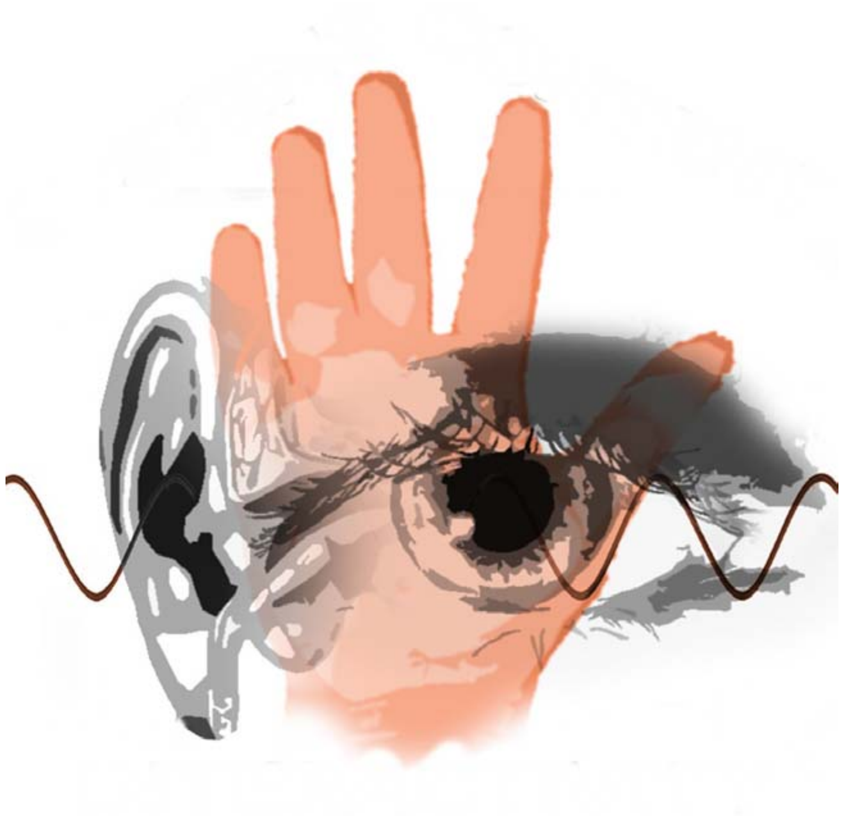
National Science Foundation

# Interactive Digital Multimedia IGERT Annual Research Review



## Year 2: 2004-2005

University of California, Santa Barbara

**Proceedings**

# Interactive Digital Multimedia IGERT Annual Research Review

# Year 2: 2004-2005

October 7, 2005
University of California, Santa Barbara

**Supported by:**

National Science Foundation Grant# DGE-0221713
“Digital Multimedia: Graduate Training in Interactive Digital Multimedia”

## Interactive Digital Multimedia IGERT Annual Research Review

## Year 2: 2004-2005

## Table of Contents

# 2005 Interactive Digital Multimedia IGERT Participants

## Fellows/Trainees:

Maria del Mar Alvarez Rohena, Computer Science
August Black, Media Arts and Technology
Laura Boucheron, Electrical and Computer Engineering
Stephen DiVerdi, Computer Science
Julie Dillemuth, Geography
Barbara Drescher, Psychology
Peter Khooshabeh. Psychology
Justin Muncaster, Computer Science
Dan Overholt, Interdisciplinary PhD
Michael Quinn, Electrical and Computer Engineering
Wesley Smith, Media Arts and Technology
Bob Sturm, Electrical and Computer Engineering
John Thompson, Music
Alex Villacorta, Statistics
Nhat Vu, Electrical and Computer Engineering

## 2005 Research Associates:

MarkDavid Hosale, Media Arts and Technology
Thomas Kuo, Electrical and Computer Engineering
Mary Li, Electrical and Computer Engineering
Lance Putnam, Media Arts and Technology
Bhaskar Rao, Media Arts and Technology
Brian Springer, Art Studio
Jason Wither, Computer Science

## High School Research Apprentices:

Max Wiedmann and Shane Kendrick

## Alumni:

Tony Boughman, Geography
Carlos Castellanos, Media Arts and Technology
Ethan Kaplan, Media Arts and Technology

## Faculty Advisors:

Kevin Almeroth, Computer Science
Keith Clarke, Geography
Miguel Eckstein, Psychology
Sara Fabrikant, Geography
Jerry Gibson, Electrical and Computer Engineering
Michael Goodchild, Geography
Mary Hegarty, Psychology
Tobias Hollerer, Computer Science
S. R. Jammalamadaka, Statistics

JoAnn Kuchera-Morin, Media Arts and Technology / Music
George Legrady, Media Arts and Technology / Art
B.S. Manjunath, Electrical and Computer Engineering
Rich Mayer, Psychology
Marcos Novak, Media Arts and Technology / Art
Lisa Parks, Film Studies
Marko Peljhan, Media Arts and Technology / Art
Stephen Pope, Media Arts and Technology
Curtis Roads, Media Arts and Technology
Kenneth Rose, Electrical and Computer Engineering
Ambuj Singh, Computer Science
Matthew Turk, Computer Science / Media Arts and Technology

## Postdoctoral Researchers:

Matthias Kölsch, now at the Naval Postgraduate School
Xinding Sun, now at Microsoft

## New IGERT Trainees in 2005-2006 (and degree institutions):

John Roberts, Computer Science
 B.S., San Francisco State University
Joriz DeGuzman, Computer Science
 B.S., California State University, San Bernardino
Emily Moxley, Electrical and Computer Engineering
 B.S., Princeton University
Jim Kleban, Electrical and Computer Engineering
 M.S., Rutgers University
Bo Bell, Music
 M.M., City University of New York, Brooklyn College
Drew Dara-Abrams, Psychology
 B.S., Carleton College
Brent Hecht, Geography
 B.A., Macalester College
Matt Peterson, Psychology
 M.S., U.C. Santa Barbara
Karl Grossner, Geography
 B.S., California State University, Chico

## External Advisory Board:

Dr. Michael Century, Chair of Arts Department, Rensselaer Polytechnic Institute
Dr. Tony Chan, Dean of Physical Science, College of Letters and Science, and Professor of Mathematics, University of California, Los Angeles
Dr. Fiona Goodchild, Education Director, California Nanosystems Institute, University of California, Santa Barbara
Dr. Ramesh Jain, Donald Bren Professor in Information & Computer Sciences, Department of Computer Science, University of California, Irvine
Dr. Alvy Ray Smith, co-founder of Pixar, retired

## Interactive Digital Multimedia IGERT Annual Research Review

## Year 2: 2004-2005

### Year in Review

As we conclude the second year of our NSF Interactive Digital Multimedia (IDM) IGERT program, we note significant growth in the number and diversity of the supported students, and most importantly, the overall high quality of the funded research. Perhaps unique within the NSF supported IGERT program, we have students from three different academic traditions on campus: Engineering, Science, and the Humanities. With an emphasis on interdisciplinary collaboration, our IGERT fellows again formed diverse teams to take on projects exemplifying multimedia research. In some cases, these projects coalesced around a specific topic spanning multiple academic fields. In other cases, student with independent topics came together, combining their results into one single technical/artistic effort. Our IDM IGERT projects address problems ranging from visualizing and auralizing datasets to interacting with computers and intelligent spaces. Three students spent the summer of 2005 interning with industry: two at IBM Almaden Research Center, and one at Los Alamos National Labs.

Also, three students received federal funding this year to continue their research in interactive digital multimedia: Julie Dillemuth received a 2005 Graduate Research Fellowship from NSF; Pete Khooshabeh received a fellowship from the Department of Homeland Security; and Mike Quinn received a fellowship to participate in the NSF LEAPS Program (Let's Explore Applied Physical Sciences) at UCSB.

