FLOX: A MULTI-MODAL APPROACH TO GROUP-EMBODIED INTERACTIVE SOUND ART

Jason Ponce

University of California San Diego
Department of Music, La Jolla, CA
`jbponce@ucsd.edu`

ABSTRACT

This article presents the interactive sound installation *Flox* as an example of a group-articulated approach to designing interactive experiences. The technology and methods used to create *Flox* are described, as are the implications of using such technology to create more enactive experiences for visitors.

1. INTRODUCTION

Flox is an interactive, multi-channel sound installation built in response to a scenario commonly observed in interactive art, in which a viewer or participant interacts with a piece in a way that emphasizes the individuality of the experience and positions the experience of the work somewhere between the participant and the work itself. While the technologies that enable interactive experiences have advanced, the benefits of this progress are typically applied towards generating a richer experience for the participant, or affording the individual more points of contact with the work. One of the classic examples of this kind of thinking is David Rokeby's *Very Nervous System*, in which the artist tracks his entire body with a video camera and uses the extracted data to articulate a piece of electronic music. Furthering this model, then, involves discovering new ways of extracting increasingly fine-grained details from a data stream. Imagine a virtual instrument that responds to every twitch and tiny gesture a performer might make. The question *Flox* poses is not whether the pursuit of such data is valuable, but rather what may be lost when the experience of an artwork is forced so thoroughly into a binary relationship.

Flox is one piece in an ongoing larger project to develop works that extend the possible artistic relationships between interactive bodies. Instead of organizing an experience solely for the individual, *Flox* distributes its experience outwards, to the group, and is realized ultimately by actions coordinated between multiple interacting bodies. It does this by presenting visitors (four at a time) with an interactive soundscape made of sonic components that their individual bodies control directly, and by giving the group the task of reorganizing a jumbled soundscape into a "resolved" state. Accomplishing this requires each person to first discover what specific aspect of the soundscape they are in control of via a free exploration of the system's various mappings of movement to sound, and then to coordinate with the group to bring everybody's part into stasis. The word "stasis" is used here to represent any number of prearranged goal states that the system might seek in a given instance. Examples of such resolution might include: rhythmic unison, harmonic convergence, active or inactive spatialization, the intelligibility of text passages, etc. Exploring the system involves physically moving a illuminated sphere, or pod, around the room, seeking the specific location in the space, or "home", where that individual person's sonic component resolves. One found, the individual is free to stand quietly and wait for the others to find their own homes, to actively assist others, or to take note of their home location and continue to experiment freely. Of course all courses of action affect the evolution of the soundscape, and in this each group's rendering of any given sequence is unique.

2. SOUND

Designing the sound for *Flox* presented challenges that differed significantly from the production of either static compositions or purely generative music. From the outset I wanted soundscapes that could be read as both pure sound and as semi-composed music. These soundscapes took the form of a series of pieces or sequences that are selected at random when a group enters the space. A sequences might have only four components (voices, layers, discrete parameters, processes, etc), or in a few cases eight or more. Given that there was no way to know how (or indeed if) visitors would resolve each sequence, it was clear that all sound components in the piece needed to be both distinctive enough that they could be individually identified with careful listening, and flexible enough in timbre that they could be mixed and remixed in any order throughout the sequence.

The sound in *Flox* is controlled and in many cases spatialized by the movements of the four participants as they move about the space. At the application level the primary voicing of the components is articulated by absolute X/Y coordinate positioning of each person, as seen by the motion tracking system mounted above the installation space. However, aside from this 2D mapping the system has access to a large amount of additional data for each visitor. These include: X-to-home, Y-to-home, cartesian distance to home, and every pod's relationship to every other pod. Also available for each pod were metrics such as mean activity, velocity and acceleration. From a compositional standpoint, these were many more metrics than could sensibly be used in any given sequence. Early testing showed that even sequences with four active components were difficult for some groups to resolve. This was mitigated somewhat by relegating more obscure metrics (inter-subjective distance, acceleration, etc) to less foregrounded elements of the sequences and leaving the primary voices to direct control. For example, one sequence that makes heavy use of beating tones triggers a dramatic modulation of the soundscape when the system detects that all four pods are physically close together. In another case, four pods moving in the same direction at the same time advances the piece which has more structural complexity to a hidden section. In yet another example, a collective stillness in the room causes four parts that are spatializing in increasingly faster circles to freeze in space until somebody moves again. In all cases it is entirely possible to resolve the sequence without discovering these hidden structures.

