A Recursive Mapping System for Motion and Sound in a Robot between Human Interaction Design

Sinan Bökesoy
sonicLAB
15/6 DoganBey Sok. 34970 Istanbul
sinan@sonic-lab.com

ABSTRACT

We present in this article an artistic human-robot interaction form, which allows us to engage and evolve new experiences through to use of artificial animate behavior, kinesis and generated sound in a feedback communication medium. This meta system, which could be regarded also as one total composed instrument, is a recursive interaction process between its agents; a robot arm, human body and a sound space being rendered artificially. The computational modeling of this embodied approach is explained and a proto-type of the experiment is shown as a real tool to test empirically this artistic concept.

1. INTRODUCTION

Andy Clark argues that our thinking doesn't happen only in our heads but that "certain forms of human cognizing include inextricable tangles of feedback, feed-forward and feed-around loops: loops that promiscuously criss-cross the boundaries of brain, body and world" [1]. And instruments are built to achieve this extension of the mind, to express ourselves and translate our thoughts which we deploy to engage the world. Early experiments that examine similar issues can be found in the seminal cybernetic hybrid art of Stelarc [2], also in recent research of human-robot interaction for artistic purpose [3]. There are many examples of robotic mechanics in interactive control for acoustic instrument performances [4] in the community.
Sinan Bökesoy

In this project; interactions between a human and robot arm, representing our two agents, lets us develop the proto-type of a correlated sound vs. motion feedback mechanism. The direct/indirect feedback behavior is aiming the system to evolve relationships without perceived repetitive tendencies. A synthetic sound space within continuous interaction will be generated as a sensory feedback.Both system components become instrumentalists acting on a sound morphing engine, and their composed actions are not a combination of programmed states offered to them. It is considered that the human does not have any previous knowledge of the interaction results.

Spatial interfacing between agents focuses on the awareness of the robot arm through a Kinect[1] camera tracking and identifying the human body surface as a 3d world object. The human agent, positioned in front of the robot arm, listens and reacts to the sonic space and his body movements are translated in real time. The presence of a human agent in this environment will stimulate responses in form of robotic motion being mapped to this sonic space (kinesis). The discrete points of this space can be addressed by multi-dimensional vectors delivering a continuous morphology in this space.

To make the robot aware of its environment, real-time Kinect camera streams are processed and coupled with blob analysis to locate the gestural activity of the human along with the depth field data of his body and his hand positions. This allows the robot to create a kinesthetic response, in the way it positions its laser tool to move the laser point within the contours of the body surface, implying a certain intelligence to its movement. Its internal state depends on the human performer's actions, and that latter is being influenced from the sonic output controlled by the robot. The movement kinematics of the robot is introduced continuously to the system as the inner parameters of the robot, such as the individual euler angles, the position data of its 6 independent axis motors and its tool coordinates in three dimensions.

The functional role of the robot’s appearance here is not the performance of a physical music instrument, which is the general tendency in robot-human interaction music experiments [3][4]. In this work, the robot's task and the concept of the sonic interaction were adapted to robot's physical structure, which is the 6DegreesOfFreedom movement dynamics (referring to the freedom of movement of rigid body in 3D space) of an industrial robot projected optimally within the interaction mechanism.

The human tries to interact with the spatial vision of the robot according to his instinct reactions in sonic output and the response of the robot e.g. the reconfiguration of the robot arm in order to position his laser tool and pointing it on the human body with specified rules alters the vector state in

---

[1] Kinect™ sensor, Microsoft Corporation 2010.

the sonic space. Likewise a recursive feedback loop is formed up with the necessary interdependency between the elements. This builds also the creative momentum. The human agent's reactions to wanted sounds drives this change.

2. A CYBERNETIC CONFIGURATION

One of the interests here is to investigate a certain human - machine interaction situation in its physical form with an audible and visual emergence. The dynamical behavior is being shaped by the flow of decision and observation channels as remaining in the system and making it coherent but temporarily unfolding for its active operative agents. The kind of sensory data exchange and transformation of the system reconfigures the agent actions.

Our recursive structure is reminiscent to the triangular interaction process in Di Scipio's work [5] encapsulating the fundamentals of the cybernetic control theory [6]. On Figure 1, we do present the interaction mechanism of both systems. The recursive eco-systemic sound interaction of Di Scipio maintains the agents; human, computer hardware and the environment hosting the audible sound content within its room acoustics.

The human and the computer agent don't interact each other directly; they do interact with the acoustical host, they perceive the information propagated from it, process it and give their decision oriented sonic interaction back to the host. The host is invisible but the emergence become perceivable by the reactions of the DSP computer.

