LICHTGESTALT: INTERACTION WITH SOUND THROUGH SWARMS OF LIGHT RAYS

Jonas Fehr  
Aalborg University Copenhagen  
jfehr13@student.aau.dk

Cumhur Erkut  
Aalborg University Copenhagen  
cer@create.aau.dk

ABSTRACT

We present a new interactive sound installation to be explored by movement. Specifically, movement qualities extracted from the motion tracking data excite a dynamical system (a synthetic flock of agents), which responds to the movement qualities and indirectly controls the visual and sonic feedback of the interface. In other words, the relationship between gesture and sound are mediated by synthetic swarms of light rays. Sonic interaction design of the system uses density as a design dimension, and maps the swarm parameters to sound synthesis parameters. Three swarm behaviors and three sound models are implemented, and evaluation suggests that the general approach is promising and the system has potential to engage the user.

1. INTRODUCTION

This paper presents the design and development of LichtGestalt: an interactive sound installation that is explored by movement. The movement aspects of the installation has been reported in [1], here our focus is on the mapping between movement and sound, sound synthesis, and dynamic generation of control.

Previously, direct manipulation of three tangible interfaces equipped with sensors was controlling the sounds and colors of the installation [2]. With collaboration in mind, the installation was designed as three tangible interfaces mounted on “branches” coming from a central trunk (Figure 1).

Inspired by works like Polymetros [3], the focus was on a minimal musical expression. Each interface was equipped to track its position in a two-dimensional space and was mapped to an individual voice allowing manipulation of timbre and the tempo of a tremolo effect. When all three interfaces were active, the audience could play a drone, with polyrhythmic qualities changing according the tempo of the tremolo. An additional sound quality, a high pitch crackle, was controlled through parameters based on the sum of all three interfaces. To provide transparency between movement and sound, visual feedback based on multicolour LEDs was designed representing both sound qualities in combination with the tracked position.

An evaluation of SonoFluo indicated issues on control. In most cases the participants did understand what was controlled and what originally was planed as musical collaboration by movement became a collaborative exploration of how the system could work. Franinovic and Salter attribute this to designers’ “... almost formulaic understanding of interaction as a series of input-output processes ... This assumes an already fixed set of relations among the user/interactor, the object/instrument/sound-making body, and the environment in which the interaction with sound takes place.” [4]. To vary these relations, in LichtGestalt we have implemented an indirection between motion and sound through a dynamic, intangible interface, movement qualities (MQ), and a virtual ecosystem (Figure 2).

2. BACKGROUND

2.1 Gesture and Sound

One approach in designing new musical interfaces is the use of gestures. With the help of machine learning, interactive systems are able to learn and recognise gestures, which are subsequently used to control sound [5].

Gesture recognition focuses on specific movements, like drawing a circle into the air, employing a variety of different algorithms. The biggest algorithmic trade off is the usability for real-time applications. For many cases a gesture is first completely recognised when finished. To be used in musical performances it is important to get a feedback as instantaneous as possible. It is also interesting to estimate the variance to a reference gesture in order to design the feedback. Many systems work with a kind of likelihood feedback, reporting the probability to be a certain movement. Some systems also report the phase of the gesture, allowing for example to control the progress of a played sound file. A good overview of different techniques is provided by [5].

Machine learning can also be used to classify different intentions of a movement [6]. Instead of focusing on a complete gesture, the quality of the movement can be extracted. The Laban Movement Analysis (LMA), after the dance theorist and analyst Rudolf Laban, is based on four main qualities: Body, Effort, Shape, and Space. Mentis and Johansson provide a good summary on the LMA, especially on the effort and present an approach using movement analysis to control sound [7].
A depth camera is used in Mentis’ and Johansson’s research [7] to estimate and classify which of the basic elements was most present and play a 15-second sequence especially for the element composed music. The system was evaluated with a LMA specialist and showed an accuracy of 67%. The audience, however, found the classification of movements in many cases not comprehensible and focused on other qualities, like for example the energy put into the movement. The authors consider the use of LMA as an element as too high a level of classification for an untrained observer.

Physical modeling, as a sound synthesis method, has an important expressive potential [8], [9]. Alaoui et al, as an indirectness, couple movement quality parameters to physical models (in their case mass spring systems) to create both visual and aural feedback for a dancer. This is of special interest, as physical models are in combination with a reasoning system a good example of dynamical systems that facilitate plausible indirection between sound and movement. Another example is an agent-based system, which is described next.

