and output.
LFO	Modulates the amplitude of the input signal. The modulation oscillates between 0 and 1 times, multiplied by an „intensity" factor which is mapped to the horizontal distance of the hands along these ranges: 0.25m … 1m, 0 … 1 The rate of the oscillation is mapped to the same input but along these ranges: 0.5m … 1m, 25hz … 4hz
Directional Volume	The volume of the source is set according to the direction the player is facing. It is mapped to the distance of the current y-rotation and a target angle. Input track one has the target angle 90° (facing to the right), input track two has the target angle 270°. So, for example, when the player is rotated to the left, input track 1 will not be audible.
Timepitch	The rate is mapped to the horizontal hands distance along these ranges: 0m … 1m, 3.125% … 100% The pitch is mapped to the size of the pose bounding box along these ranges: player height * 0.25 … player height * 1.5, -1000% … 1000%
Bitcrusher Distortion	Bit-depth is mapped to the vertical hands distance along these ranges: 0m … 1m, 16 … 4 Frequency is mapped to the horizontal hands distance along these ranges: 0m … 1m, 22050hz … 441hz

Table 3. Description of tonal tracks in Nefrin.

<table>
<tr><th colspan="4">Dynamic drum machine</th></tr>
<tr><th colspan="4">Percussion 1</th></tr>
<tr><td>Instrument</td><td>Pattern</td><td>Trigger</td><td>Details</td></tr>
<tr><td>Cabasa</td><td>x·······x···x···············x···</td><td>Dominant hand</td><td rowspan="5">The layers are played at normal or double speed, depending on the body diagonal of the pose bounding box. The threshold is 1.5m</td></tr>
<tr><td>Open hi hat</td><td>····················x···········</td><td>Secondary hand</td></tr>
<tr><td>Noisy click</td><td>········x···············x·······</td><td>Both hands together</td></tr>
<tr><td>Noisy click 2</td><td>x···x···x·······x···············</td><td>Right knee</td></tr>
<tr><td>Low noisy click</td><td>x···············x···············</td><td>Left knee</td></tr>
<tr><td>Tonewood</td><td>x·x·x···x·x·x···x·x·x···x·x·x···</td><td>Rotation gesture</td><td>This layer does not change tempo</td></tr>
<tr><th colspan="4">Percussion 2</th></tr>
<tr><td>Instrument</td><td>Pattern</td><td colspan="2">Trigger</td></tr>
<tr><td>High tom</td><td>··x·············</td><td colspan="2" rowspan="5">Speed of both feet and hands higher than 0.1m per 10 frames, but lower than 0.2 m per 10 frames</td></tr>
<tr><td>Piccolo tom</td><td>·····x········x·</td></tr>
<tr><td>Percussive string</td><td>········x··x·x··</td></tr>
<tr><td>Piccolo tom 2</td><td>x···x···x···x·x·</td></tr>
<tr><td>Wooden tom</td><td>x···x···········</td></tr>
<tr><th colspan="4">Percussion 3</th></tr>
<tr><td>Instrument</td><td>Pattern</td><td colspan="2">Trigger</td></tr>
<tr><td>Synth kick</td><td>x·······x·······</td><td colspan="2" rowspan="8">Speed of both feet and hands higher than 0.2 m per 10 frames</td></tr>
<tr><td>Synth cabasa</td><td>····x·······x···</td></tr>
<tr><td>Wooden rim click</td><td>···x···x··x···x·</td></tr>
<tr><td>High tom</td><td>··x·············</td></tr>
<tr><td>Piccolo tom</td><td>·····x········x·</td></tr>
<tr><td>Piccolo tom 2</td><td>········x··x·x··</td></tr>
<tr><td>Wooden tom</td><td>x···x···········</td></tr>
<tr><td>Type writer</td><td>xx··xx··xx··xxx·</td></tr>
<tr><th colspan="4">Groovy Rhythm</th></tr>
<tr><td>Instrument</td><td>Pattern</td><td>Trigger for all layers</td><td>Layer-specific trigger</td></tr>
<tr><td rowspan="3">Kick drum</td><td>x···············x·····x·········</td><td rowspan="11">Any knee is moving OR both hands are moving</td><td>Motion eccentric < 0.5</td></tr>
<tr><td>x···············x·····x·········</td><td>Motion eccentric < 0.75</td></tr>
<tr><td>x·······x··x····x·······x··x····</td><td>Motion eccentric >= 0.75 OR at least one hand raised over the chest</td></tr>
<tr><td>Snare drum</td><td>········x···············x·····x·</td><td>Motion eccentric < 0.5</td></tr>
<tr><td>Clap</td><td>········x···············x·····x·</td><td>Motion eccentric < 0.75</td></tr>
<tr><td>Snare drum</td><td>····x·······x··x····x·······x··x</td><td>Motion eccentric >= 0.75 OR at least one hand raised above the chest</td></tr>
<tr><td>Cabasa</td><td>x·x·x·x·x·x·x·x·x·x·x·x·x·x·x·x·</td><td>Motion eccentric < 0.5</td></tr>
<tr><td></td><td>x·xxx·x·x·xxx·x·x·x·x·x·x·x·xxx·</td><td>Motion eccentric > 0.5</td></tr>
<tr><td>Wood</td><td>····x·······x·····x·······x·····</td><td>Motion eccentric > 0.5</td></tr>
<tr><td>Drum broom</td><td>··x···x····x······x···x····x····</td><td>Motion eccentric > 0.5 AND at least one hand raised above the chest</td></tr>
</table>

Table 4. Description of rhythmic tracks in Nefrin.

4.3. User interface

The graphical user interface is deliberately kept simple, with the focus on the song selection. The first time the app is started, players go through an“onboarding”, i.e. an explanation of how to use the software. However, in this case this only means the basic functionality, the space requirements and the setup, not an explanation of the interaction methods.

Once the user has selected a piece, an instruction appears on how to set up the device. This initial setup process is a critical point. The device has to be leaned against the wall on the floor with the rear camera facing the players. This is not intuitive for users as they are used to looking at the screen and seeing themselves in it when using the camera. Initially I tried to explain the setup with text and an illustration, but it turned out that users ignored this information and failed during setup. Finally I made the textual explanation shorter and clearer and, probably even more important, integrated a video loop that shows the setup. [38]

Until the device is in the correct position, music with a waiting loop character is played. Once the device is set up, a voice will ask the user to position themselves at least three steps away from the phone. The app automatically detects when the user has reached this position and the voice prompts them to raise any hand above their head. As soon as this happens, the piece starts.