In our system, this host is being replaced by the 3D spatial environment; the human body spatial information, and in response to it the translated kinematic information of the robot with its inner dynamics, which is the most interesting quality of a 6DOF robot arm.

This mapping of the kinematics data in usable vector format will serve to the DSP computer to produce the emergent sonic process.

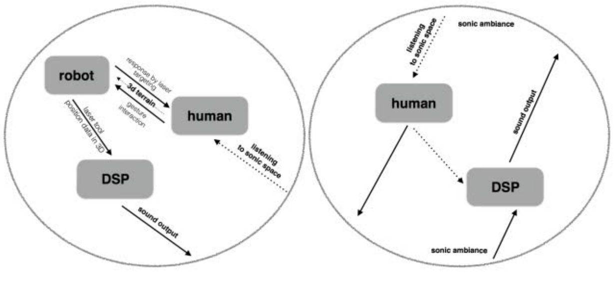


Figure 1. The comparison of recursive interactive systems (our experiment and DiScipio's interaction design)

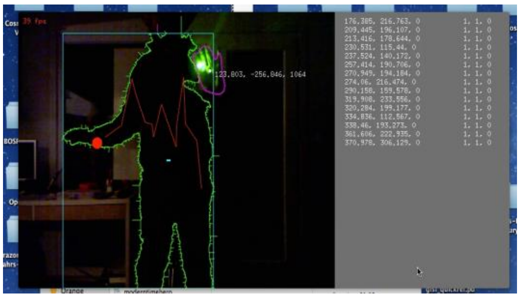
Again the human and the sound DSP hardware don't interact directly to each other. The sound space is being evaluated by the human. In a manner, he replaces the DSP hardware in Di Scipio's application. His actions involve a decision towards his preference in the sonic space transmitted as body gestures again sensed by the robot.

For a 6DOF robot arm, a linear movement produces non-linear kinematic expressions to be solved for the robot axis configurations [7]. Therefore we can have benefit of complex layers of motion and sound space mapping, where for example the 3D tool coordinates of the robot can be mapped directly to the morphing pointer of sound space but the data from the axis configurations of the robot can alter the micro level parameters of the sound defined around the morphing pointer. Likewise we can combine various configurations.

3. THE INTERACTION COMPONENTS

Industrial robots have certain kinematic algorithms where they can position the tip of their tool in a defined distance (or touch with it) to a work object, which is defined as a 3D geometry object. They do automatically arrange their axis manipulation in order to move their tool position with the appropriate orientation. Any path, defined with start and stop points on this work object will be traversed by the robot by successful interpolation of its tool position and orientation. (linear motion or joint motion [8].)

For our robot arm, a ABB IRB120, we have implemented a simple laser pointer tool, attached it to its gripper, which the robot can move precisely (0.01mm accuracy). Our aim is to move our laser tool by pointing it on the human body, namely by touching it virtually, on calculated curve paths inside the human body contour (Figure 2). The mechanism of this approach is explained in section 4.3.



```
176.385, 216.763, 0          1, 1, 0
209.445, 196.107, 0          1, 1, 0
213.416, 178.644, 0          1, 1, 0
230.531, 115.44, 0           1, 1, 0
237.524, 140.172, 0          1, 1, 0
257.414, 190.706, 0          1, 1, 0
270.949, 194.184, 0          1, 1, 0
274.06, 216.474, 0           1, 1, 0
290.158, 159.578, 0          1, 1, 0
319.908, 233.556, 0          1, 1, 0
320.284, 199.177, 0          1, 1, 0
334.836, 112.567, 0          1, 1, 0
338.46, 193.273, 0           1, 1, 0
361.606, 222.935, 0          1, 1, 0
370.978, 306.129, 0          1, 1, 0
```

Figure 2. Blob analysis of the human body and the hand in separate planes. The red curve is the robot path inside the human body contour.

During this movement, its tool position and orientation changes by its intelligent algorithm. This arm movement has an anamorphic structure and esthetically appealing as an artificial motion design. The data, which can be retrieved from the robot motion dynamics is the 3D tool position, the 6 axis angles (in Euler angles) of the arm and the 3D joint positions of the 6 axis [8] (Figure 3). We use this multi-dimensional vector to transcribe it to our sound morphing engine in various configurations.

The robot responds also to the human hand/palm by tracking its position for a specific reason. When the hand is raised with the palm directed towards the robot as seen on Figure 2 with the violet blob contour, it is the moment where the human agent shows his preference about the sonic timbre quality. The robot moves the laser point to the blob centroid, and continues tracking it. There, the human can traverse the sonic space around this point of his preference. This is the only moment for the system where the human can directly manipulate the sound-morphing engine by still interfacing through the robot.