### IGERT Seminar Series

Our seminar series proved to be very well-attended in 2004-2005. Here is a list of our invited speakers:

10.15.2004 Prof. Michael Zyda, MOVES Inst., Naval Postgraduate School
10.29.2004 Prof. Hong Tan, ECE, Purdue University
11.05.2004 Prof. Shawn Brixey, Digital/Experimental Arts, University of Washington
01.14.2005 Prof. Mihaela van der Schaar, ECE, UC Davis
01.21.2005 Prof. Michael Goodchild, UC Santa Barbara
01.28.2005 Dr. Ilya Goldberg, National Institute of Health
02.04.2005 Mr. Florian Hollerweger, Tech. University of Graz (MATP Visiting scholar)
02.11.2005 Dr. Keith Butler, Boeing Phantom Works
02.18.2005 Prof. Ram Seshadri, Materials (Ethics Seminar), UC Santa Barbara
02.18.2005 Dr. Xavier Amatriain, Pmpeu Fabra University, Barcelona
03.04.2005 Prof. Stephen Pope and Student Guests, UC Santa Barbara
04.01.2005 Prof. Ismo Rakkolainen, FogScreen, Visiting Researcher in 4Eyes Lab
04.08.2005 Prof. Hari Sundaram, Arts, Media, and Engineering, Arizona State University
04.15.2005 Prof. Ulrich Neumann, Computer Science, USC
04.22.2005 Prof. Lisa Jevbratt, MAT & Art Studio, UCSB
05.06.2005 Prof. Mohan Trivedi, ECE, UC San Diego
05.20.2005 Prof. Elaine Chew, Industrial and Systems Engineering, USC
05.27.2005 Prof. Mary Hegarty, Psychology, UCSB

## The Future

We end our sophomore term with 20 participating students (13 Trainees and seven Research Associates). Our third year brings nine new Fellows into the program and the guarantee of more Associates later on. Furthermore, on September 20, 2005 the proposal for a PhD in Media Arts & Technology was approved by the University of California. We expect this PhD track will provide the Media Arts and Technology Program at UCSB with greater representation among IGERT Fellows, and we are enthusiastic about what this means for the relationship between MAT and the IDM IGERT.

Although we are still in the process of outfitting our central lab, we are pleased with the progress we have made in terms of research and interdisciplinary collaboration. One of the student research projects, resulting from the work of five students from Art, Computer Science, ECE, and Media Arts and Technology, will be featured in the main UCSB library as an interactive installation. Another, a technique for interacting with biological datasets, will be featured at one of the upcoming concerts in the Center for Research in Electronic Art Technology. The research of our IGERT certainly represents a diverse look at the field of digital multimedia.

Following is a summary of the recent research being conducted by the IGERT students, the initial ideas of which were discussed during the one-day project retreat on May 13, 2005. Attached is a collection of our students' publications to date.

## Acknowledgements

We would like to thank Dean Matthew Tirrell, College of Engineering, Prof. Charles Li, former Dean of Graduate Division, Dean David Marshall, Humanities and Fine Arts, Dean Martin Moskovits, Mathematical, Life, and Physical Sciences in the College of Letters and Science, and Chancellor Henry Yang, for their continued encouragement and support. We would also like to thank Prof. JoAnn Kuchera-Morin, former Chair of Media Arts and Technology Program, Prof. Umesh Mishra, Associate Dean, College of Engineering, and Prof. Tim Cheng, Chair of Electrical and Computer Engineering Department, for their help in building this program. Finally, we wish to give special thanks to our external board of advisors for their guidance. This IGERT project is supported by the NSF Award #DGE-0221713.

**Principal Investigators:**

B.S. Manjunath, Director, Professor of ECE, manj@ece.ucsb.edu
George Legrady, Professor of Media Arts & Technology and Art, legrady@arts.ucsb.edu
Stephen Travis Pope, Media Arts & Technology, stp@create.ucsb.edu
Kenneth Rose, Professor of ECE, rose@ece.ucsb.edu
Matthew Turk, Chair, Media Arts & Technology, Professor of Computer Science, mturk@cs.ucsb.edu

**Education Coordinator:**

Tim Robinson tim@ece.ucsb.edu

# Interactive Digital Multimedia IGERT
# Student Research Projects - 2005

# Out of the ether: A system for interactive control of virtual performers using video and audio streams

JoAnn Kuchera-Morin MAT    Mary Li ECE    John Thompson Music    Jim Kleban ECE    Lance Putnam MAT

*Abstract*—**We present a real-time multimodal system that performs along with a human instrumentalist triggered by a combination of features from captured video and audio data. Three distributed components, an "Eye," "Ear" and "Brain" simulate the perceptive process of an accompanyist. A "Voice" component resynthesizes analyzed audio from the performer triggered by messages from the Brain. An implementation of the system responds to visual gestures by a flutist performing a music piece.**

*Index Terms*—**Human Computer Interaction, Gesture Recognition, Score Following, Pitch Detection, Multimodal Analysis**

## I. Introduction

WITH the recent advances in technology and human computer interaction a new realm of creativity has opened to both musical composers and performers. This paper describes work towards implementing a networked computing system that can provide the framework for original interactive compositions for a wide variety of musical styles. Gesture recognition techniques coupled with audio analysis can provide novel and robust ways for computer systems to "play" along with one or more human musicians, to much greater extent than simply adding prerecorded tracks to a performance. A variety of real-time computer resynthesis techniques are enabled through a virtual performer such as parameterized reverbs, counterpoints, or sound spatialization and these techniques can be controlled directly and indirectly by the human performer using visual and audio cues.

### A. Background

Music written for musical instruments and electronic sounds has traditionally used a pre-recorded tape for the electronic accompaniment. The performer is made to synchronize to the fixed media, restricting expressive control that live performance typically allows. Attempts to create a virtual performer more sensitive to the human performer have focused on extracting information from a human performer's actions on the instrument, in combination with analysis of the resulting audio signal.

In the 1970s Morton Subotnick attempted to derive control information for analog envelope generators from the performance audio signals. The complexities of controlling the various parameters on analog equipment led Subotnick to work with the computer in the 1980s. Subotinik sought to make the computer "respond properly as if it were a musician following a conductor." Working with Barry Vercoe, they pursued methods to allow the computer to follow the performer's position in the musical score [9]. The research into score following techniques has been furthered by Roger Dannenberg's extensive work involving automated accompaniment[8].

Methods to capture performers' actions have often employed instruments modified with various sensors meant to grab control data (typically MIDI data) from the performer. Work at IRCAM has made possible sophisticated compositions using these modified instruments, such as a the use of a modified flute in Philippe Manoury's "Jupiter." Unfortunately, these modified instruments remain experimental and have not become a standardized aspect of the musical world. An exception to this are the electric guitar and the midi-outfitted piano, the Diskclavier.

Frustrated by the inability to sway the tradition toward incorporation of electronics, some musicians have abandoned the widely available instruments of the standardized repertoire in favor of individually tailored hyperinstruments. The radical departure from the acoustic tradition presented by some hyperinstruments is rewarded by the ability to easily interface with the computer in novel ways; however, it pushes away many of the most talented and traditionally trained performers who remain rooted in the acoustic tradition.

Extracting control features from a musical performer without the use of modified or hyperinstruments, has focused primarily on deriving information from the performance audio signal. Cort Lippe's "Music for Clarinet and ISPW" capitalizes on the work of Miller Puckette at IRCAM in the late 1980's. In this piece the virtual performer is informed entirely by features extracted from the clarinet's performance audio signal. Score following is achieved by deriving the pitch of the audio signal, and parameters of compositional algorithms are driven by features of the amplitude and spectral domains.

Recently, a multimodal analysis of the performance space attempts to use both visual and audio information to control the virtual performer[2].