There is also the option to record a video of the performance. This can then be edited, saved and shared by the user. More than half of the testers have expressed desire for such functionality.

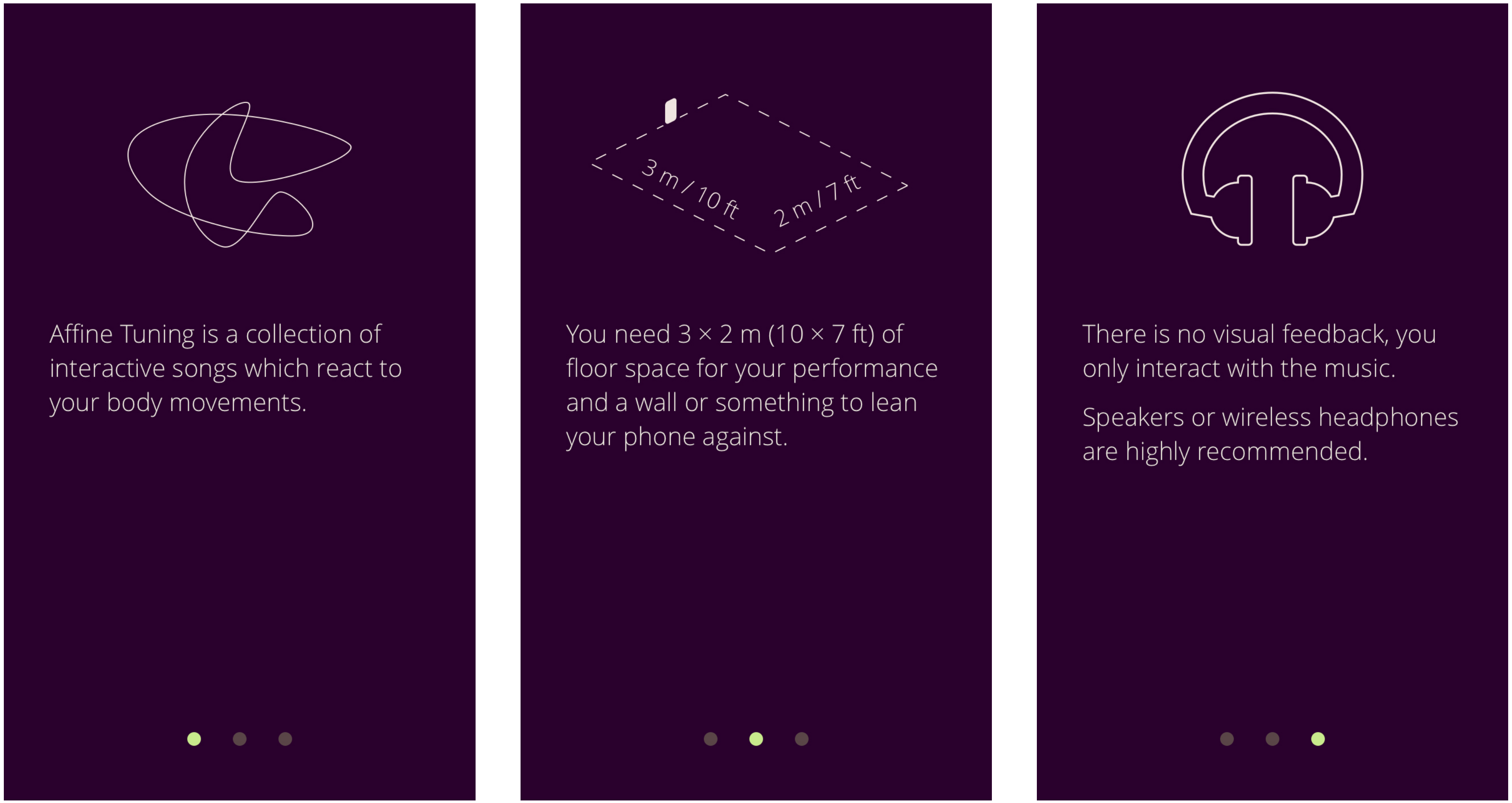


Figure 19. Onboarding screens provide an introduction.