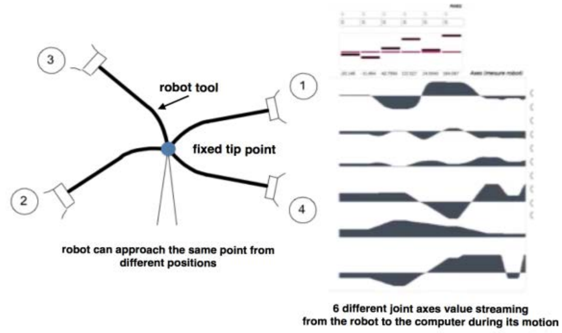


Figure 3. The robot orients its tool at different configurations to approach the same point and the change in joint angles of the 6DOF arm is shown on the right.

The hand position serves as a new sound morphing reference state. When the hand leaves the scene, this point will be remembered as a reference vector state which represents a corner of the tetra-hadron (see Figure 4) and the action goes on in the new spanned sonic space.

The transformative processes which are changing the internal states of the robot, are the 3D world positioning of the body and the attempt of robot laser targeting on it, with the specified rules. The robots adaptive behavior to its surrounding (the spatial definition of the human body) interferes directly with the morphing sound engine, which again is evaluated by the human agent by his intuitive reactions.

3.1 The Sound Morphing Engine

We define a sound space, where each entity of this space can be addressed with a linear relation of its parameters. Each dimension of the entity vector may stand for a perceptual parameter, sonic characteristic (low or high level) or a parameter of the synthesis engine under hood. An efficient morphing process provides an optimum control space mapped to the sonic space, which it does render continuously [9]. We have chosen the Cosmos$f$ stochastic synthesizer with its 3D morphing engine [10], which has a spherical representation of its morphing tool and the sonic entities inside in 3D visuals, hence offering a corresponding environment to map the 6DOF movement data of the robot. (see Figure 4).

In Cosmos$f$, one can define a preset which includes 100s of parameters of its synthesis engine as the snapshot of the current parameter space. The morphing engine can calculate a morphing pointer as the euclidian distance between for defined presets of Cosmos$f$ as being the morphing reference points in the parameter space.

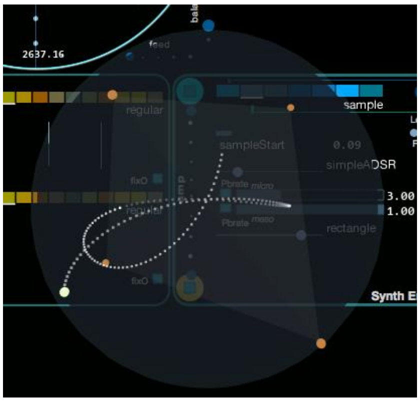


Figure 4: A sound morphing tool representing the state of the morphing pointer within a sphere and between 4 reference vectors (corners of a tetra-hadron).

The morphing pointer points to a multidimensional vector which continuously tranverses inside this parameter space, which can be visualized as a sphere. The 4 presets as the morphing reference states are represented as the corners of a tetra-hadron, which fits in this sphere (Figure 4). The white dots on Figure 4 are the discrete morphing states and therefore the parameter states of Cosmos$f$ rendering the sound space. The morphology is directly controlled through continuously altering of a 3D morphing vector pointer in the sonic space, which represents in our case a specific state of the underlying synthesis engine with hundreds of control parameters.

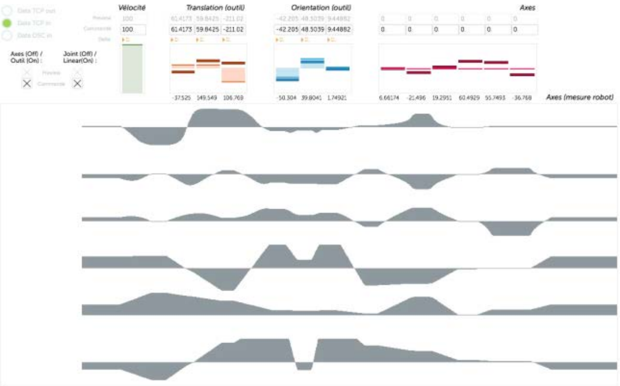


Figure 5: A simple linear motion of the robot gripper causes complex non-linear axis angle transformation

The morphing becomes indeed the interpolation of these parameter states in real-time and with audio rate precision between the four reference preset states (corners of the tetra-hadron). The robot tool coordinates are in 3D and can be easily translated to the polar coordinates of the morphing sphere and mapped as a vector pointer to trigger the morphology of the sound. On a previous research [11], we have tried to visualize the robot axis information in real-time as shown on Figure 5. The all 6 axis angles can be used here to map directly to a group of synthesis parameters of Cosmos$f$. This gives a narrower mapping field in the parameter space. However the attraction here is that the robot delivers a non-linear data output from a linear movement of its gripper e.g. the laser tool, this hidden quality becomes audible through out this mapping process between the axis angles and the chosen Cosmos*f* parameters. The software responds finally by pointing to the sonic space with reference vectors mapped through the interaction of the robot arm and the human body. This spatial interaction in real world will be heard as a sonic travel in continuum through out the sound space.