### B. Motivation

The difficulty of deriving meaningful and predictable control data from only audio and visual signals has encouraged many to again embrace an indeterminate approach when working with the virtual performer. A loose interpretation of the actions results in an incidental human-computer interaction system in which a great deal of variety and complexity can be generated at the cost of deterministic control. This work describes the implementation of a robust system of multimedia sensing and interaction for the creation of music and multimedia of a more deterministic nature.

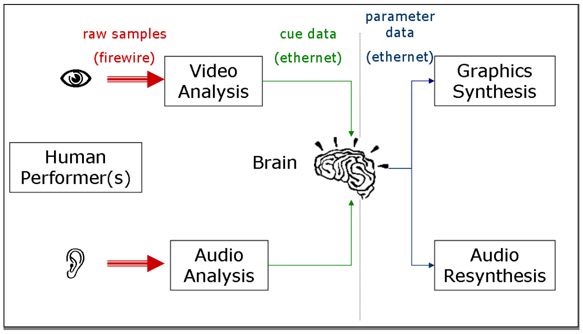



Fig. 1. Block diagram of the virtual performer system.

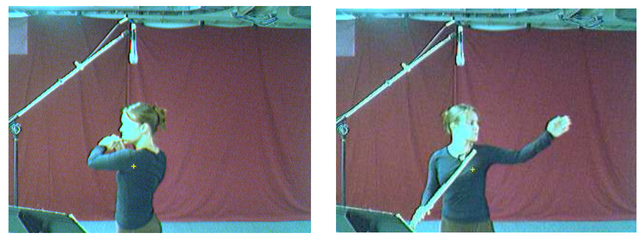
## II. System Overview

The virtual performer system (Figure 1) consists of four distributed systems: an image processing gesture recognizer (Eye), a pitch detector (Ear), a score follower/integrator (Brain) and an accompanying audio resynthesizer (Voice). The visual recognizer detects predefined gestures in the video stream captured in real-time from a camera centered on the human instrumentalist. The Eye sends gesture messages upon detection via Open Sound Control (OSC) packets to the Brain integrator. The pitch detector determines the presence of a prevailing pitch in the real-time audio input stream from a microphone placed in proximity to the performer. The Ear sends its pitch estimates with corresponding times to the Brain over the network. The Brain has apriori knowledge of the musical score and thus can track the instrumentalist as she works through a piece, score following. When visual gestures from the performer are detected that occur at the correct positions in the score, the virtual performer has then been cued by the instrumentalist to respond. The Brain then sends control messages to the Voice for executing sound processes/transformations determined in advance by the composer. One typical response used in many electroacoustic works is to playback recorded audio buffers of the performer's output at various points along the piece usually with some kind of sound transformation applied.

The "Out of the Ether" virtual performer is a robust, portable, real-time system operating on four distributed components. This specific hardware implementation uses two mobile computing platforms: a notebook PC running the Brain and Eye and a Mac Powerbook G4 running the Ear and Voice. A Unibrain FireEye camera captures 320x240 resolution video at 15 fps and transfers data over Firewire. A high quality microphone captures the performer's audio which is then sampled at 44.1 kHz. This system has been tested using a flutist performing John Thompson's composition "Etude for Flute and Virtual Performer."

## III. The Eye

Much work has been done in gesture-based human computer interaction, however, there is only a limited amount in the context of music. Some previous work resembling ours is described in [1]. A gesture not only serves the purpose of communication between computer and human, but also to present an expressive element of musical performance as studied by Vines et al. [2]. A similar system aimed at using expressive gesture analysis to generate real-time visual and audio content was developed by Camurri et al. [3]. Our approach to the gesture recognition was to use representative and reduced dimension feature models to train and to develop a performer neutral recognition system. The series of gestures we defined not only feel natural for the performer, but also have an element of drama to further engage the audience with the musical performance. Two of the gestures are shown in Figure 2.



(a) (b)

Fig. 2. Sample gestures of (a) a right turn and (b) a left arm extension used to cue the virtual performer.

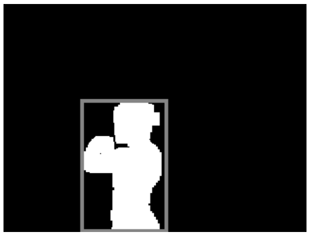
### A. Feature Extraction

In order to determine what good features are for a gesture recognition problem, one needs to investigate the context of the problem. In our scenario, a musical performance, the background is assumed to be a stage and therefore static. This implies with the help of a good background subtraction algorithm [4] we can segment out the motion of the human player. We also need features that possess characteristics that are representative of the gestures and computationally inexpensive. It would also be ideal if the features were invariant to color and suitable for different players. After considering the above constraints and testing on a few different types of features, we chose 2D projections as the feature based on a binary image that only contains the silhouette of the human player. A horizontal projection combined with a vertical projection captures the shape of the silhouette in orthogonal directions. To reduce computation, we use the result from 2D motion tracking and only work on pixels within a bounding box defined by the predominant moving object in the image. Each element of the vector is computed as the summation of the non-zero pixel values along the row (or column) inside of the bounding box of the silhouette. This is shown in Figure 3.

$$Hp_{r_i} = \sum_{c \in r_i} I_{r_i,c_j} \qquad (1)$$

where $Hp_{r_i}$ represents the horizontal projection along the row $r_i$, $I_{r_i,c_j}$ is the intensity value at pixel $(r_i, c_j)$, namely 1 or 0.

After each frame is processed for background subtraction and feature extraction, the features then are used as input



(a)

$$f = \begin{bmatrix} Hp_1 & Vp_1 \\ \vdots & \vdots \\ Hp_{240} & Vp_{240} \\ 0 & \vdots \\ \vdots & \vdots \\ 0 & Vp_{320} \end{bmatrix}$$

(b)

Fig. 3. (a) Silhouette representation from which the (b) feature matrix is extracted

to the recognition routine to determine if there is a gesture present in the frame. In the next section, we will present two methods used.

### B. Gesture Recognition

In order to make use of the features and as well as reduce computation cost, we chose the following two methods: Minimum Mean Square Error (MMSE) criteria; (Equation 2 below) and correlation-based template matching (Equation 3 below)

$$gest1 = min\,|f_c - f_g|^2 \quad (2)$$

$$gest2 = bestMatch(f_c, f_g) \quad (3)$$

where *gest1* is gesture determined by method 1, and *gest2* is by method 2. In the first method, the system outputs the gesture that minimizes the squared distance between the feature points $f_c$ in the current image and the defined gesture $f_g$ based on an empirical threshold derived by running a series of experiments. This empirical threshold also helps reduce the false alarm produced by the second method. In the second method, the best match was obtained by two 1D cross correlations. A gesture is produced by the overall recognition system whenever *gest1* equals *gest2*. The recognition system thresholds are designed to eliminate misses while allowing for less serious false alarms. This design gives leeway to the Brain in making decisions as to what action the virtual performer should take along with the audio information sent from the Ear.

## IV. The Ear

In musical performance, the predominant auditory cue from one performer to another is that of frequency, both on the meso time scale of rhythm and on the micro time scale of pitch. The Ear part of the system analyzes incoming audio from the human performer using a fast Fourier transform and makes a decision on whether there is a pitch present and if so what the pitch is in Hertz. Our application required a pitch estimator that would be robust against noise, computationally efficient, and take advantage of the fact that the phase vocoder [7] would be used as a compositional device. In order to distinguish between a noisy signal and a periodic signal, a thresholding mechanism was used that dropped spectral bins below a certain decibel level of the maximum magnitude bin. The spectrum of a pitched sound has peaks at periodic intervals above a certain noise floor based on characteristics of the analyzed signal. A periodic signal will have a few regularly spaced peaks while a noisy signal will exhibit a higher density of irregularly spaced peaks around the maximum magnitude bin. For estimating an actual pitch, the harmonic product spectrum (HPS) was used [5]. The HPS is a product of a spectral frame by integer down-sampled versions of itself. The resulting spectrum generally has a peak at the fundamental frequency of the signal due to the fact that harmonics align when the compressed spectra are multiplied.