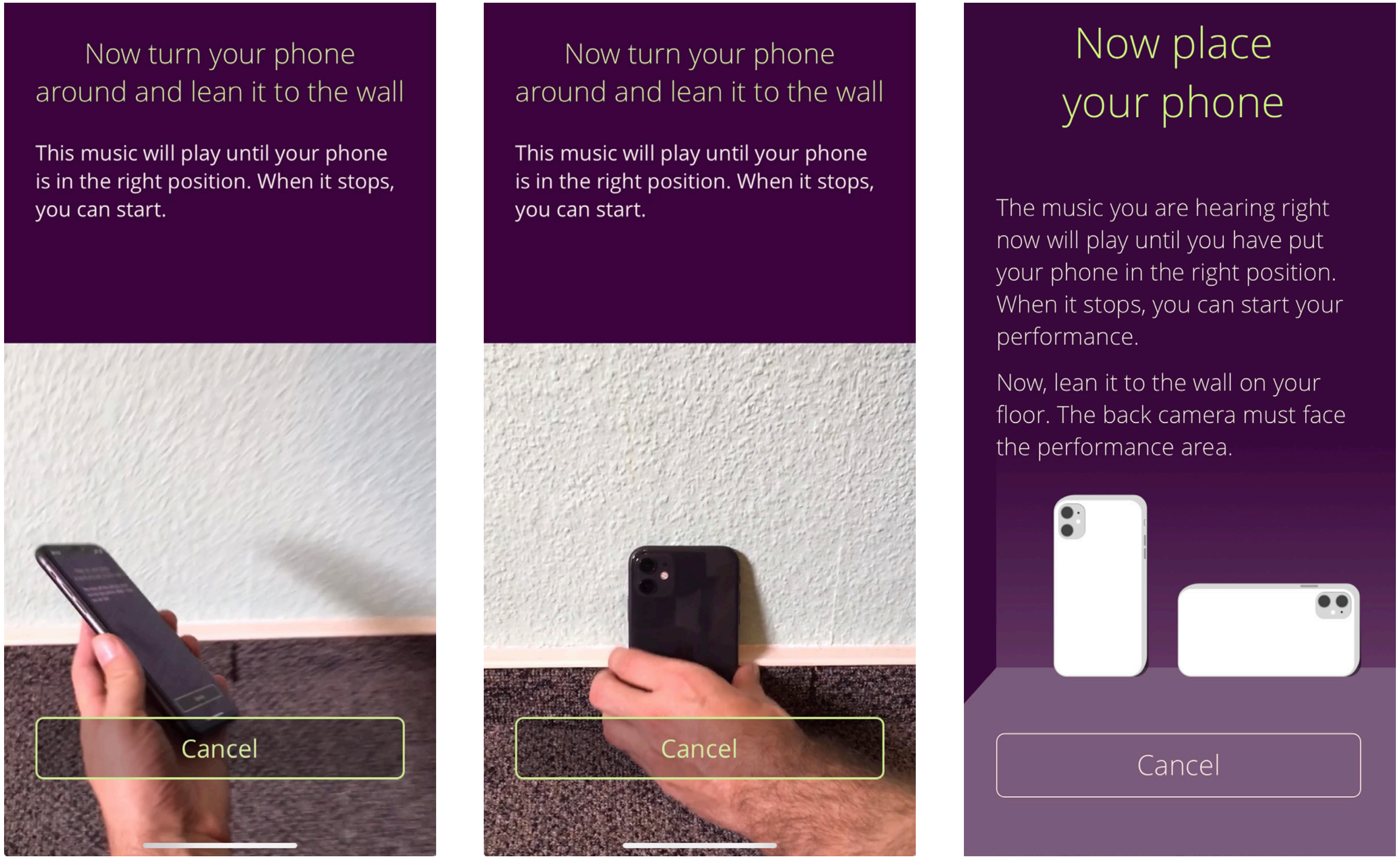


Figure 20. Left, middle: A looping video helps with the phone setup. Right: At first, I used an illustration for this, but testers could not easily understand the information.

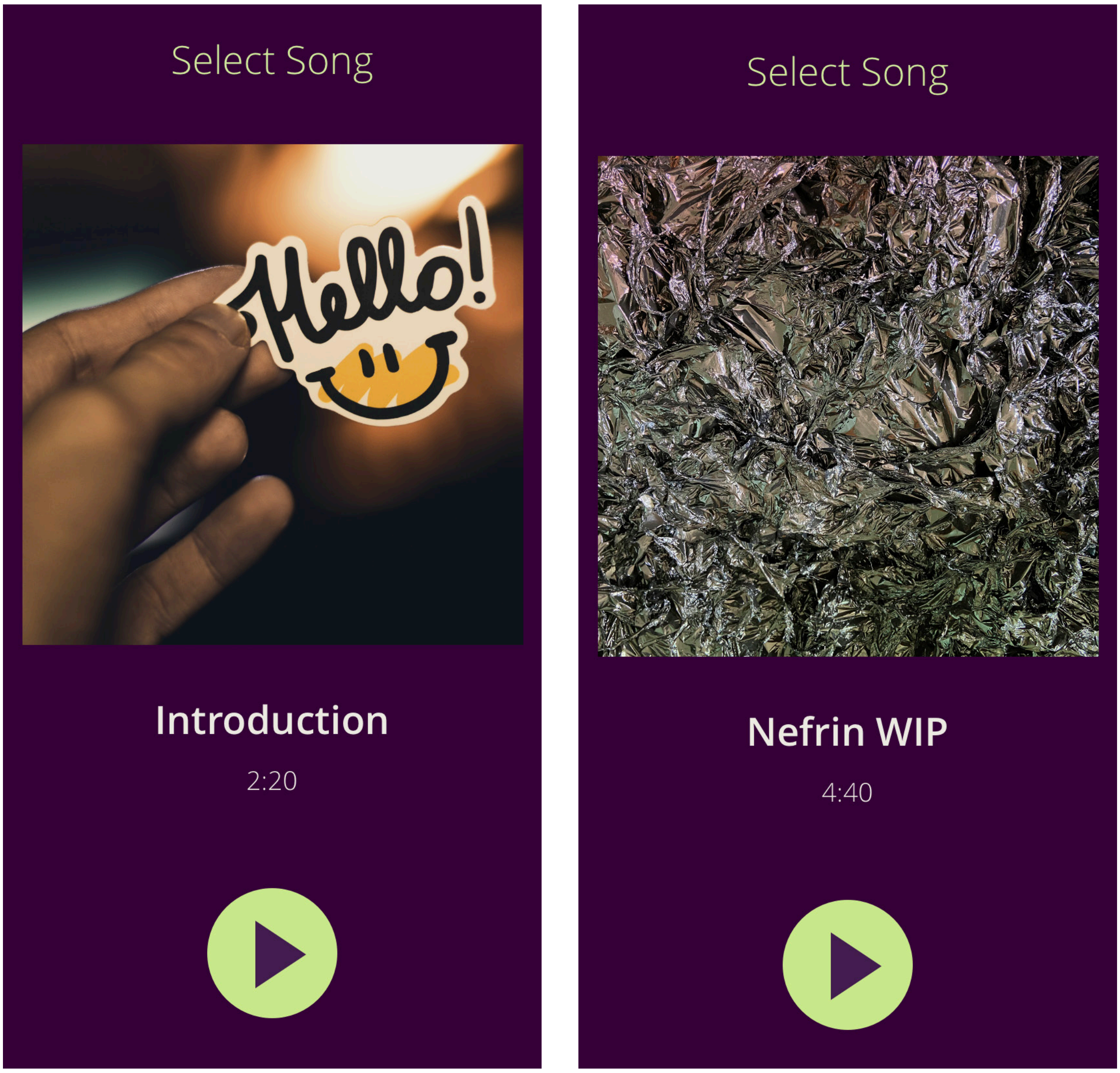


Figure 21. The song selection UI is kept minimal as well.