2.2 Sound and Agents

A software agent is an autonomous entity that observes and acts upon an environment and directs its activity towards achieving goals. Agents, for instance, enhance the musical performance possibilities: different approaches where some of the musical parameters are controlled by agents in parallel to the user have been described in [10]. In their implementation, agents acting in the same space manipulated virtual instruments based on physical models. In different scenarios, agents were influencing the outcome of an action with reaction, for example damping a string after it played or changing the string’s properties while playing.

With the use of agents as procedures, the instrument takes on a life of its own and enhances the possibilities of the player. The installation Room #81 is a good example, where an agent is used to create both a soundscape and light changes to frame the user’s interaction [11]. In different setting agents are guided by the user to produce different sounds when interacting with other objects in the game. In this case, the agent is not just a trigger/control parameter but also visually represented on the screen.

Shacher and his colleagues recently suggested a classification of fundamental mapping relationships with the help of swarm simulations [12]. These include a number of strategies that relate to musical practice, highlighting the role of swarm simulations in mapping, especially in making the mapping relationships less predictable and more organic. Other studies on agents and sound include [13] and [14].

3. CONCEPT

In our installation, we consider an audio-visual composition as a complex virtual ecosystem embedded in its enactive landscape (see Figure 2). The user as part of the enactive landscape is able to directly manipulate this ecosystem.

The ecosystem is based on a flocking algorithm by Reynolds [15], which is represented in space through light-rays thrown by a projector. When the user enters the space of the ecosystem, he becomes a part of it, and the boids react towards him as a physical object. In parallel, the movement quality (MQ) is estimated in order to change the flocking behaviour and therewith the way the user interacts with it. The details of the MQ estimation have been presented in [1]. In parallel the position and
contour of the hands are translated into the virtual space of the flock in order to provide direct interaction possibilities with the boids.

Currently, the reaction of boids is considered as one of the three main conditions: hunted, passive and curious: hunted is a basic reaction towards fast and directed movements of the user. The boids start to organise in swarms, and seek distance to the user. Passive: when nothing of importance happens, the boids act independent (similar to a heard of wildebeests grazing). They move slowly, and seek distance to each other. The user has to move randomly and not too fast. Curious: when moving very slowly, the boids become interested in the user and come closer (similar to insects seeking a light source).

To create the aural feedback, the position and acceleration of each boid is forwarded to a Max/Msp patch using OSC. After updating the flocking system, the boids are rendered and projected, which closes the circle (see Figure 2). Currently, all simulation and visual computing is done in C++ on openFrameworks\(^1\), including the add-on openCV. The sound is generated in the environment of Max/MSP\(^2\).

4. SONIC INTERACTION

In this section, we describe how we tackle the sonic interaction design of LichtGestalt. Our project is artistically driven and we do not rely on generative design approaches. For instance, we did not conduct design workshops or participatory design sessions. Yet, we find the action/sound relationships obtained as sonic interaction models in [16] relevant for our work. Especially, their conducting model that introduces the symbolic and semantic meaning of gestural interaction with sound becomes important when interacting with swarms in LichtGestalt. Conducting, with the help of dynamic control structures becomes a way to address the need of different levels between control and sound synthesis [17], [18].

4.1 Design Constraints

The user should easily understand how to interact with the virtual ecosystem. For simplicity, here we consider a case where the LichtGestalt is placed in a room akin to a large digital musical instrument, such as the ReacTable [19]. This is to shorten the process of audience interactions with generic artworks [20]. Because of the intangibility of the light rays, however, the generated sound in LichtGestalt should be very directly coupled to the visual representation. The user should get an idea when a sound occurs and what is generating it.

As a first design, we have considered three simplistic sound sources to experiment with the dimensionality of mapping, as suggested in [12]. Currently, not all of their parameters are used. Furthermore, they are all monophonic, in order to simplify the sound synthesis and explore the action/sound mapping. In later stages of design, we will also consider sound spatialization, as the position is the most common property in swarm visualization and auralization [13].

4.2 Density as a Design Dimension

Density is the degree to which something is filled, crowded or occupied. It is described as the quantity of something per unit measure. In our design, we consider the concept of density as a rich source of possibilities to combine the visual and the corresponding aural feedback. The visual cues regarding to the density are:

- the amount of boids per area, visible as bunches of light-rays,
- the closeness of boids in a swarm, measures as the amount of boids per area of the swarm, visible as the intensity of the bunches, and
- the amount of times the boids hit another object, visible as direction change in the movement.