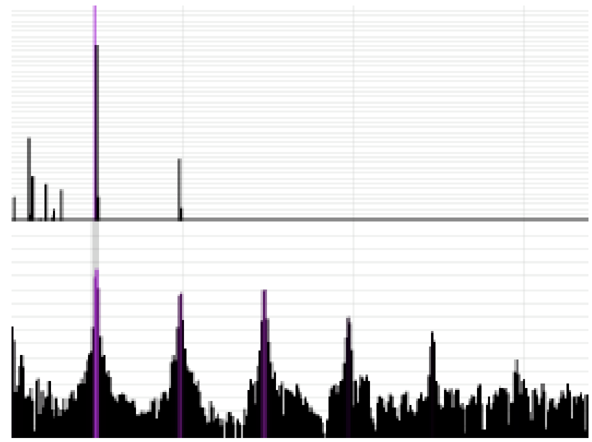


Fig. 4. The frequency (bottom) and harmonic product (top) spectra from 0 Hz - 3.5 kHz of a flute sound.

Due to the fact that a Fourier transform by itself is limited to a quantized set of frequency values, the instantaneous frequencies of the sinusoids [6] were calculated to give a more accurate pitch estimate. One prevalent problem with pitch detection algorithms is octave errors. To circumvent this, a filter was employed that would require at least two consecutive frames to have similar pitch before a new pitch was considered valid.

## V. The Brain

The Brain's function is to acknowledge dramatic moments specified in the performance and activated by explicit gestures from the peformer and then activate the response of the virtual performer. Thus, the Brain performs the task of integrating the audio and visual information from the pitch detecting Ear and the gesture detecting

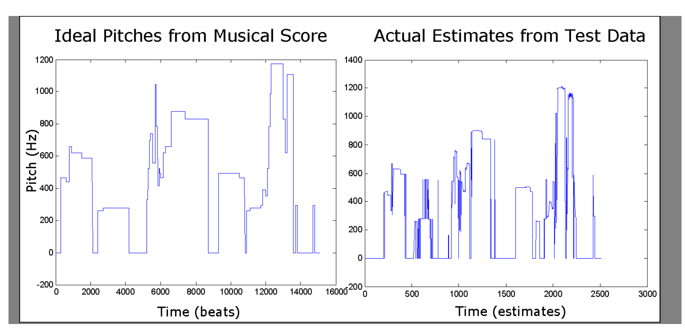



Fig. 5. Ideal pitch sequences versus actual pitch sequences: input from the Ear to the Brain.

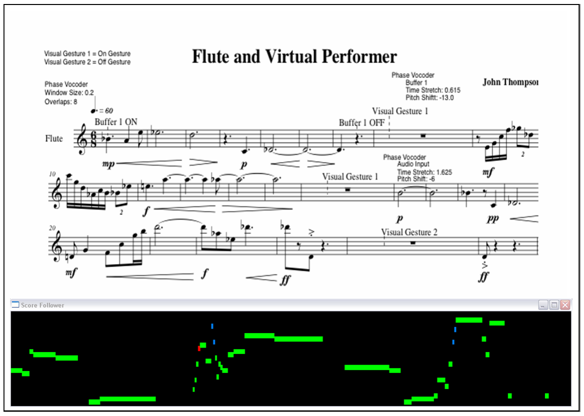



Fig. 6. Musical score for an example piece with corresponding score follower visualization

Eye. Given that the flutist will be performing a specified score we have apriori knowledge about the sequence and durations of the pitches. However, performers make errors (omissions and extraneous notes) and pitch events also may not be detected by the Ear. The Brain must therefore utilize score following techniques robust to these errors in order to properly anticipate the arrival of visually-gestured cues to the virtual accompaniment.

Score following has been actively researched for over twenty years [8], [9]. The "Out of the Ether" system has the luxury of a high signal-to-noise ratio (SNR) audio signal from the flutist and assumes for the time being monophonic music. A simple algorithm using a pointer to match to expected note events while looking ahead to account for note omissions is found to be sufficient. For more complex pieces or polyphonic music, statistical approaches to score following such as Hidden Markov Model techniques[10] may be preferable. A visualization of this system's score follower has been developed. Figure 6 shows John Thompson's piece and the accompanying visualization colors each bar as a 'matched' or 'missed' note.

Additionally, pitch estimates from the Ear can be noisy. Figure 5 shows a typical plot of ideal and actual pitch estimates. A third-order averaging filter is used to smooth the input data before finding the minimum L2 distance to determine the corresponding note. Tempo can be estimated by using the ratio of onset times defined by the score to observed onsets. This tempo estimate is available to the Voice of the virtual performer for accompaniment.

The Brain integrates the audio and visual inputs using composer-defined windows in the score to allow or disallow visual gesture cues. From score following the virtual performer can be confident but not certain about proximity of a visual gesture event, so the composer-defined windows currently need to be larger than one note. Even so, the Eye is provided flexibility in receiver operating decisions to allow 'false detections' which would normally disrupt a real-time musical performance. Therefore more subtle visual gestures, if desired, are possible.

## VI. The Voice

The final section of the overall system, the Voice, is a sound generating component that is controlled via messages from the Brain. For our system, we are transforming time-domain samples into frequency-domain samples for both the practical purpose of audio feature extraction and for the artistic purpose of transforming and resynthesizing the data as the "voice" of the virtual performer. The Voice and Ear will ideally be running on the same machine since the spectral data used for analysis by the Ear will also be used by the Voice. Spectral data unfortunately requires an immense amount of bandwidth (8 or more times as much as simple waveform data) so we would like to avoid streaming the data over a network to avoid dropouts in the audio. We also do not want to compress the spectral data, since the artifacts introduced are unacceptable for a musical performance. At the moment, we are storing spectral frames in a buffer so that processes such as time-scaling and time-delaying the human performer's output are possible. Messages from the Brain can tell the Voice when to start recording spectral data into a buffer and when to play back the buffered data with spectral transformations.

## VII. Conclusion

We have created a moderately robust system that responds to a human performer based on a pre-arranged score of musical notes and physical gestures. This has provided us with a useful platform towards creating a more robust system that can better integrate features of the audio and video streams to drive processes within a score. This research also shows promise for less intrusive mechanisms for human and computer interactive computer music performance. The system works well within a distributed framework, so it leaves the option open for more demanding feature extraction and score following algorithms on multiple processors. We have also collected a small database of video recordings of a flutist performing the test piece twelve times with three different sets of gesture. This data can be used to research novel ways to integrate the audio and video data for better cue detection.

## References

[1] P. Modler, T. Myatt, "A Video System for Recognizing Gestures by Artificial Neural Networks for Expressive Musical Control," Gesture-Based Communication in Human-Computer Interaction, 5th International Gesture Workshop, Genova, Italy, April 2003, LNAI 2915, pp.541-548, 2004.

[2] Brad Vines, Marcelo M. Wanderley, Carol Krumhansl, Regina Nuzzo, and Daniel Levitin. "Performance gestures of musicians: What structural and emotional information do they convey?," In A. Camurri and G. Volpe (eds.) Gesture-Based Communication in Human-Computer Interaction - 5th International Gesture Workshop, GW 2003, Genova, Italy. Springer-Verlag, pp. 468-478, 2004.

[3] Antonio Camurri, Barbara Mazzarino, Gualtiero Volpe, "Analysis of expressive gesture: The EyesWeb expressive gesture processing library," in A. Camurri and G. Volpe (eds.) Gesture-Based Communication in Human-Computer Interaction - 5th International Gesture Workshop, GW 2003, Genova, Italy. Springer-Verlag, pp. 460-467, 2004.

