EXTRACTION OF GESTURAL MEANING FROM A FABRIC-BASED INSTRUMENT

Doug Van Nort[*†], *David Gauthier*[†], *Sha Xin Wei*[†], *Marcelo M. Wanderley*[*]
[*]Input Devices and Music Interaction Lab/CIRMMT
McGill University, Montreal, Canada
[†]Topological Media Lab, Concordia University, Montreal, Canada

ABSTRACT

This paper presents an approach to the analysis of gestural data and extraction of related features from a cloth-based instrument. Issues surrounding the meaning of gesture and intentionality in such a performance environment are discussed, and we present a solution to analyzing and extracting information in a way that leverages the inherent quality of the cloth-as-controller. Other factors are considered in the system design, including the performance context in which the goal is to elicit improvised play from participants who do not possess an a priori model of interaction or vocabulary of acceptable gestural input.

1. INTRODUCTION

In the creation of interactive musical performance systems, there are often design constraints imposed by virtue of one's desire to map the expressivity of imagined and performative gestures into coherent sound results. Both augmented acoustic instruments as well as novel instrumental systems tend to look towards a performance paradigm based on a proscenium setting, built on a particular gestural vocabulary. With such constraints in place, interaction designers can make informed choices which deeply consider the selection of sensor technology, human-computer interaction model , mapping choice and including presentational and artistic choices [7].

In addition to the effect of presentational issues of music performance, a priori cognitive models of interaction are often physically introduced by an instrument's form factor, or the use of sensors and controllers (buttons, sliders, etc.) that are themselves separable or constrained to certain specialized actions and degrees of fredom [6] [8] [10] . In contrast, our design objective has been to augment improvised play through fabric-based interfaces that do not rely on knowledge of "instrumental gestures", nor do they segment or recognize motion based on underlying musical structures. Rather the goal is to promote the salient features of the textile (flexibility, stretchiness, texture, etc.) as determinant of the possible modes of interaction without directly referencing familiar cloth or fabric-based artifacts, or other human-computer interfaces [9]. This has led us to the design of a particular cloth controller (known as "the Blanket") as well as feature extraction methods that are the focus of this paper.

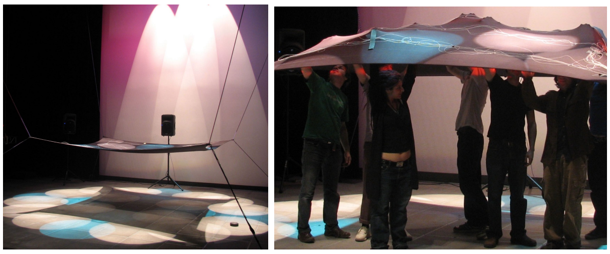
2. INTERFACE DESIGN

We constructed the Blanket from a 10' x 10' piece of highly stretchable Lycra fabric, on which is sewn a 5 x 5 grid of light-dependent resistors (LDRs) that span its surface. It is hung from the ceiling (or surrounding walls) through its corner rivets and positioned horizontally 6' above ground level, as in fig 1a. The sensing surface of the Blanket faces the ceiling and an array of lights are projected in parallel so that the intensity increases from faint at the Blanket surface to intense several feet above this. Players are underneath or to the side, and interact with the instrument by rising up and applying their upper body to it or by shaking the instrument from the boundary. This changes the shape of the surface and consequently the output of the light-dependent resistors, which then transmit to an Arduino [1] sensor interface for sampling and conversion. Through this design the Blanket does not represent an overly-suggestive artifact such as clothing, yet provides an eminent potential for interactiion and exploration simply through the manipulability of the fabric itself.

3. THE ROLE AND MEANING OF "GESTURE"

The notion of "gesture" has been studied from a variety of perspectives, considering its communicative aspects, the interaction within a cultural context as well as its functional properties. A clear division can be seen in the writing and classification of free-hand gestures and gestures engaged in the manipulation of an object. In the former case, gesture has often been considered as a channel of communication that augments speech, ranging from continuous support of vocal utterances to symbolic meaning that might take over the spoken channel [5]. In the case of object manipulation, it has been given thorough consideration in the case of instrumental gestures in music, leading to a structural description of gesture and its function in direct sound production as well as communicator of expressivity [2].

In the context of the Blanket and our larger system design, we focus on the *result* of participants' individual and collective input gestures. Our ultimate concern lies with the entire feedback loop of human-fabric-sound wherein the fabric acts as the primary transduction and determinant of the interactive potential. Thus, we consider gesture from a third point of view: those gestures that are



(a) (b)

Figure 1. The Blanket instrument (a) sans human interaction (b) collective play along the interior.

embedded in the fabric itself, as well as in the resultant sound. To this end, we consider the sum total of all individual and collective gestures as they emerge from the fabric, and gestural extraction approaches that are appropriate given these goals.