3.2 The human agent

Galeet BenZion asserts that kinesthetic and tactile learning are separate learning styles, with different characteristics [12]. She defined kinesthetic learning as the process that results in new knowledge (or understanding) with the involvement of the learner's body movement. In our case, the human and the robot are not in touch but physically aware of their spatial existence. The human existence provides the data for the robot to evaluate and move its tool so that the laser perfectly points on the target of 3D sensed terrain field.

The laser of the robot will be pointing on random paths to his body surface. The human while listening to the sound space can move and alter the terrain field, therefore interact indirectly with the morphing pointer generated by the laser tool coordinates and robot joint kinematics. The human learns about the qualities of the sonic space; the synthesis method and the mapping process is unknown to him. The human also can interact with the robot by holding his palm towards to robot the way he can alter the reference preset vectors in the sound space towards his preferred choice. Likewise the sonic space will be narrowed according to the target preferences of the human. This is a direct manipulation possibility of the system with an exterior input.

4. HARDWARE HOOK UP

The system structure can be categorized as following;

4.1 3D environment scan

For this purpose, the popular Kinect camera has been used and attached to the bottom part of the robot such as the robot tool, the work object (human body) and the Kinect coordinate system matches. (see Figure 6). They look towards the same x-y plane while the z-axis being the depth field of Kinect and also the axis where the robot is moving its laser tool towards the human agent. The Kinect camera provides a depth field of 640X480 resolution in 30fps guaranteed; and within the programming environment Openframeworks[2], one can apply the blob analysis for detecting the necessary contour information with the OpenCV[3] library (see Figure 2). The 3D terrain data obtained by the Kinect camera is transformed into a work object for the robot where it is using his laser tool to target on. Likewise we can keep the interest of the robot inside the body contours of the human, extract the centroid information from the palm contours of the human, +/- 2mm. accuracy depth information for adjusting the robot tool and the work object (human body) z-axis coordinates on the robot side. All the 2D points which we see on the RGB camera output of Kinect are represented in 3D world coordinates and being sent to the robot for as targeting assistance.

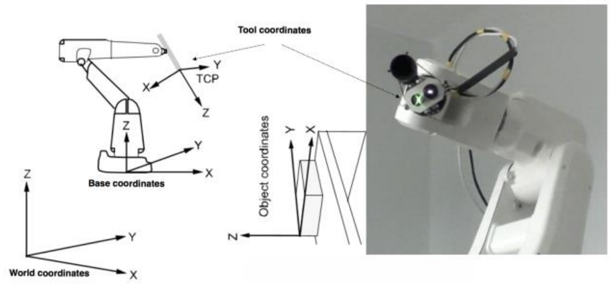


Figure 6: The orientation of the x,y,z axis for the robot tool and the work object,.

4.2 Robotic motion path

The computer software controlling the robot generates dynamic 3D curve data spanned inside the human body contour. The points of the curve should be reachable for the robot. When the body moves, the blob contour will change and the curve will adapt itself, brake into multiple parts etc, (see Figure 2) so that the segments remain inside the body contour. The curve data and the centroid data of the hands (if available) will be updated continuously.

4.3 Laser tool and targeting

The robot holds a laser tool (Figure 6), which it can turn on and off and point in real 3D world coordinates with a dedicated algorithm. Industrial robots do usually operate on surfaces with their tools attached to their hand. The robot will be given the definition of the surface (called work object) by its programming tools and it can position its tool by calculating the necessary kinematics (positioning in Cartesian coordinates/orientation in Euler angles) by itself to move smoothly along the surface [8]. But in our case; the surface, which is the human body, is not directly being touched and it is a dynamically changing one. The terrain data is being updated with the Kinect 3D scan and the robot has to adapt itself to this change of the surface. Therefore we change the *z* axis definition of the work object in the robot program according to the data coming from the Kinect in following steps;

[2] www.openframeworks.cc
[3] www.opencv.org

- The robot will be given its next path to follow on the human body as 3D point data.
- The work object *z* coordinate is updated according to the 3D point target and the laser tool will be given a virtual length, long enough that the distance between the tool *z* coordinate and the target point *z* coordinate can be covered by the movement of the robot arm on the *z* axis.
- With the virtual length given to the laser tool and having the work object *z* coordinate updated, the robot achieves a 'touching' gesture on the body movement with the laser pointing on the target.
- The complex kinematics data as a result of the movement of the robot gives the unpredictable output as tool coordinates, orientation, and robot joint data being used for the mapping process.