4.4. Testing

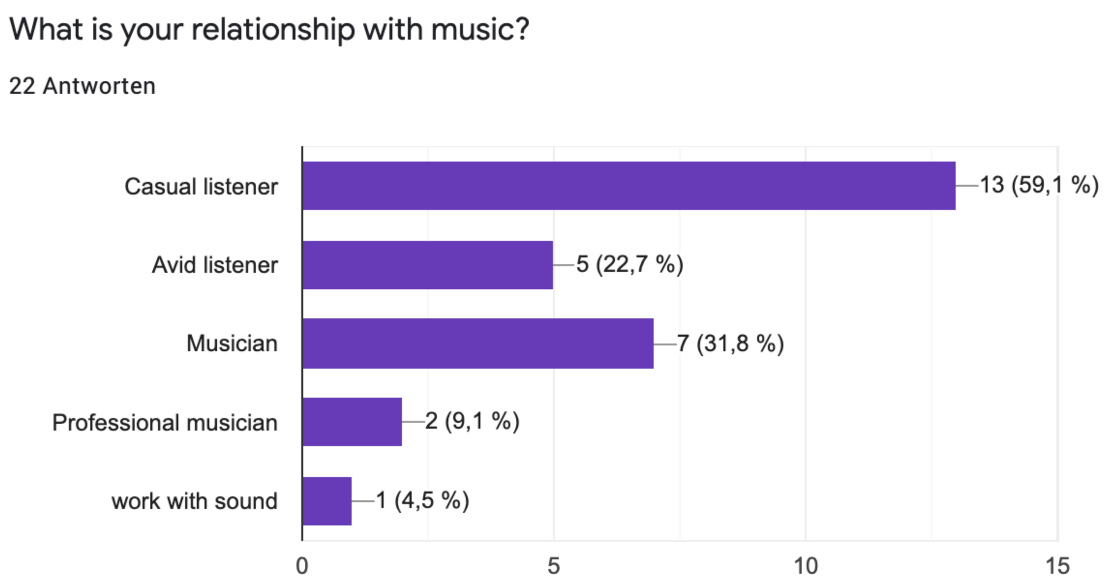
I had more than 20 people test the software during the development process. Many, but not all, were students of HfK. The testing took place remotely as well as during physical meetings. The remote testers received a link where they could download the app via Apple's "Testflight" platform. They could try the app at a time convenient to them and as often as they wanted. There was a button in the interface that led to a feedback questionnaire. Later I started to meet with testers personally. There were several reasons for this: firstly, many of them didn't have a suitable device, so we used my iPhone. Furthermore I had to find out that only a few of the remote testers actually installed the app and also filled out the questionnaire.

When testing in person, I paid attention to the following standardized procedure in order not to influence the testers as much as possible:

1. **Briefly explain the general purpose of the software.**
2. **Open the app and hand them the phone. Observe how they read the instructions and manage the setup process.**
3. **Leave the room so as not to influence the performance by my presence. Testers have the option to record their session.**
4. **After the performance, testers fill out the feedback survey.**

Afterwards there was time for a personal conversation in which further feedback, ideas and thoughts about the project could be expressed. The feedback was mostly positive. On the following pages are the questions and answers in collected (therefore partly shortened) form.

The piece played during testing was not Nefrin but an intermediary composition I had developed during my research and development process. It still mostly represented the described interaction methods.



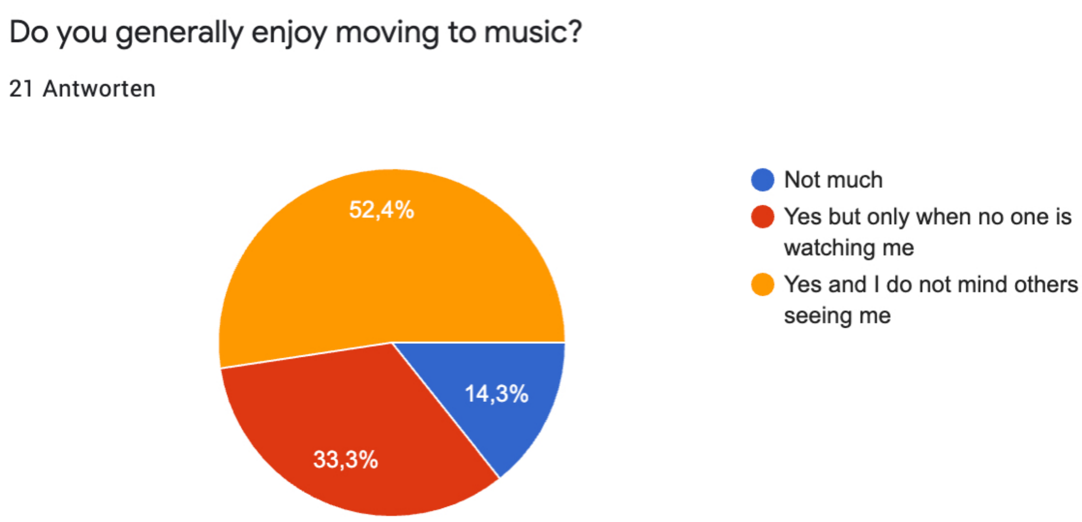
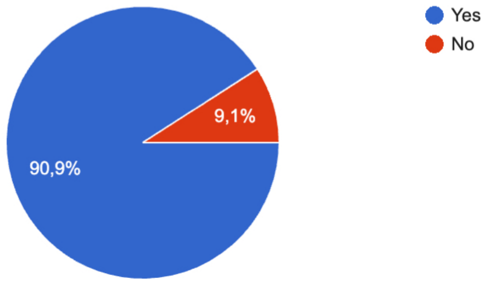


Figure 22. Feeback results (this and the next page).

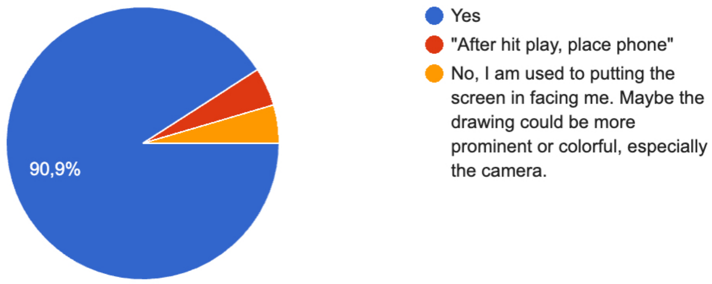
Did you play the song until the end?