[4] Mary Li, John Thompson, Mike Quinn, "An integrated system of 3D motion tracker and spatialized sound synthesizer," Interactive

Digital Multimedia IGERT Annual Research Review, University of California, Santa Barbara. pp. 34-39, 2004.
[5] A. Michael Noll, "Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum and a maximum likelihood estimate," Proceedings of the Symposium on Computer Processing in Communications, Vol. XIX, Polytechnic Press: Brooklyn, New York, pp. 779-797, 1969.
[6] M. R. Portnoff, "Time-frequency representation of digital signals and systems based on short-time Fourier analysis," IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-28(1): 55-69, 1980.
[7] J. L. Flanagan and R. M. Golden, "Phase vocoder," Bell System Technical Journal, pp. 1493-1509, 1966.
[8] R. Dannenberg, "An on-line algorithm for real-time accompaniment," Proceedings of ICMC. Paris, France, pp. 193-197, 1984.
[9] B. Vercoe, "The synthetic performer in the context of live performance," Proceedings of ICMC. Paris, France, pp. 199-200, 1984.
[10] C. Raphael, "Automatic segmentation of acoustic musical signals using hidden Markov models," IEEE Trans. Pattern Analysis of Machine Intelligence. (21:4), pp. 360-370.

**Mary Manli Li** received her B.S. degree with honor in Computer Science and minor in Applied Mathematics from San Francisco State University in 2003. She is now a third year MS/PHD student at the Department of ECE at UC Santa Barbara. Her research interests including image analysis, pattern analysis and computer vision. Her research advisor is professor B.S. Manjunath.

**Jim Kleban** begins the Ph.D. program in Electrical and Computer Engineering at the University of California at Santa Barbara in Fall 2005. He received his M.S.E.E. from Rutgers University in 2000. His current research interests include multimodal signal processing, real-time audio streaming, evolutionary algorithms for musical synthesis and speech interfaces to databases. He is an IGERT NSF fellow in digital multimedia and a member of the IEEE.

**Lance Putnam** Lance Putnam received his B.S. in Electrical and Computer Engineering from the University of Wisconsin - Madison in 2002. He is currently pursuing his Master's degree in the Media Arts and Technology at the University of California - Santa Barbara. His main areas of interest are interface design for computer composition and digital signal processing for music and audio. He has been known at times to be a practitioner of extreme programming.

**John Thompson** is a PhD candidate in the Department of Music. He is keenly interested in the synaesthetic properties of various media and works to algorithmically explore the engines of cross-disciplinary creativity. While still active in the composition of both electronic and acoustic music, he is currently investigating techniques and methods for the creation of interactive media works.

# Nuclear Segmentation of Multispectral Histopathology Images

Laura. E. Boucheron
IGERT Fellow*
Electrical and Computer Engineering

***Abstract*—The overall goal of this research is the investigation and development of techniques for higher-level (beyond pixel-level) image analysis for improvement of automated detection and identification of cancer in multispectral histopathology images. The work summarized here represents an initial step in this process, namely segmentation of nuclei.**

***Index Terms*—Histopathology, Multispectral Imagery, Medical Image Analysis**

## I. INTRODUCTION

CURRENT studies in cancer classification have demonstrated the utility of morphometry [1], texture analysis [1, 2], and graph-theoretic methods [3–6]. While these metrics have proven useful in the study of cancerous tissue, there is a lack of unifying structure to these analyses as well as a lack of automation in many cases. It is an overall automated approach to high-level medical image analysis that is sought in this project.

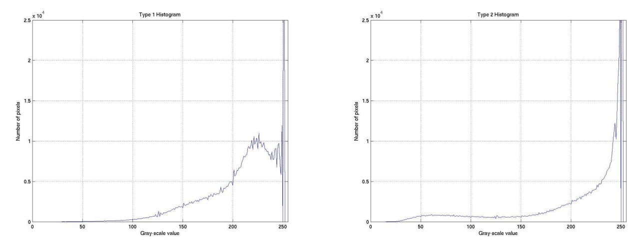
The dataset used for this research is Hematoxylin and Eosin (H&E) stained breast biopsy sections, courtesy of the Department of Pathology at Yale University. H&E staining is a common technique to elucidate pathological diagnoses from tissue sections; it stains genetic material blue, and membranes and fibers pink [7].

The imaging of these tissue sections was performed by the Pathology Department at Yale University. Each image in this multispectral dataset contains 29 bands, 10 nm between bands, covering the visible spectrum (420 nm – 700 nm).

## II. NUCLEAR SEGMENTATION

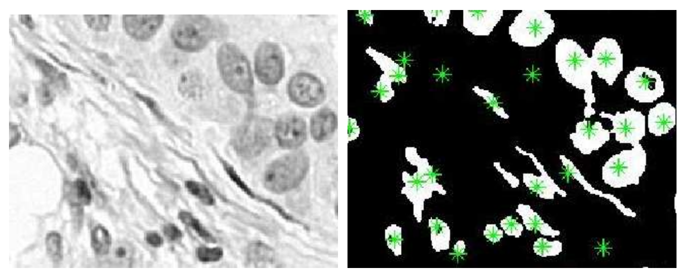
Important for much of the higher-level feature analysis described earlier is the segmentation of cell nuclei. Many published studies use a simple histogram-based thresholding followed by morphological operations [1–3, 6, 8, 9]. These methods are easy to implement and computationally efficient, but do not incorporate (much) spatial information; more complex segmentation methods may provide better results.

Certain of the multispectral bands display better contrast between nuclei and the background. Empirically, band 23 (640 nm) was chosen to be the "best" band, in that it had reasonably dark nuclei with a minimum of dark extracellular componenents. Further analysis of band 23 indicates that image histograms of this band display one




(a) Type 1 Histogram (b) Type 2 Histogram

Fig. 1. The two types of band 23 histograms



(a) Band 23 (b) Thresholded image

Fig. 2. Result of thresholding segmentation. The asterisks in (b) correspond to a manual definition of all nuclei centers.

of two structures as shown in Figure 1. Type 1 histograms (Figure 1(a)) display a steady positive slope followed by a sudden upturn. Empirically, the optimal threshold for these images is just below the knee of the histogram. Type 2 histograms (Figure 1(b)) have an approximately flat slope before a more gradual knee; the optimal threshold for these images is in the middle of the flat portion.

Threshold determination for images follows the basic observations of Type 1 and 2 histograms noted above (space does not permit more detailed description of the algorithm). After thresholding, the segmented image is cleaned up with basic morphological operations. An example result can be seen in Figure 2. It is important to note that a threshold is determined for each individual image; histogram thresholding methods mentioned previously use one threshold for all images in the dataset.

The dataset used in this research contains 58 multispectral images, classified by Yale pathologists as malignant (26 images) or benign (32 images). Overall segmentation re-

*This work was funded by Los Alamos National Laboratory and completed there as a graduate research assistant. Partial support provided by IGERT Grant DGE-0221713.

sults can be seen in Table I. False negatives refer to nuclei that were missed by the segmentation algorithm, and false positives refer to segmented blobs not corresponding to nuclei. Preliminary results indicate that this method may generalize to other H&E stained multispectral imagery.

TABLE I
OVERALL SEGMENTATION RESULTS.

| Dataset | Detection Rate | False Negatives | False Positives |
|---|---|---|---|
| Malignant | 88.8% | 11.27% | 4.55% |
| Benign | 94.5% | 2.44% | 3.25% |

## III. WORK IN PROGRESS

Currently being investigated are improvements in the threshold determination algorithm to yield more robust and consistent behavior. Also, further anlaysis must be performed on the segmentation results to verify optimal parameter values (a few heuristically chosen constants), and to compare this algorithm to others.

The second main area of work in progress deals with the need to investigate the utility of multispectral data in medical image analysis. Several recent studies have demonstrated the diagnostic capability of spectral information independent of spatial information [10–12]. In particular, there is a need to define optimal number, width, and resolution of spectral bands and to incorporate spatial support of the spectral features.