4.4 Data Exchange

There is an adequate communication protocol to maintain an efficient flow of parameters in real time and continuously. We have implemented a circular buffer for the data read/write operation between the visual analysis and the robot control application. The data exchange between system components will is structured as following;

- OSC[4] transfer between the visual analysis application (Kinect) and the robot driving software following the UDP protocol.
- TCP/IP connection between the robot and the robot driving software. Also a serial communication is possible between the two. The 3D coordinates of the tool and the 6 joint angle data is transferred as a RawBytes package all together.
- OSC transfer between the robot driving software and the audio morphing software.

We do need just one notebook computer running all the applications to drive the system. The robot receives the Kinect depth field data for its laser tool and the curve path it should follow with a RawBytes package transfer in 1024bytes at once.

5. PRELIMINARY RESULTS

We have seen another example of how to interact a human with a robot and establish a successful awareness between the agents in spatial dimensions. We have tried to document this proto-type as a paradigm in a system-theory view where a recursive loop between the output and defined inputs cause the creative development. It becomes a possible observation of a multi layered feedback mechanism serving as a real world environment of informational and sensory landscape. We state that the supplementary information as in feedback form does approach us gradually at the consciousness of the hidden values of the system. If our mind can experience new aspects of this physical environment, this very physical environment we do create reconfigures our creative capacity as well. Some improvements are already in schedule. From the development of the live performance arose the necessity to improve the motion path complexity showcasing a more attractive and intelligent response to the changes in its 3D environment. Response to human body in terms of cognition of its separate parts is possible through the Kinect SDK. Choosing the target points and drawing intelligent paths on the body surface will be developed further. The 6DOF robot arm motion already has its intelligent motion mechanism inherent, but we do try to augment it according to the sensory inputs that it receives.

Another point to consider is; choosing the robot targets and positions according to some prior or developed knowledge of the sound field, which it does influence. (memory/learning mechanism). This implies a two-sided enhancement where the robotic motion will be enhanced by the improved knowledge of its environment it does interact with. As stated, the human has no previous knowledge about the sound field but his cognitive responses develop quickly a memory based on his heuristic decisions; driving the sound space towards its preference.

6. REFERENCES

[1] A. Clark, *Supersizing the Mind*. Oxford University Press, ISBN 9780195333213, 2008.

[2] Stelarc, *The Monograph*. MIT Press, 2005.

[3] Buch, P. Coussement, L. Schmidt, “playing robot, an interactive sound instalation in human-robot interaction design for new media art” in *Proc. of the NIME conference* , Sdyney, Australia, 2010.

[4] A. Kapur, “et al. The Machine Orchestra” in *Proc. Int. Computer Music Conference*, New York City, 2010.

[5] A. DiScipio, “Sound is the Interface” *Journal Organised Sound*, vol. 8, no. 3, pp. 795-825, 2003.

[6] N. Wiener, *Cybernetics (Control and Communication in the Animal and the Machine).* MIT Press, Cambridge MA, 1948,1961.

[7] P. Melin, O. Castillo, *Modeling Simulation and Control of Non-linear Dynamical Systems.* Taylor & Francis, ISBN 0-415-27236-X, 2002.

[8] *Technical Reference Manual, Rapid Instructions, Functions and Data Types*, ABB Robotics, SE-721 68 Vasteras, 2012.

[9] M. Caeteno, N. Osaka, “A Formal Evaluation Framework for Sound Morphing” in *Proceedings of the ICMC conference,* Lubliana, Slovenia, 2012.

[10] S. Bokesoy, “Development of a Morphing Tool for the Cosmosf Synthesizer” in *Proceedings of the Journees Informatique Musicale Conference* , Paris, France, April 2013.

[11] Jacqumin G., Coduys T., “Partitions retroactives avec Iannix” in *Proceedings of the Journees Informatique Musicale Conference* , Bourges, France, April 2014.

---

[4] a content format developed at CNMAT

[12] G. BenZion, *An Analysis of Kinesthetic learners responses: teaching mathematics through dance*. Doctoral Dissertation. American University, Washington D.C.