22 Antworten



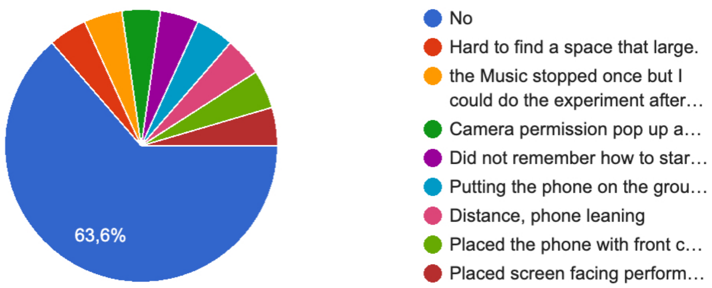
Was the setup process explained well?

22 Antworten



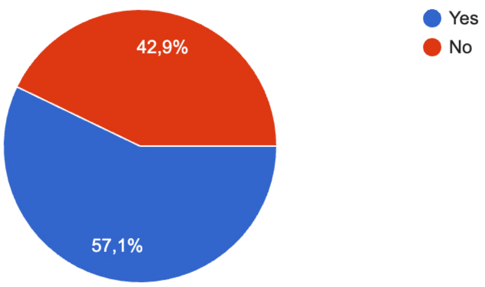
Did you face any problems in the setup process?

22 Antworten



Would you like to watch a capture of your performance afterwards? (As a future feature, your performance has not been captured!)

21 Antworten



Could you recall how you interacted with the app during your performance? (what kind of body movements, use of space...)

21 Antworten

when i stop moving, the music stops

Durch den Iphone Fokus auf mich selbst, hatte ich das Gefühl bzw. wusste ich, dass ich mich nicht großartig entfernen kann, somit bewegte ich mich innerhalb meines eigenen kleinen Kreises.

I danced to the music. I didn't move much space because I didn't know how much camera caught me.

Moved a lot forward and backward, it felt like I could never really control it. Lots of arms movement, also near the ground. Hard and soft movements with the arms, circular movement with my whole body. Stopped every now and then to see what happens then. Every time I thought I "got" it, something else happened.

Standing mostly, but not still, for short time sitting and moving my upper body parts. I moved along the room.

moved arms, hands, leg (works well with kicking!), moved around the space but not much happened

Moving slowly seemed to create an ethereal/atmospheric vibe. Speeding up seemed to introduce a beat and increased the tempo.

used every part of my body. rotated, used the hands a lot. sliding, used the whole space. was quite expressive.

I started slow and moved more and more. Partially I thought the music would go faster but I could not realise changes so easily.

Walking around, raised arms, randomly doing stuff around. Not lifted feet, mainly with hands.

I felt like I have to move. very nice it was very responsive. there were no controls to learn. I just started to dance. the timing felt perfect, the rhythm of my body was connected to the music. But after a time I tried to explore the space. Some things were unexpected: I stopped moving, but then some drums played, then I continued to move.

Figure 23. Feeback results (this and the next three pages).

How did (or not) the musical feedback motivate/inspire your movements?

21 Antworten

somewhat

Ich empfand den Rhythmus als sehr fliessend, somit war ich selbst "im Fluss". Es gab nicht wirklich Pausen, somit wurde ich mit meinen Bewegungen gefordert weiter zu machen.

I was motivated because I could move with the rhythm changing.

Felt like when spreading arms it changes to drumnbass. I think when I moved a lot there was drumnbass, when I was still it was more ambient. Could be more interactive. It was a kind of searching, exploring. Last time it was more playful.

A lot, it was inspiring me to explore my body movements.

Clear relationship with the melody on the hands and bass line on the legs, allows me to decide what to do with the music

I felt my movements had not much of an impact, especially when moving fast. Maybe with more tries I could understand better. At one point I felt that the music controlled me, when it got more upbeat.

I moved more when the music turned louder.

Did stuff to see what happens. Not even the speed. I did things but not because of the music.

the rhythm moved me and I could not really say if I generate it or not. I felt like I have some control but it also pushed me.

Tried with hands first to find out mechanics, then with one leg, there was the bass, then moved sideways and the drums started. Thought that each part of my body was a certain instrument, but then in the last part the drums were playing, which was confusing. At some point in the end there were dissonant sounds which I thought are an indicator that I did something wrong.

Any other thoughts, comments, suggestions?

14 Antworten

Danke, dass ich es ausprobieren durfte.

I liked the information about using the app because it was nicer than before.

Speaks to me, bringing dance into everyday life… Like a mirror.

Just wanted to give feedback quick about the sound playback while phone muted. I always keep my phone on silent and so I heard no music during setup

It was a nice experience, would be really nice to see the performances.

Volume is slightly low on my headphone, could be clearer to the user that purpose of the music at the beginning is used for optimising the volume during the 'performance'.

I could see this being used in dance/yoga studios! Really cool stuff :)

more direct connection of movement and music would be more rewarding.

Just wanted to give feedback quick about the sound playback while phone muted. I always keep my phone on silent and so I heard no music during setup

It was a nice experience, would be really nice to see the performances.

Volume is slightly low on my headphone, could be clearer to the user that purpose of the music at the beginning is used for optimising the volume during the 'performance'.

I could see this being used in dance/yoga studios! Really cool stuff :)

more direct connection of movement and music would be more rewarding.

The actual interaction could be improved. Was not weird to look at the screen.

I dont know how to approach it as a product. it could have many purposes, for dance practice, for entertainment. or as a performance tool. it could have more music and sound options. How to control the interaction with the device. I missed the device at some point.

What do you think about the music?

22 Antworten

comforting

Sehr angenehm und eine gute Begleitung um überhaupt erstmal in die Bewegung zu kommen.

There were various rhythms in the song, so it was interesting to dance. It was a pity that the song suddenly stopped.

Covers different areas/styles, ambient, drumnbass, plays with contrast, expected even more different styles. Could have more bass.