GENetic Imagery Exploitation (GENIE) (Los Alamos National Laboratory, Los Alamos, NM) is a software system designed to evolve feature extraction algorithms from multispectral imagery. By looking (statistically) at the data planes chosen in GENIE solutions, one can get an idea of the importance of individual data planes. For more information on GENIE, see Ref. [13]. GENIE runs are currently in progress to gather such data about the relative importance of the different spectral bands for the segmentation of nuclei.

## IV. FUTURE WORK

It would be possible to spend an entire thesis investigating nuclei segmentation techniques, and while the segmentation aspects will remain an integral part of this research, there is a need for research on higher-level analysis, namely the modeling of feature distributions in cancer imagery. This seems to be an area of research that would be of great interest to those in cancer research, but is rather lacking in publications.

Future research will look to take this medical image analysis to the object level: to look at imagery in terms of objects (e.g. nuclei, cytoplasm) instead of analysis based on individual pixels and simple pixel neighborhoods. Analysis will focus on breast cancer, with imagery coming from multiple imaging modalities (H&E light microscopy, immunostaining, and fluorescent imagery).

## V. ACKNOWLEDGEMENTS

Many thanks to the Department of Pathology at Yale University, particularly Cesar Angletti and David Rimm, for the images used in this study.

## REFERENCES

[1] B. Weyn, G. van de Wouwer, M. Koprowski, A. van Daele, K. Dhaene, P. Scheunders, W. Jacob, and E. van Marck, "Value of morphometry, texture analysis, densitometry, and histometry in the differential diagnosis and prognosis of malignant mesothelioma," *J Pathol*, vol. 189, pp. 581–589, 1999.

[2] G. van de Wouwer, B. Weyn, P. Scheunders, W. Jacob, and D. van Dyck, "Wavelets as chromatin texture descriptors for the automated identification of neoplastic nuclei," *J Microsc*, vol. 197, pp. 25–35, 2000.

[3] B. Weyn, G. van de Wouwer, S. Kumar-Singh, A. van Daele, P. Scheunders, E. van Marck, and W. Jacob, "Computer-assisted differential diagnosis of malignant mesothelioma based on syntactic structure analysis," *Cytometry*, vol. 35, pp. 23–29, 1999.

[4] J. T. G. Smith, G. D. Lange, and W. B. Marks, "Fractal methods and results in cellular morphology – dimensions, lacunarity, and multifractals," *J Neurosci Methods*, vol. 69, pp. 123–136, 1996.

[5] B. Weyn, W. A. A. Tjalma, P. Vermeylen, A. van Daele, E. van Marck, and W. Jacob, "Determination of tumour prognosis based on angiogensis-related vascular patterns measured by fractal and syntactic structure analysis," *Clin Oncol*, vol. 16, pp. 307–316, 2004.

[6] J. W. B. Spillman, J. L. Robertson, W. R. Huckle, B. S. Govindan, and K. E. Meissner, "Complexity, fractals, disease time, and cancer," *Phys Rev E*, vol. 70, pp. 1–12, 2004.

[7] Michael H. Ross, Gordon I. Kaye, and Wojciech Pawlina, *Histology, A Text and Atlas with Cell and Molecular Biology*, Lippincott Williams&Wilkins, 4th edition, 2003.

[8] J.-P. Thiran and B. Macq, "Morphological feature extraction for the classification of digital images of cancerous tissues," *IEEE T Bio-Med Eng*, vol. 43, pp. 1011–1020, 1996.

[9] M. Sammouda, R. Sammouda, N. Niki, N. Yamaguchi, and N. Moriyama, "Cancerous nuclei detection on digitized pathological lung color images," *J Biomed Inform*, vol. 35, pp. 92–98, 2002.

[10] R. Jaganath, C. Angletti, R. Levenson, and D. L. Rimm, "Diagnostic classification of urothelial cells in urine cytology specimens using exclusively spectral information," *Cancer*, vol. 102, pp. 186–191, 2004.

[11] R. A. Shaw, J. R. Mansfield, S. P. Rempel, S. Low-Ying, V. V Kupriyanov, and H. H. Mantsch, "Analysis of biomedical spectra and images: From data to diagnosis," *J. Mol. Struct. (THEOCHEM)*, vol. 500, pp. 129–138, 2000.

[12] I. Barshack, J. Kopolovic, Z. Malik, and C. Rothmann, "Spectral morphometric characterization of breast carcinoma cells," *Br. J. Cancer*, vol. 79, pp. 1613–1619, 1999.

[13] N. R. Harvey, J. Theiler, S. P. Brumby, S. Perkins, J. J. Szymanski, J. J. Bloch, R. B. Porter, M. Galassi, and A. C. Young, "Comparison of genie and conventional supervised classifiers for multispectral image feature extraction," *IEEE T Geosci Remote*, vol. 40, pp. 393–404, 2002.

**Laura E. Boucheron** received her B.S. and M.S. in Electrical Engineering from New Mexico State University, Las Cruces, NM in 2001 and 2003, respectively. Her research interests include image processing in general, specifically image understanding, object-level image analysis, and medical image analysis. She is currently pursuing her Ph.D. in Electrical and Computer Engineering at UCSB. Her academic advisor is Dr. B. S. Manjunath.

# Biometric Techniques for Classifying Pianists (Summer 2005)

MarkDavid Hosale
IGERT Associate
Media Art and Technology

Bhaskar Rao
IGERT Associate
Media Art and Technology

Justin Muncaster
IGERT Fellow
Computer Science

*Abstract*— **This research meets in the middle of staff notation and biometrics. We wish to develop techniques to identify pianists based on stylistic features extracted as they play a piece. As stylistic features greatly resemble the features one find's in a keystroke dynamics problem, work in one area benefits work in the other. We wish to have a system that can classify pianists in order to allow an apprentice pianist to see how closely they match the style of their mentor as well as provide algorithms to exploit the potential of keystroke dynamics as a continuous biometric.**

***Index Terms*—Biometrics, Keystroke Dynamics, MIDI, Score Following**

## I. Introduction

A musical score is not a literal inscription of the music it is intended to represent. Behind every musical score there is an unsaid expectation of the manner in which a performer will play the work. The well-trained musician brings to the score years of experience as an interpreter of the music they are playing. The issue of musical score interpretation falls under the vastly studied area in music known as performance practice.

It is because of these hidden performance practices that the computer has a difficult time interpreting musical scores based solely on the information present in the score document. The computer lacks all of the information needed to allow it to be an adequate tutor for a musical instrument, such as the piano. One solution to this problem is to provide the computer with a model of an interpretation of a score in a particular genre, which the computer could use as a basis for an evaluation.

By analyzing information gathered in the performance of a piece by an expert performer it is potentially possible to use this information as a model by which practicing performers will be compared.

## II. Classifying Individuals

The field of biometrics attempts to identify individuals using biological or behavioral characteristics [8]. Biological characteristics include fingerprints, retinal patterns, and facial features. Behavioral characteristics can range from the way a person walks (referred to as gait) to the way a person types. Behavioral characteristics are useful biometrics because they are usually difficult to fake and non-intrusive to measure. Keystroke dynamics, or the analysis of one's typing habits, is a particularly useful behavioral characteristic. It is inexpensive to measure, imperceptible to the subject, and continuous in the sense that measurements are available so long as the user is typing.

Classifying users of a computer keyboard is in many ways a similar problem to classifying pianists. In both cases the system is supposed to process a stream of events indicating the state of a key and in both cases features can be extracted based on the timing of these events. As keystroke dynamics has been shown to be a useful biometric [8], we use techniques there to develop a system that can classify pianists.