3. IMPLEMENTATION

3.1. Technical Requirements

As an interactive environment *Flox* has unique technical requirements. These are as follows:

1. *Accuracy*: The experience of continuous control over any given sound parameter is central to the piece, so it was determined that positioning needed to be accurate to within .3 meters, or about a single footstep in any direction.

2. *Reliability*: Since one of *Flox*'s central conceits is the direct embodiment of electronic sound, the tracking mechanism needed to be unobtrusive and highly reliable. It was critical that whatever tracking method was employed it provide persistent sorting over timet[1], and not interfere with the experience in the process. As tracking errors in such systems are inevitable, when such errors occur it was critical that recovery be graceful and predictable.

3. *Flexibility*: The system must be capable of generating accurate tracking data within an area as large as 15 square meters (the dimensions of the gallery space where *Flox* opened). Additionally, for the sake of portability the system needs to be resilient against significant differences in other physical spaces. These include the physical dimensions of the space itself, unpredictable ambient light levels, and the presence or absence of various structural features within the space (beams, vertical support columns, etc).

4. *Cost*: Budget was of course an important consideration, and the design phase went forward with the intention of finding the most inexpensive method that satisfied all the above criteria.

3.2. Video Tracking

The above requirements ruled out a number of technical approaches to solving the tracking problem. Approaches considered and rejected included RFID, and ActiveBadge scheme, a combination radio/ultrasound approach similar to that used by the MIT Cricket positioning system[2], Bluetooth, passive proximity beacon schemes, RF received signal strength (RSS) techniques and a traditional single camera tracking system. Ultrasound, a relatively cheap approach that looked promising in the early design phase, was deemed unsatisfactory due to reliability issues, and the fact that depending on environmental factors accuracy could not be ensured under +/- 1 meter. See Dijk [1] and Randell [2] for a thorough perspective on ultrasound tracking. A top-down video solution was ultimately selected as the tracking method.

Figure 1. The installation in action.

3.3. Multiple Cameras

Video tracking was initially written off as a viable solution due to the difficulties of reliably tracking a large space in a top-down fashion. These include visual distortion from the use of wide angle lenses, and line-of-sight issues produced by a central column and by the bodies of the participants themselves. Ultimately an acceptable solution was found by employing four relatively inexpensive dual-port firewire image sensors that were networked together in a star configuration. This approach allowed the 15 square meter gallery space to be divided into four equal regions, each tracked with its own camera. Using smaller regions in this way avoided problematic lens distortion, and provided flexibility when adjusting for the particulars of a given physical space. For example, the

---

[1] That is, once a sound or parameter was assigned to a certain individual upon entering the space, it was critical to that person's experience that the assignment remain consistent over time, even in the event of tracking errors.

[2] CSAIL, MIT http://cricket.csail.mit.edu/

gallery at High Concept Laboratories in Chicago, where *Flox* opened, provided a large space with a central structural column. Tweaking the four input matrices allowed me to crop out the center column altogether, providing continuous, relatively undistorted top-down tracking of the entire space. The four firewire streams are sent to a hub mounted on the ceiling, and then transferred to a firewire-ethernet repeater which feeds a computer in another room running a Jitter application. This application takes the stream from each camera and joins them into a single image which is then fed into a tracking algorithm and processed as if from a single camera. The algorithm makes use of Jean-Marc Pelletier's cv.jit library as well as a set of custom Max/Jitter externals I created that assist with tracking and error management.

Figure 2 shows the main interface panel for the installation. The black area depicts the floor of the installation space, as seen by the four cameras mounted on the ceiling. Visible are the white blobs ringed by numbered circles. These are the objects that are indexed and currently being tracked by the computer. The red numbered circles are the corresponding ”home” site for each pod. These home locations are changed every time a new group enters the space so that the following group is free to make their own discoveries. The grey circle in the lower right and the blue box in the upper right shows the spatialization configuration for sequence 5. In this example it can be seen that spatialization is following the locations of the pods directly, but any passive or active configuration is possible.