Haven't gotten there yet

It was interesting for me to see how I can change the music rythm, I liked it!

The length of the music is slightly on the short side (or didn't know when it is coming to an end), would have love to have a little longer (30 more seconds for example) to explore around - and development to an ending feeling in the music could perhaps help.

It was generally interesting and harmonious. I wondered if there were any other styles of music/instrumentation.

Enjoyed different themes, ups and downs, felt organic with the movements.

A bit too slow. Non-instrumental.

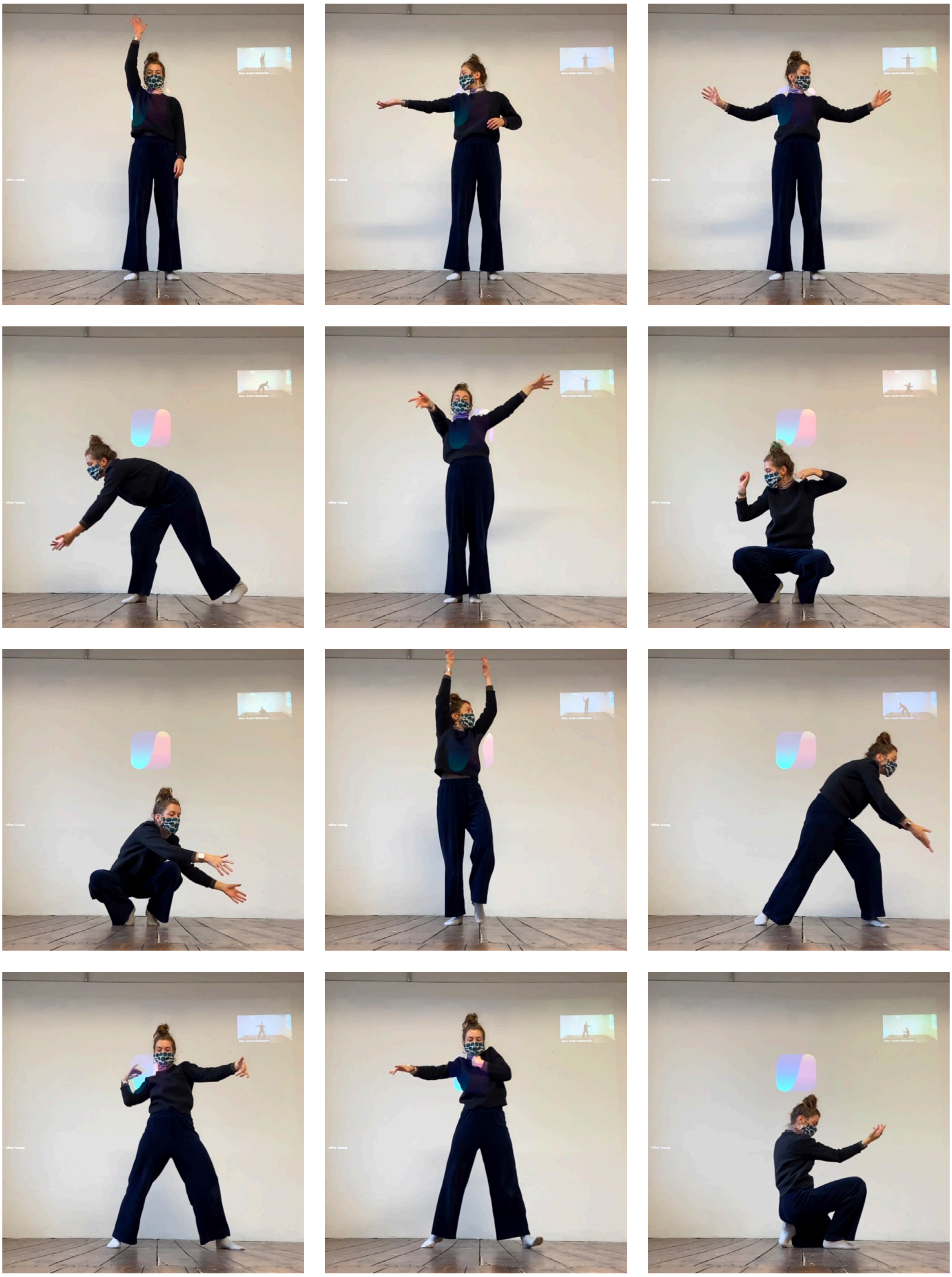
Fine. Sounded good. Not anything for or against it.

liked the sounds and harmony. drums could be more organic. I would like to be part of the music.

Was cool. Not sure if I can influence the tempo. Music should react to sudden acceleration.

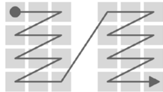
5. PRESENTATION

The work has been presented on November 20th 2020 at the University of the Arts Bremen. The presentation started off with a performance of Nefrin by Sarah Finja Rost, which has been captured in this video. [39]



Presentation

Figure 24. Stills from the performance video, which was captured directly in the app with the built-in recording feature. The original format is 9:16 (portrait orientation), but the images have been cropped from the center for layouting reasons.



6. OUTLOOK AND REFLECTION

6.1. Outlook

It is planned to release the software in the iOS App Store in early 2021. Before the release, at least one more piece should be added which focuses on rhythm instruments. Personally, I do not think that this app can be successful as a product. I consider the app format itself here only as a channel to deliver the experience of the artwork. Because I think that there is not really a market for it as a "product", I plan to release the app for free. At this point it is much more important for me reach as many people as possible than to earn a few cups of coffee's worth through sales. I will promote the app online on the website affinetuning.com and want to post inspiring videos to popular social media sites to attract players.

In the future, I would like to work with musicians on more tracks or features. Before that, some work still has to be done: The current architecture of the software is not ready to be used by outsiders or even uses any of the formats digital composers work with. This means that right now there is no sustainable workflow. Another crucial aspect is the lack of a common language, like a notation for such interactive and dynamic compositions which includes both the movement and the music notation. It will be interesting to research if there are choreography notations which also include music which could be modified for this purpose. So far, I did not look into this because of time constraints.

Figure 25. The app icon is inspired by a wave and the contrast of warm and cold.