4. EXTRACTION OF MEANINGFUL GESTURES

4.1. Interaction and Intentionality

We utilize the statistical framework of correlation analysis and related methods. This allows us to explore possible continuous fabric topologies and their temporal behavior without committing to a single model through the use of high-level or overly-specialized data. It further allows us to explore the notion of *intentionality* in gesture, the markers of which stem from actions such as repeated or concurrent motions.

The LDRs are distributed uniformly across the Blanket surface, leaving no directional bias and allowing us to choose to track those areas which are most contextually relevant. The simple sensing modality of vertical displacement does not afford a rich interaction itself, but the number of available sensors and a proper consideration of their own inter-channel interaction can greatly enhance this. In particular, we "listen" to a given subset of the sensor grid, and analyze various spatial and temporal correlation sequences that are extracted from this.

4.2. Spatio-Temporal Correlation

Consider the raw control input stream as a single time-varying vector $X[n] = \{x_1[n], x_2[n], \ldots, x_N[n]\}$, wherein each dimension represents a different point on the sensor grid. From this information, we build a collection of estimators and feature extractions techniques based on the general idea of cross-correlation across spatial channels as well as temporal autocorrelation at given points along the cloth surface. Considering $X[n]$ as a wide-sense stationary stochastic process [4], we can express its generalized spatio-temporal correlation sequence as

$$R_n[k,i,j] = E(x_i[n]x_j[n-k]) \tag{1}$$

giving us an expression of the dependence between variables across space and time.

Now, this leaves us with a statistical framework from which we must build a proper real-world estimate, as well as a potentially intractable amount of data. The former problem is dealt with by looking at time-smoothed as well as instantaneous estimates of the data streams, which results in the two respective approaches:

$$\hat{R}_n[k,i,j] = x_i[n]x_j[n-k] \tag{2}$$

and

$$\tilde{R}_n[k,i,j] = \sum_{l=n-L}^{n} x_i[l]x_j[l-k] \tag{3}$$

where L is the window size of observation for the input streams of control data. The problem of reducing the amount of data from which we extract meaningful gestural features is dealt with by consideration and observation of the manner in which one interacts with the Blanket instrument, including some observations on the set of gestural actions that it affords and elicits.

4.3. Fabric-Based Interaction and Resulting Feature Extraction

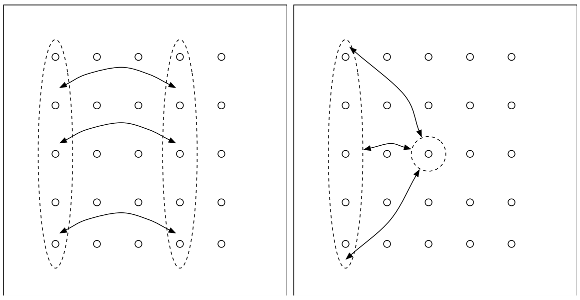
We don't make any strong modeling assumptions because of our goal to not strongly enforce cognitive models or schemas such as one would have in a classic instrumental performance context. Yet there are certain modes of interaction that we consider as indicators of intentional movement, including periodic motion and wave-like or repeated movements of the Blanket.

4.3.1. Multi-dimensional, Area-Based Correlates

With this in mind we consider certain areas of our cloth surface as having particular importance due to the shape and installation of the Blanket. This leads us to extract the multi-dimensional cross-correlation of these areas of interest. Topologically speaking, this relation does not have to constrain itself to the underlying sensor grid. Defining other correlative structures allows us discover many more natural and organic gestures, as these often do not arise in perfect orthogonality to the Blanket surface dimensions. For example, the interaction between boundary and center is of importance to the Blanket, as these two represent perceptual limits of the surface as well as natural points of interaction for individual (waving of the blanket) as well as collective play ("covering" of an inner participant, sending gestural waves back and forth). The fundamental difference between this approach to feature extraction and that of grid-based column correlation is depicted in figure 2. Further, this extraction approach differs from the expression of the previous section in that we consider an entire area of fabric space as a single entity, and examine its relation to another section of the soft controller over time. For example, the general interaction between two columns of the MxM grid becomes

$$\bar{R}_n = \sum_{i=0}^{M-1} x_i[n]x_{i+kM}[n] \tag{4}$$

meaning that we are taking the inner product of two columns, in this example the first and the kth. In the case



(a) (b)

Figure 2. Different Approaches to the Blanket topology include (a) the rectangular interaction between columns and (b) the circular interaction between boundary and center.

of the interaction between boundary and center, we take M copies of the center point and treat this as a vector of size M, then taking the scalar product with the boundary in question.