Figure 2. Flox interface depicting four tracked objects.

3.4. Error Handling

Predictably, early testing showed that temporary tracking errors were common and indicated that the system needed a robust way of managing error conditions.[3] Due to the nature of the installation itself it was particularly important that indexing be reliable and persistent; that is,when a lost object is recovered it is critical that the object be re-coupled with its original tracking index in order for the the continuity of the piece to be preserved. This is accomplished by examining the video stream for known objects on a frame-by-frame basis. When a new frame arrives, the index and coordinates of all known objects are stored in an array, along with a short historical snapshot of their vector representations averaged over a user-definable number of frames:

$$\frac{1}{n}\sum_{i=1}^{n}(\overrightarrow{P_nQ_n}) \tag{1}$$

In the event of a lost object or tracking error, the system retains the index and last known coordinates for the lost object in separate array, and compares it against any new objects that may appear subsequently. When a new object does appear, if only one object is stored as missing then the lost object is considered recovered, and the object is reassigned its original index. In the case of multiple simultaneous lost objects, when any new object is detected, the system defers the assignment of an index and a series of calculations are made for each of n lost objects:

$$\min\sqrt{\sum_{i=1}^{n}(p_i-q_i)^2} \tag{2}$$

wherein the coordinates of the newly arrived object are compared against the coordinates of each of the objects in the array of last known good locations. The index of the stored object with the smallest Euclidian distance is then assigned to the newly arrived object and the index is then removed from the array. In the unlikely case that two or more stored points are close enough in distance to make the match ambiguous, the results are sent through an additional stage to calculate

$$\theta_n(\overrightarrow{P_nQ_n}) \tag{3}$$

or the angular direction of the vector of each missing object. Another round of comparisons is made, this time against the prevailing direction of the newly arrived object vector, and the stored object with the greatest directional similarity is chosen as a match. In short, the algorithm assumes that a missing object will be recovered close to where it was lost, and in the event of an ambiguous match the object with the greatest similarities in both proximity and bearing is chosen.

This algorithm proved highly effective at maintaining persistent indexing in an unpredictable visual environment, and thereby a coherent experience for each visitor.

3.5. Pods

Each individual that enters the installation space is invited to carry a illuminated, radio-controlled pod with them, and informed that to goal is to return each pod to it's unique home location. As each person moves about the space, the pod they carry produces colored light that gives feedback about the state of the piece. Although it is the

---

[3] The most common error condition was the system temporarily losing a tracked object. This could be the result of many different factors, from changes in ambient light levels to a user stepping out of the tracking boundaries.

body of the person and not not the pod they carry that is tracked by the computer, being asked to carry a physical object with them in this manner provides a few key benefits within the installation. First, the pods create a constant tactile reminder that the actions of the participants are intimately linked to what is being heard at any given moment. Second, they provide initial encouragement in the form of visual feedback when the soundscape may be particularly chaotic earlier in the sequence. Listed here are a few of the conditionals, and the resulting visual feedback from the pods:

1. SEEKING: a pod that is actively seeking its home location will glow different shades of purple depending on how close or distant to home or resolved it is at any given moment
2. HOME: pods will fade from purple to slowly pulse blue light when it is moved inside a user-defined home radius
3. RESOLVED: when all pods are home and the soundscape is resolved all pods rapidly blink green as a visual indication that the sequence has completed
4. LOST: any pod that has been moved outside the boundaries of the installation space or which has been temporary been by the tracker will rapidly blink red until it has been recovered

In the event that a RESOLVED condition is achieved (all pods are successfully moved to their home locations), the system is able to reward the group effort by taking advantage of the fact that it knows where each person is in the room, and where they are in relation to one another. This takes the form of a spatially coordinated light sequence before each pod dims to indicate the group should exit the installation to allow another group of four to experience the piece.

But perhaps most importantly the pods afford a very strong ancillary benefit to the rest of the visiting audience. Although there are only four individuals allowed inside the tracking area at any given moment, there is a very strong performative component to the piece. Those visitors who chose not to directly participate in *Flox* or who are waiting to do so can easily follow the progress of those inside the piece, and so function as audience to a very particular kind of performance event. The controlled lighting, the colored pods, the slow, deliberate walking and the resulting music/soundscape create a distinctly ritualistic atmosphere that invites everybody in the space into the experience, not just those inside the tracked area itself.