6.2. Reflection

One challenge about reflecting on this work is that the experience is individual for everyone. I can only write truthfully about my own experience which is probably very different from everybody else's, because I have lived through the whole development of the software and music. I have my certain way of interacting with the system which in turn shaped how it works. I realized this time and again when I saw other people making drastically different experiences, sometimes "struggling" to make a connection at all. This is of course a result of my decision to design a "free form" system with not many rules to follow, which also results in less guidance. One interesting consequence of this approach are the connections which players try to make retroactively between the music and the movement. This is one of aspects which was most interesting for David Rokeby. One player, who was an amateur dancer, told me that she found it rather difficult and even exhausting because she could not just move *to* the music, and felt that she should be alert the whole time. It basically meant more work for her. In fact, by now I am quite sure that this system is something which offers a developing experience and different levels of reward over time. But I am also aware that most people won't have the patience, motivation or simply interest for this process. Some examples: Initially, most people stay in the same spot the whole time, standing up straight, oriented towards the device. But some musical effects rely on the player to rotate, go low, or raise their feet. Also, utilizing starting, stopping, and the speed of movement as an interaction method also takes time to discover – not because it is difficult, but requires a certain mindfulness.

From a composers perspective I can only share my own experience so far, which is probably not a very relevant insight because I had almost no prior experience in music composition. The process is also very tightly coupled to the interaction design, thus I usually considered both aspects at the same time. One thing I noticed is that movements are often "quicker" than musical structures. One example: I wanted to couple the virtual room size of a reverb effect to the size of the bounding box of the players current pose. But people do not stay in a very "large" pose for more than a few split seconds, a much shorter time than the effect of reverberation needs. I ended up to use not the bounding boxes of the motion paths instead. As a result, this feedback feels much more indirect. I expect that most players will not see a connection between movement actions that occurred more than a few seconds ago and musical structures based on them. One thing which I found exciting — and in this case I think this holds true for "real" musicians as well — was that the interplay of sound, especially when using effects like a bitcrusher distortion, was often not predicteable; and this does not even include the player's interpretation of the concrete sonic result.

Advantages and disadvantages of the platform

The app has not been released at the time of writing, so I cannot evaluate feedback from a large number of users. However, during the testing phase, I noticed that users have certain expectations of an “app” that differ greatly from, say, an installation in an exhibition. An app is perceived as a product, and serves a clearly identifiable purpose. Both interface and feedback are expected to be clearly recognizable/operable and unambiguous. Users already know before making an input what its purpose is and what result is to be expected. If this result does not correspond to the expectation, this is seen as an error in the software. The gaming sector is less strict, but the principle also applies here on the whole. Another common user expectation ran counter to my intention and the technical necessity of using the rear-facing camera: looking at and interacting with a screen. I took it for granted after a short while not to use visual feedback, but for many testers this became a real problem when positioning the device: they almost always tried to point the screen towards the room.

I had not realized another drawback until I started looking for testers: The required devices are still comparatively expensive. The cheapest compatible new iPhone costs 467 € in Germany at the time of writing. There were many people interested in testing, but they didn’t have a suitable iPhone. I had virtually traded the exclusivity of the exhibition format for another, mainly monetary threshold. In this respect, I can only hope that the technology will become more affordable over time (which is to be expected).

Technical Constraints

One challenge which consumed a lot of time was a certain unreliability in the pose detection. Time and again I tried to find algorithmic solutions especially to detect motion onsets, but they were never reliable and consistent enough. I suspect that this is also just a consequence of me not having a computer science or signal processing background and the according experience and knowledge of suitable algorithms. But even with that I am sure that some problems would not have been solved. For now, this shortcoming is something to be accepted; I must live with the noise. I tried to react to this artisticly in the last part of my composition, when noise and distortion are an essential part of the music, thinking of the interesting sonic results as a glimmer of hope.

A feature which was prevented by the tracking quality was stereo sound which changes its balance based on the players’ rotation. The result was more confusing than anything else because over time the rotation data would drift. Another technical limitation which influenced my concept early on is that the pose detection only works for one person at a time. It would have been very interesting to include interaction between two or more players in the same physical space.

Apart from that, I identified two other technical constraints. One is audio performance when it comes to realtime synthesis and effects. Like in graphics, I imagine one develops an intuition for the performance ceiling, but I was lacking the experience needed for this. Unfortunately, the penalty here is much more punitive than with graphics because it results not just in a few graphical glitches or framerate stutter, but can be very unpleasant and even harmful for the ears and the equipment. Also timing problems can add up which can stretch envelopes, which in turn leads to more audio playing at the same time, effectively destroying the composition.

Another constraint which seemed rather minor at first was the fact that the pose detection only works with the back-camera. It turned out to be a real hindrance for usability. It would be much easier for users if they could see some kind of visual feedback at least for positioning themselves in the camera image. I still think that the approach of concentrating on sonic feedback is valid, but I also want the threshold of entry as low as possible. I tried to mitigate this problem with voice feedback in the setup process but it often happened that users would not comprehend this the first time and be lost.

Process

During the process I faced a number of challenges. Each on their own were not major setbacks, but, in hindsight, I was a bit oblivious to the fact that the topic of my thesis had both quite some breadth and depth. This led to some non-ideal workflows, fragmented periods of working in different areas (i.e. jumping around quite a bit) and a convoluted schedule. Ideally, this work would have been done by a team of specialists. And as everyone else during this time, I was also affected by the COVID-19 pandemic. Early on I planned to work with a choreographer from the Theater Bremen. But just before starting the collaboration, the lockdown happened, and shortly after, he fell sick. And it did not help that my partner was stuck abroad two months because of flight restrictions. It was also planned to show a version of the work at the conference "Kunst der Erfahrung" ("Art of Experience") at the university Witten, but this event had been cancelled. After the collaboration had not happened, I worked on my own, only reaching out to testers a few months into the process. In hindsight I think the project would have benefited from earlier, continuous testing.

I realize that parts of this last chapter read a bit bleak. After all, it was a long and exhausting process and pretty isolated at times. But I still had a lot of joyful and stimulating moments as well. There were many instances when interacting with sound through my body simply felt magical. Maybe the most peculiar and interesting thing was when I got aware of the fact that I had created this weird system for me, in the first place. I have interacted with it not only in the short term feedback loops of my performances, but also at the larger timescale of development. In a way, this is true for many kinds of artworks, but the quality of the relationship in this case was pretty remarkable. I distinctively remember one time, when lagging feedback and unreliable detection actually evoked a sensation of weight and stiffness. Even though this feeling was primarily a negative one, I found the effect pretty profound, especially because there were no visuals involved, only sound.