4.3.2. Instantaneous Spatial Correlates

The above approach examines the instantaneous cross-correlation of two areas of the Blanket surface. In a similar - but fundamentally different approach - we examine the momentary correlation between a collection of different Blanket locations. This gives insights into concurrent motion and the phase relationship between different points along the Blanket surface. In terms of the underlying Blanket gestures, this relates to both concurrent motion from group participation as well as waves that result from oscillations of single or multi-users. From a software design perspective, we construct a matrix that allows us to "tap" the Blanket surface at various locations, giving us a time-varying function that expresses the point-wise correlation (to augment the area correlation of above) as well as giving information about the directionality of movement.

4.3.3. Concurrent Autocorrelation Analyses

Finally, in this iteration of our system design and gestural analysis platform, we extract autocorrelation sequences from each spatial location along the Blanket surface. While the spatial correlates provide information about directionality and the "gestural shape" or contour of the fabric, this approach provides the temporal behavior and regularity of said contours and further augments our knowledge about the direction and degree of wave-like motion of the Blanket. The most essential information provided us is the degree of periodicity, presumed in our work to be a strong measure of intentionality. This use of the autocorrelation sequence is in parallel to that of fundamental frequency detection in music analysis [3]. We further note that at an even more basic level, regularity of motion results from intentional and collective gestures. In general this most basic form of intention results (paradoxically) in a more complex waveform along the Blanket surface than a simple periodic motion. As such, we examine the degree of harmonicity as well, usable both as a continuous parameter or as a change of our sound instrument's state. In a sound/music analysis context, the autocorrelation sequence has also been used to extract features such as the ratio of odd/even harmonics or of voiced/unvoiced parts of speech [11]. Our work has similar motivations in that we seek to uncover deeper salient features from temporal wave patterns, yet we search for other idiosyncrasies unique to our fabric control surface which may not take on meaning in an audio stream per se. We further have the added information of multidimensional input streams to consider.

While we do not classify collective gestures in our work, this approach allows said gestures (a marker of group intentionality) to elicit a meaningful response through consideration of spatial phase relationships. Though we look for fundamental modes of interaction that manifest on the cloth surface, any gestural input will result in a continuous response. This allows a participant to explore the system's response, finding the "resonances" that arise from our assumptions about intentionality as well as other meaningful responses that we have not considered, thereby adhering to a continuous human-in-the-loop processing chain.

4.4. Tuning of the Response and the Effect of Scale

There are several ways that the response and overall feel of the Blanket interface can be tuned. In particular, there is a dependency in regards to how the sensor streams are windowed across time and space. For example, the window size for each incoming stream in the case of temporal autocorrelation has a strong effect on the detection of periodicity as well as the time scale over which information propagates across the surface. There is a tradeoff between bias and variance in the data streams depending on the estimate used (e.g. equation (3) vs. (4)), which also extends to the assumptions of the data that is windowed for the cross-correlation sequences - such as our implicit assumptions of what occurs outside the given window of observation [4]. We treat these questions of windowing vs. instantaneous estimate, window type and length as free parameters that we use to adjust the system response. One interest of this research is how such adjustments might elicit different gestures to emerge as well as affect one's attention to multiple time scales, with both of these phenomena engendering different approaches to mapping fabric actions to sound synthesis.

4.5. Approaches to Gestural Analysis

After extracting spatio-temporal correlation sequences over varying windows of time and subsets of cloth, there still remains an interpretive layer needed to map this information to perceived gestures along the Blanket surface. For a detection of periodicity, and further finding the fundamental period, we use classic techniques [11] [3] of peak picking, particularly finding the argmax of the first peak of non-zero lag within the normalized autocorrelation sequence. The degree of of *periodicity* is tied to the value