These visual cues can be related to the sonic cues based on the following properties:

- **temporal**: the amount of events per time, ranging from continuous to single events, from drumroll to single hits, spectral: the spread of the sound in the spectrum, and layers: a music is considered as more dense, when multiple instruments play simultaneously than when just a single does. This property is close to the spectral density, but can also be considered as multiple melody lines like for example in polyphonic music.

4.3 From Swarms to Sound

The tree conditions of swarm behaviour we have outlined in Sec. 3 serve as basic starting point in sound design. While hunted was hinting towards a climax, the other two conditions are considered as calmer soundscapes. In the hunting part, the boids organise in flocks. A computer vision based evaluation was programmed estimate the flock formation, together and with their density and group velocity.

Additionally, a temporal approach was implemented by the detection of collisions. Collisions happen from time to time and by distinction of different actors (boid/boid; boid/wall; boid/user) different sound qualities can be mapped towards them.

With the change of the condition either of the sound generations are addressed differently. This allows keeping the sound generation in general simplistic and static; only through the change of the general condition the user’s focus switches towards the different feedbacks.

4.4 Sound Synthesis

So far, we have discussed the density coupling between three swarm behaviours and three sound sources. The first of these sound sources correspond to the swarms and implemented as a classical simple two-oscillator FM Synthesizer [21]. The other two correspond to collisions and

\(^{1}\) http://openframeworks.cc
\(^{2}\) https://cycling74.com/products/max/
implemented by physical models from the PeRColate library in Max/MSP: user/boid collisions trigger a marimba tone, whereas other collisions trigger a bowed bar tone. Specifically, marimba is triggered at a high pitch, by varying the pitch randomly over a half octave and the stick’s hardness and position. In this way the created sound is always minimally varied. When multiple boids collide with the hands in very short time intervals, a ringing sound texture is created. The bowedbar, on the other hand responds to the collisions with the wall and creates a warm and deep sound, again with a random pitch in the range of three semitones. We next describe the continuous control of the swarms.

4.4.1 FM Synthesis of Swarms
Our system detects each group of boids as an individual flock, for each flock we use the same synthesis and mapping, so each flock has its own voice. For each swarm the system detects its velocity, its area, and the number of boids in it. By dividing the area through the number of boids, a measure of density is calculated and used to vary the amplitude of the modulator. The velocity is directly mapped to the frequency of the carrier. The loudness is calculated from the area covered by the swarm. The ratio between the modulator and the carrier is chosen non-integer (2.04) as it creates an interesting, gnarly sound. The movement activity of the swarms thus reflects in the pitch change; if the flock moves slowly the pitch stabilizes. In states where no flocks are formed, no sound occurs. Instable conditions where flocks are formed for very brief moments create transient sounds. The creation of an individual voice per flock addresses the layered density, while the pitch change is connected to the spectral density.

5. EVALUATION
A test session was conducted with 11 subjects, in a room at the university over a time period of 3 days. As the installation is developed for a museum setting, this is not ideal. However, to find out how an individual subject reacts towards the installation and how she personally perceives it, the isolated setting seemed reasonable as it provides a controlled setting and better observation possibilities of the procedure. During this test session, the subjects were asked to investigate the installation and to fill out a questionnaire afterwards. In order to determine how consciously the subjects could control the installation, the subjects were asked to reproduce specific reactions. The gathered data was qualitatively evaluated. The three stages of the test procedure are described below.

5.1 Investigation of the installation
The subjects entered the room and were asked to explore the installation. They were left to explore freely, until they stopped by themselves or seemed to run out of ideas, repeating the same movements. The subjects were observed in terms of where they looked and which reactions the installation produced upon their actions. The observations were captured with a standardised form.

5.2 Questionnaire about the experience
After the investigation the test subjects were asked to fill out a questionnaire. The questions were designed towards following topics:
- Personal questions, to find out if the subject has any knowledge and background relevant to the evaluation.
- A question about how the different reactions were perceived, assuming the subjects recognise different reactions.
- Questions towards the sound and the connection towards to the light rays, and how the touch of the light rays was experienced.
- A question about the control experience; what could be controlled, and if they have an idea of how the system works.
- Through a question about the associations, the focus was shifted towards the aesthetics.
- Questions regarding the installation’s sound aesthetics and visual appearance, their personal engagement and general impression.

Seven of the questions, especially those regarding aesthetics and interests, were posed with a Likert psychometric scale model, to provide a common measure. The subject was asked to rate within a range of 1 to 10; 1 standing for not at all, while 10 meant quite a lot/very. The liker-scale answers were followed by a text-field with the request to comment.