Writing this, I am excited again to continue with this work.

7. BIBLIOGRAPHY

All URLs have been last visited at January 8th 2021.

1 Krueger, Myron W. 1977. Responsive environments. In: Proceedings of the June 13-16, 1977, national computer conference. AFIPS '77. Dallas, Texas:ACM

2 Universal Everything: *Super You* AR app: https://universaleverything.com/projects/super-you

3 *Very Nervous System* video: https://vimeo.com/8120954

4 Entry of *Very Nervous System* on Medienkunstnetz: http://www.medienkunstnetz.de/werke/very-nervous-system/

5 David Rokeby. The Harmonics of Interaction. MUSICWORKS 46: Sound and Movement. 1990. Online article: http://www.davidrokeby.com/harm.html

6 *softVNS* Max/MSP modules by David Rokeby: http://davidrokeby.com/softVNS.html

7 *Just Dance* gameplay video: https://www.youtube.com/watch?v=hJ6xS1pDeo8

8 https://www.ubisoft.com/de-de/game/just-dance-2020/

9 *Nagual Dance* video: https://www.youtube.com/watch?v=UMKwKagpxLA

10 Nagual Sounds website: http://www.nagualsounds.com/

11 Introducing Motion Composer: https://vimeo.com/158513377

12 Vogt, Katharina, Pirrò, David, Kobenz, Ingo, Höldrich, Robert, Eckel, Gerhard. 2009. "PhysioSonic - Evaluated move- ment sonification as auditory feedback in physiotherapy." In: Auditory Display. CMMR 2009, ICAD 2009. Lecture Notes in Computer Science, vol 5954. Springer, Berlin, Heidelberg

13 *Motion Composer V3* leaflet: http://motioncomposer.de/wp-content/uploads/2020/08/ENG_MC3.0_2020_4.pdf

14 Official *Posenet* repository: https://github.com/tensorflow/tfjs-models/tree/master/posenet

15 Microsoft Azure Kinect product page: https://azure.microsoft.com/en-us/services/kinect-dk/

16 Dushyant Mehta, Oleksandr Sotnychenko, Franziska Mueller, Weipeng Xu, Mohamed Elgharib, Pascal Fua, Hans-peter Seidel, Helge Rhodin, Gerard Ponsmoll, Christian Theobalt. XNect: Real-time Multi-Person 3D Motion Capture with a Single RGB Camera. 2020. ACM Trans. Graph., Vol. 39, No. 4, Article 1.
Project website: https://gvv.mpi-inf.mpg.de/projects/XNect/

17 Apple. Bringing People Into AR. 2018.
Video: https://developer.apple.com/videos/play/wwdc2019/607/

18 Video documentation of iOS ARKit Noise and Errors: https://www.youtube.com/watch?v=LRADlzGPNFU

19 Rudolf von Laban. The Language of Movement - A Guidebook to Choreutics. 1974. Plays Incorporated

20 Rett, Jörg, Dias, Jorge. 2007. "Human Robot Interaction Based on Beyesian Analysis of Human Movements." Coimbra, Portugal. Conference Paper. DOI: 10.1007/978-3-540-77002-2_45

21 He, Jing. 2014. "Paint with Dance: An exploration in creating abstract animation with movement." Master Thesis. Parsons School of Design. New York

22 Antonio Camurri, Gualtiero Volpe, Stefano Piana, Maurizio Mancini, Radoslaw Niewiadomski, Nicola Ferrari, Corrado Canepa. The Dancer in the Eye: Towards a Multi-Layered Computational Framework of Qualities in Movement. MOCO '16: Proceedings of the 3rd International Symposium on Movement and ComputingJuly 2016 Article No.: 6 Pages 1–7. https://doi.org/10.1145/2948910.2948927

23 ARSkeleton demo repository by user "iamfine": https://github.com/iamfine/ARSkeleton

24 My visualisation tool in use: https://vimeo.com/375224596
25 openFrameworks project website: https://openframeworks.cc/
26 AudioKit project website: https://audiokit.io/
27 Curtis Roads. Composing Electronic Music: A New Aesthetic. 2015. Oxford University Press, Madison, New York.
28 Aniruddh D. Patel, John R. Iversen. The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. Frontiers in Systems Neuroscience, Volume 8, Article 57. 2014 DOI: 10.3389/fnsys.2014.00057
29 Douglas Cooper. Very Nervous System. Wired Magazine, 1995.
Online version: https://www.wired.com/1995/03/rokeby/
30 LFO mapping: https://vimeo.com/471993445
31 Macintosh Plus: https://www.youtube.com/watch?v=aQkPcPqTq4M
32 *Your Move* timepitch: https://vimeo.com/471990901
33 *Halo* mappings: https://vimeo.com/471991521
34 Luke Dahl. Triggering Sounds From Discrete Air Gestures: What Movement Feature Has the Best Timing?. NIME'14, June 30 – July 03, 2014, Goldsmiths, University of London, UK.
35 Charles E. Clauser, John T. McConville, J. W. Young. Weight, Volume, And Center of Mass of Segments of the Human Body. Aerospace Medical Research Labs. 1969. https://ntrs.nasa.gov/citations/19700027497
36 Marc-André Weibezahn. Structuring Multi-Layered Musical Feedback for Digital Bodily Interaction: Two Approaches to Multi-layered Interactive Musical Feedback Systems. ICMI '20 Companion: Companion Publication of the 2020 International Conference on Multimodal Interaction. Pages 446 - 448. October 2020. https://doi.org/10.1145/3395035.3425970
37 My interpretation of *Pachelbel's Canon*: https://vimeo.com/479520369
38 Affine Tuning setup video loop: https://www.youtube.com/watch?v=5JbYB-n4ACE
39 Colloquium performance of *Nefrin*: https://youtu.be/d93fXQ9IOns
40 Sonic orb experiment: https://vimeo.com/471992132