3.6. Construction

The pods themselves are made of spherical translucent plastic enclosures, each containing an Xbee 802.15.4 radio module, an Arduino microcontroller, 3 superbright LEDs (red, green, blue), and a battery pack. Although using Xbee modules added slightly more to the total build cost of each pod, they provided significant benefits. In addition to a traditional serial AT command protocol the Xbee modules provide a frame-driven API mode that exposes each module's hardware ID, allowing for the construction of communication packets that include a header capable of referencing unique source and destination addresses. This in effect allows multiple unique Xbee modules to be addressed using a single channel or frequency, greatly simplifying the radio communication between the computer and the pods. A custom Max external was created that is responsible for parsing raw data from each tracked object, generating source and destination IDs, constructing a checksummed payload and sending it to a tethered Arduino via a hard serial (USB) connection, where it is transferred to the Xbee radio for transmission to the appropriate pod.

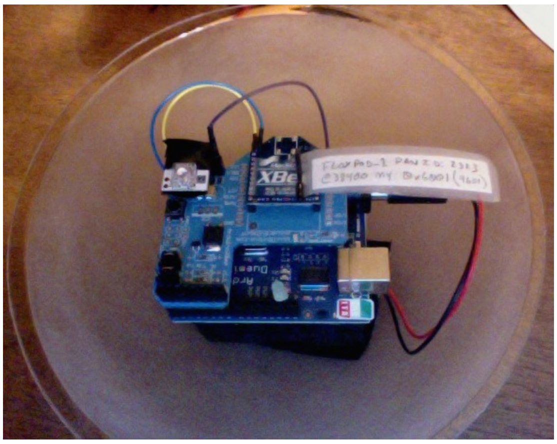

Figure 3. Inside a floxpod.

Figure 4. Outside the tracking area.

4. DISCUSSION

A look back reveals *Flox* achieved a number of its goals. The primary aim of using multiple interacting agents to articulate a piece of sound art was appreciated and well-received. I received much positive feedback about the social components of the piece, and it was fascinating to observe groups of friends negotiate the piece compared with groups of strangers. As might be expected, those groups who were more attuned to the actions of the group as a whole and who were more interested in communicating with each other were more ”successful” in resolving the sequences. In a number of cases groups who did not communicate well or who avoided communication altogether were unable to isolate individual components and were unable to resolve their given sequence. Having a musician’s ear also seemed to be an advantage, especially when negotiating sequences with more oblique mappings of movement to music.

As an experiment in whether it is possible to reframe the way electronically generated sound is embodied or experienced by an audience, *Flox* produced mixed results. It was reported by numerous visitors that the simpler, more literal pieces or sequences produced the strongest body-sound connections. For example, those sequences that made use of diverging/converging beating in different frequency ranges were easiest for people to identify, physically connect to and solve. But these were perhaps the least interesting musically. More complex mappings, such as schemes that attempted to tie musical structure to space were less successful, and in some cases simply confusing. Although some of this difficulty was foreseen, the degree to which the perception of embodiment was bound to initial musical decision-making was striking. Future productions of *Flox* will certainly take this into account.

I would like to express my deepest thanks and gratitude to Kevin Simmons and all the staff and artists at High Concept Laboratories in Chicago, IL where *Flox* was first produced. This project would not have been possible without their very generous support.

5. REFERENCES

[1] E. Dijk, K. van Berkel, R. Aarts, and E. van Loenen, “Ultrasonic 3d position estimation using a single base station,” in *Ambient Intelligence*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2003, vol. 2875, pp. 133–148, 10.1007/978. [Online]. Available: http://dx.doi.org/10.1007/978

[2] C. Randell and H. Muller, “Low cost indoor positioning system,” in *Ubicomp 2001: Ubiquitous Computing*, ser. Lecture Notes in Computer Science, G. Abowd, B. Brumitt, and S. Shafer, Eds. Springer Berlin / Heidelberg, 2001, vol. 2201, pp. 42–48, 10.1007/3-540. [Online]. Available: http://dx.doi.org/10.1007/3-540-45427-6_5