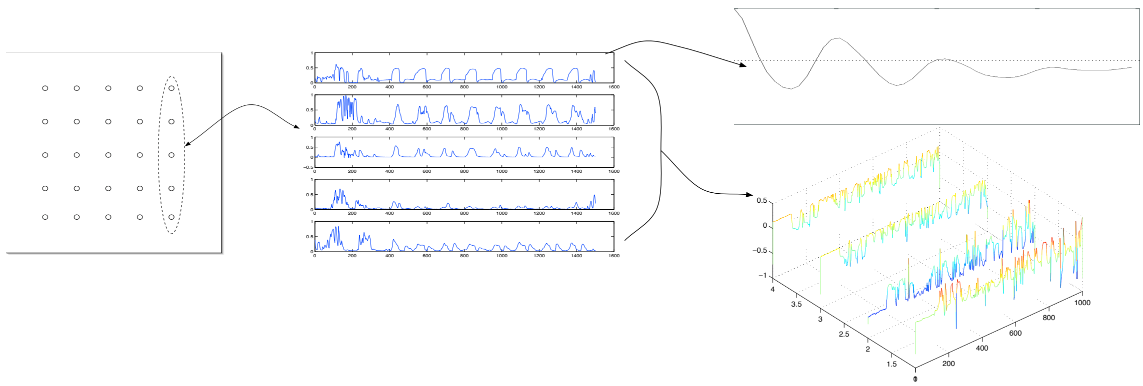


Figure 3. Sensor data from Blanket (left) is "tapped" as time series (center) and autocorrelation (right top, function of lag time) and cross correlation (right bottom, function of spatial lag and time) series are extracted in order to provide information about regularity and phase relationships, respectively. Above data results from flapping the Blanket edge closest to the chosen column.

of the given peak, and we further extract the degree of *harmonicity* from the combined values of all peaks that exceed a user-defined threshold - where this threshold is again a free parameter that can tune the system response and is highly composable. If we take the harmonicity normalized by the total power of the signal, then we have a measure of the *harmonic-to-noise ratio*, which can be used as a measure of how many people (in a multi-user context) are engaged in producing collective gestures. Finally, we consider the *directionality* of Blanket gestures. This feature can be extracted in several different ways, such as looking at the gradient of the projection "surface" produced by measuring the degree of periodicity at each location, or by cross-correlating with appropriate kernel matrices.

In sum, these measures provide a mapping into gestural phenomena, given by contours, paths and fluctuations of the Blanket surface. In terms of the Blanket architecture this can be seen as a meta-layer that we consider in terms of human enaction, the physical qualities of our fabric and ideas of collective and intentional meanings and is considered as somewhat independent of the sound synthesis methods. Our explorations into the mapping into sound parameters – discussed in detail elsewhere [9] – include mapping the behavior of the sensor "grains" into granular synthesis parameters, using global and local harmonicity to control tuning of granular parameters, mapping regions of activity to intensity level of different sound processes, source localization to spatialization and so on.

5. FUTURE WORK AND CONCLUSIONS

The presented framework of spatio-temporal correlation analysis was chosen because of its flexibility and limited assumptions. Extracted features related to intra-cloth interactions that describe phase relationships, regularity of motion, direction, etc. reveal the complex interdependence across the fabric surface and between collaborative players. Further gestural meaning can be considered from the combinations of these features, considering the notion of intentionality in gesture and the physical affordances of a given fabric, in this case the Blanket instrument.

6. ACKNOWLEDGEMENTS

This work has been conducted in the context of the *WYSIWYG* project, and includes Freida Abtan, David Birnbaum, Rodolphe Koehly and Elliot Sinyor. The project has been funded by Hexagram and the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT).

7. REFERENCES

[1] Arduino. http://www.arduino.cc/. *Accessed July 5, 2007*.

[2] Claude Cadoz and Marcelo M. Wanderley. Gesture-Music. In *Trends in Gestural Control of Music*, pages 315–334. IRCAM – Centre Pompidou, 2000.

[3] A. de Cheveigné and H. Kawahara. Yin, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. Am.*, 111:1917–1930, 2002.

[4] Simon Haykin. *Adaptive Filter Theory*. Prentice-Hall, fourth edition, 2002.

[5] Adam Kendon. *Gesture: Visible Action as Utterance*. Cambridge University Press, 2004.

[6] Benoît Maubrey. Audio jackets and other electroacoustic clothes. *Leonardo*, 28(2):93–97, 1995.

[7] Eduardo Reck Miranda and Marcelo Wanderley. *New Digital Musical Instruments: Control And Interaction Beyond the Keyboard*. A-R Editions, Inc., 2006.

[8] Maggie Orth, J. R. Smith, E. R. Post, J. A. Strickon, and E. B. Cooper. Musical jacket. In *SIGGRAPH '98: ACM SIGGRAPH 98 Electronic art and animation catalog*, page 38, New York, NY, USA, 1998. ACM Press.

[9] Sha Xin Wei, Marcelo M Wanderley, Freida Abtan, David Birnbaum, David Gauthier, Rodolphe Koehly, Elliot Sinyor, and Doug Van Nort. WYSIWYG: Wearable Sounds, Gestural Instruments. *Submitted for Publication*.

[10] Gili Weinberg, Maggie Orth, and Peter Russo. The embroidered musical ball: a squeezable instrument for expressive performance. In *CHI '00: CHI '00 extended abstracts on Human factors in computing systems*, pages 283–284, New York, NY, USA, 2000. ACM Press.

[11] Udo Zölzer, editor. *Dafx: Digital Audio Effects*. John Wiley & Sons, Inc., New York, NY, USA, 2002.