5.3 Performance session
The performance test was an experimental approach to see if the subjects were able to reproduce requested reactions. This was attempted, due to the fact that the vocabulary, which describes one’s experience, is often difficult to find. With the performance session, a specific reaction could be addressed and the result directly be observed.

The requested tasks were designed after specific characteristics of the three conditions, hunted, passive and curious. They were as follows:
- Can you collect the light rays? There were two solutions to solve this task: With slow movement, the passive condition would allow to gather them in a corner, or when keeping the hand still, the curious condition would make the boids become attracted. Can you move the light rays around? This task can be seen as a follow up to the previous and if they are able to play with the conditions given. When in passive mode, they already had to move conscious. When using the curious condition, the hand could be moved very slowly and the boids would follow.
- Can you make the installation create the high pitch sound texture? The high pitch sound texture is just created in the hunted condition. If they apply the curious condition the drone had a lower pitch, so they were additionally asked if they could modulate the pitch.
- Can you create the high pitch bell like sound? This was an attempt to see if the user would understand that the collisions with the light rays are causing the sound.
- *Do you have any idea when the lower pitch percussive sound occurs?* The last is not really a task, as it does not expect any action of the user. The question was posed to see if the test subjects understood the connection between collisions with the edges and the therewith trigger sound.

### 6. RESULTS

Seven of the eleven participants reported some kind of technical background or have had some previous experience with interactive installations. Their age was widely spanned from 21 to 56 with an average age of 36 years. Three of the participants have heard about the installation before, while just one of them had some closer knowledge about the system. However, their results show no significant difference to the other participants. Their data are therefore evaluated equally with the others. The approximate duration of each session was about 30 minutes, most of the time taken by the questionnaire.

All qualitative data obtained has been coded by the first author, and complemented with his observations during the investigation and performance. The quantitative results are meant to complement the coding, by themselves they not reliable due to small sample size; nor it was the aim to gather statistics for interaction with an experiential installation like the LichtGestalt. We nevertheless present them in the sequel, as they are easier to summarize and draw conclusions compared to a qualitative data and observations. The full data including codes and implications on the design loop, and user associations such as “Very romantic; Hot summer night and bugs flying. Quiet space in a not polluted nature by water” or “I tried to play a wizard, waving his hands to shoot out fireballs...” can be assessed from the master thesis of the first author [1].

#### 6.1 Aesthetics

Under the topic aesthetics, the answers from the questions regarding the sound, the visual appearance and their interplay are evaluated (see Fig. 3). The visual appearance was rated rather positively; the reasons behind the ratings varied considerably. Some mentioned the subtle and simplistic design as very pleasing, while others just call it nice. Negative points are regarding the projection on the ceiling and other possible improvements.

The interplay between the sounds and the visual appearance, on the other hand, may show how interaction fidelity may modulate even crude sound designs.

#### 6.2 Experience

The answers regarding the experience and engagement were quite positive, as illustrated on Fig. 4. Many participants mentioned the different reactions to be explored as the main motivation, while others found the abstract design nice to give space for imagination. Also the sensation of manipulating sound is mentioned several times. The question about a possible recommendation to friends is borrowed from [3]: it gives a different angle about the experience.

![Figure 4. Experience evaluation results.](image)

#### 6.3 Control

The answers to the control question were positive but not consistent (Fig. 5). Control, as an abstract measure, seems subjectively perceived. When asked for elaboration, some of the subjects report less control due to the semiautomatic behaviour of the boids, while others seem to master the dynamics of the interaction with high ratings.

![Figure 5. Control evaluation results.](image)

### 7. CONCLUSIONS AND FUTURE WORK

We presented the LichtGestalt: a new interactive sound installation to be explored by movement via a dynamical system (a synthetic flock of agents). In the installation, the relationship between gesture and sound are therefore mediated by synthetic swarms of light rays. We have presented the sonic interaction design of the installation, which uses density as a design dimension and maps the swarm parameters to the sound synthesis parameters.

The system as a running installation has been tested on 11 subjects at the university. A qualitative evaluation of the results, which was not discussed here in detail, has indicated that the general approach is promising. The evaluation highlights also the potential of the system to engage the user, as it seems pleasing to explore and discover different dynamic reactions.

However, the evaluation has also disclosed some potential improvements: the current swarm behaviors and sound models can be extended based on the presented groundwork in the future. Also the swarms can be spatialized, and more elaborated sound synthesis and sonic in-
teraction design [22] can be incorporated in LichtGestalt installation.

8. REFERENCES