## III. Application Area

Our application is intended to assist a novice in learning to play more like an expert musician by matching their performance to a model of how the same score might be interpreted by an expert musician

Our application is intended to augment one on one instruction so that students can receive training during their own practice sessions. One of the hardest things that novice musicians have to overcome is practicing their mistakes. A system like we're proposing can help reduce those mistakes by helping to improve a student's practice habits

Like a continuous biometric, our system is designed to work in real time. Students will receive feedback as they are playing and can adjust their performance to better match the model in the system.

There are three components to the system, comprised of two data sets, an expert data set and a student data set, and the keystroke dynamic analysis system. First, an expert muscian plays into the keystroke dynamic analysis system, where their performance is analysed and stored. After a substantial database of music has been entered into the system it is ready to evaluate the performance of student pianists. Then, when a student plays a piece into the system it is evaluated against the 'Expert' data set and the results are fed-back to the student.

## IV. Architecture Overview

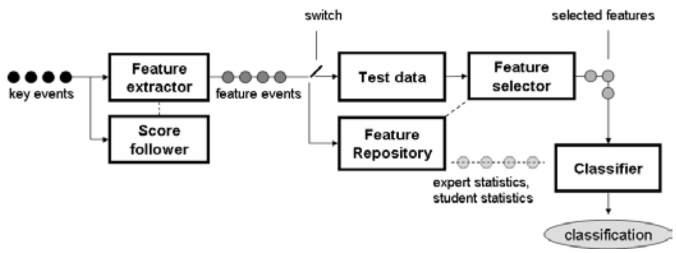



**Figure 1. This figure shows an overview of the system architecture. "Key events" are analyzed by a feature extractor which performs various measurements based upon where the user is in the score. During training, these events are labeled and sent to the feature repository. During testing, the features go to the test data and then particular features are selected and sent to the appropriate classifier.**

### A. Feature Extraction

Standard MIDI data comes serially in small packets, or "events", that correspond to key presses and key releases. Key events contain information about which key was pressed, how hard the key was pressed, and how much time has passed since the previous event. In our system, the first place these events go to is the feature extractor and score follower, as denoted in Fig. 1.

The feature extractor's purpose is to take measurements on the data coming in based on where the performer is in the score. To achieve this, the feature extractor consults a score follower which attempts to keep track of where in the score the performer is. Although the score follower is essentially a "black box" in our diagram and can be any suitable algorithm, our particular score follower aligns the pressed notes with their corresponding notes in the score using a dynamic programming algorithm.

Once the extractor knows which note in the score the performer is intending to play it is able to determine precisely how much the performer varied from the score. Variations from the score include such measurements as a key being held down for too long or too short (relative to the current tempo) or a key being pressed too hard or too soft. These measurements are taken as the user plays and pushed through the pipeline as feature events.

### B. Directing Feature Events and Feature Selection

If the system is in training mode then feature events are labeled and pushed into a database called a feature repository. The repository holds all of the training data for the classifiers and can be queried by both the classifier and feature selectors.

If the system is in testing mode, the data is sent into the test data. The feature selector then monitors the test data and when the test data meets the criterion for selection, it is taken. The criterion for selection can vary depending on the training data available or the classifier being used.

For example, a particular classification technique may only work well if there is a lot of test data and a lot of training data for a given feature. Thus, the feature selector would buffer feature events until such criterion is met. Finally, the selected features are sent to the classifier.

## V. Classification Techniques

We have formulated a two-class classification problem. Based on input feature data the classifier should tell us if the performer is playing like a novice or like an expert. This classifier will be trained on feature data from an expert playing the same piece. Looking at the feature space and after some initial data exploration we adopted a multiple classifier approach to the problem similar to our approach to keystroke dynamics.

Different classifiers have different strengths and weaknesses based on amount of training data available and the characteristics of the feature data. We divide the feature space based on this knowledge to get good classification results. We use a Naïve Bayes classifier, a Nearest Neighbor classifier and a Support Vector Machine. Taking a weighted output of the continuous classification results from the three classifiers we estimate how well the pianist is playing.

Naïve Bayes Classifier works very well for those features that have enough instances in the training data for the validity of the Gaussian assumption and sufficient statistics for estimating the mean and the variance of the distribution. Nearest neighbor works on distance metric proximity assumption and does not try to estimate the distribution. Support Vector machines (SVM) find the optimal classifying non-linear hyper-plane. SVMs and Nearest neighbor methods generally do not require as much training data as Bayes classifier and do not make any assumptions on the underlying probability distributions of the training data or distance metrics.

## V.Conclusion

MIDI timing data from a pianist's performance of a piece can provide us valuable information on a pianist's style. Our system does a score following and translates MIDI messages into a format of keystroke timings for a Keystroke Biometric system. The keystroke analysis system using labeled training set is able to distinguish between an expert and a novice pianist in real-time, therefore act as a piano tutor.

## Acknowledgment

Partial support provided by NSF IGERT Grant# DGE-0221713

## References

[1] Bloch, J.B. and Dannenberg, R.B., "Real-Time Computer Accompaniment of Keyboard Performances," in *Proceedings of the 1985 International Computer Music Conference*, (1985),

International Computer Music Association, 279-289. http://www.cs.cmu.edu/~rdb/bib-accomp.html#icmc85.

[2] Cano, P., Bonada, J., and Loscos, A., "Score Performance Matching Using HMMs," in *Proceedings of the 1999 International Computer Music Conference*, (1999), International Computer Music Association

[3] Dannenberg, R.B., and H. Mukiano, "New Techniques for Enhanced Quality of Computer Accompaniment," in *Proceedings of the 1988 International Computer Music Conference*, (1988), International Computer Music Association, 243-249.

[4] Leggett J, Williams G, Usnick M, Longnecker M. *Dynamic identity verification via keystroke characteristics*. International Journal of Man Machine Studies 1991;35:859e70.

[5] Obaidat M, Sadoun S. , *Verification of computer users using keystroke dynamics.* IEEE Transactions on System Man Cybernet Part B 1997;27(2):261e9.

[6] Yu E, Cho S. Novelty detection approach for keystroke dynamics identity verification. Fourth International Conference on Intelligent Data Engineering and Automated Learning IDEAL 2003 Hong Kong, China : 2003

[7] Schölkopf B, Platt J, Shawe-Taylor J, Smola AJ, Williamson RC. *Estimating the support of a high-dimensional distribution.* Technical Report MSR-TR-99-87. Microsoft Research, Redmond, WA; 1999.

[8] A. Jain, R. Bolle, S. Pankanti, "Biometrics: Personal Identification in a Networked Society", Kluwer Academic Publishers, 1999.

**MarkDavid Hosale's** interests are indisciplinary, but the connecting tissue comes from thinking of art and music as a narrative. Particularly, the kinds of narrative that are structured using non-linear representations of information, time and space. He received his B.A. in Music from the College of Creative Studies at UCSB in 1999, and a Masters of Music in Music Composition and Theory from the University of Illinois at Urbana-Champaign in 2001. Currently, he is pursuing a Ph.D. in Media Art and Technology at University of California in Santa Barbara. His IGERT advisors are Marcos Novak and Stephen Travis Pope.

**Bhaskar Rao's** interests are in pattern recognition especially applied to Multimedia data and biometrics. He received his Bachelor of Engineering degree in Computer Science from the Vishveswaraiah Technological University in Bangalore, India in 2002, and a Master of Science in Media Arts and Technology at the University of California, Santa Barbara in 2005. His IGERT advisor is Matthew Turk.

**Justin Muncaster** is a Ph.D. student in the College of Engineering at the University of California. His interests lie in multimodal data fusion with an emphasis on probabilistic inference. He currently is working on using multiple weak biometrics to classify individuals. His advisor is Matthew Turk.

# Spheres of Influence

August Black, Thomas Kuo, Marko Peljhan, Mike Quinn, Brian Springer, Jason Wither

***Abstract*—Spheres of Influence is a multimedia installation at UCSB's Davidson Library in which a person uses the position of their body to manipulate still satellite images and international TV news programs.**

***Index Terms*—parallel graphics, motion detection, computer vision, interaction, aesthetics.**

## I. Introduction

"Spheres of Influence" is a collaborative project to produce an interactive screen-based installation at UCSB's Davidson Library. The initial prototype consists of 4 large plasma displays, a graphical visualization including segments of global news footage, and a multi-camera position tracking system for user input. In combination, the system as a whole proposes a novel "touchless" physical interface to multidimensional audio/visual information.

Fig. 1. Spheres of Influence interface.

The display screens are mounted side by side to create a wide panoramic view with a wing span of over 4.2 meters. This distance is significant to the proposed location and use of the installation, providing ample room for natural movement in the library corridor. Within this 4.2 meter breadth in front of the plasma screens, the motion and position tracking system captures the movements of the viewers and delivers this information to the visualization application. The visualization application consists of a single viewing screen rendered seamlessly over all 4 horizontal displays using the server-client structure of the OpenGL industry standard and the Chromium parallel graphics engine [5]. It shows abstract cut-outs of satellite photography of the earth, which are animated and can be interactively selected, zoomed, and scrolled on screen based on the 2 dimensional user input - their position (x,y) in the space. Additionally, video sources are compiled from a live satellite source as well as from the SCOLA[1] archive of news footage from around the globe [8]. As a user zooms in on a selected region of satellite imagery, a video from that area is spawned on screen for the viewer to see. Each of the cut-out segments of satellite photography are based on one of the 70 countries in the SCOLA TV news archive, and all TV news clips are in the native language of the clip's originating country . In total, over 30 different languages are represented in the project. However, the dominant language of global commerce - English - is not represented.

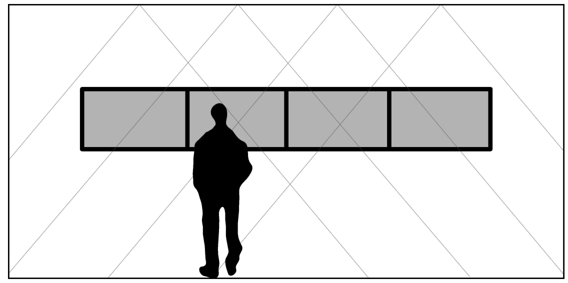


Fig. 2. Camera tracking using 4 cameras mounted on the ceiling with overlapping views.

## II. Tracking System

The tracking system consists of 4 IEEE 1394 cameras, each attached to a computer, and a central tracking computer. All computers run Ubuntu linux. Capture is performed with the dc1394 libraries [9]. Video processing is performed using Intel's OpenCV libraries [2].

The cameras are mounted on the ceiling along a line which is approximately 170cm from the displays. Figure 2 shows an illustration of the camera setup. The overhead placement was chosen because of the size limitations of the space. This arrangement minimizes the possibility of occlusion of the subjects in the space and provides accurate location in the horizontal plane.

By utilizing a running Gaussian average background model [7], each camera locates the subjects in its view and then creates a feature vector for each. This feature vector consists of a timestamp, an identification number, x and y positions, area, and average R, G, and B values. These features are used to maintain consistent identification of subjects between frames.

Once identification is complete, the subjects' locations are translated to real-world coordinates using camera parameters which are calculated offline [3]. The list of subjects is then transmitted via the OSC (Open Sound Control) protocol to the central tracker [1].

Because of overlap between neighboring cameras, the central tracker locates and eliminates duplicate subjects. Once any issues are resolved, the final list of subjects is transmitted to the display system, again via OSC as illustrated in Figure 3.

## III. Parallel Graphics Engine

To display the application over four displays we are using the open source Chromium package [5]. Chromium intercepts OpenGL calls before they are sent to the graphics card and is able to then modify them as specified. The Stream Processing Unit (SPU) that we are using is a tiling routine that distributes the geometry and textures from the application to the appropriate machine for display. Using this we are able to transparently

[1] SCOLA is a non-profit educational organization that receives and retransmits television programming from around the world in native languages.

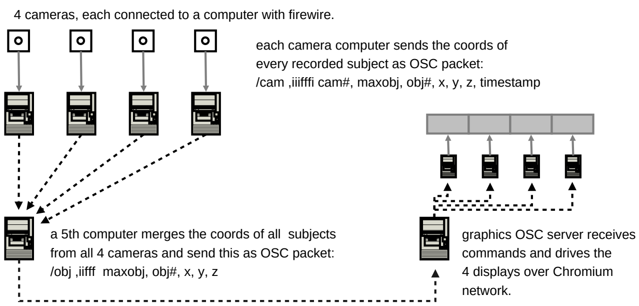



Fig. 3. Spheres of Influence network.

Fig. 4. Spheres of Influence graphical interface: zoom mode.

run our OpenGL application over four displays without modifying the original code.

One of the most important aspects of the display system is the satellite map imagery. We are using 1 km per pixel resolution data, which means that our entire dataset is close to two gigabytes. To display both highly detailed and localized map data, as well as larger areas we use a pyramidal tiling system. For this we've broken up the map data into several layers of resolution with small tiles of the image at each resolution. We can then switch between layers to display different amounts of detail. We also intelligently pre-load tiles into memory to avoid any slow-down from large amounts of texture loading when switching between levels.

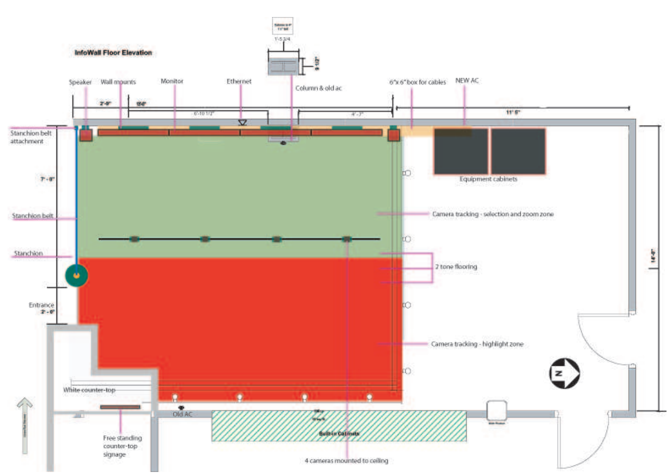



Fig. 5. Diagram of floor layout.

Our interface is designed to be simple and intuitive enough for visitors to use without any training or advanced knowledge. We have separated our space into two halves as shown in the floor diagram of Figure 5. The back half can be thought of as the highlight space where users can highlight areas of interest. In this half of the space the users position controls the left to right speed of the map strips as they pan across the screen. The closer the user moves to the middle of the space, from the back the slower the maps will pan. In this view the map directly in front of the user are highlighted for selection. As the user crosses the front / back halfway line they enter the enhanced detail viewing space. In this space the rest of the non-highlighted strips become night-time views, while the selected strip zooms in further towards the center of broadcast associated with that strip as the user walks forward. As the user continues to walk forward a video from that broadcast appears, as well as metadata associated with the currently showing programming. To exit the enhanced detail mode the user simply needs to recross the center line to re-enter the highlighting mode.

## IV. Video Analysis

The SCOLA video archive is downloaded daily at night and batch processed, searching for handheld camera segments in the video, which we feel more closely represents scenes of action and interest. To extract handheld camera footage from a news program requires two steps, shot detection and camera motion detection. We assume that if a camera is moving during a shot, then the shot is from a handheld camera. An additional motion analysis step is added to rearrange the shots from most to least activity. Shot boundaries are detected based on the output of a Canny edge detector for adjacent frames of the video. From these frames, we get two ratios that describe the number of incoming edges by computing the dot product of the first edge image and a dilated version second edge image. The dilation provides a margin of error for the for small shifts of the edges. A similar ratio is computed for the outgoing edges. The maximum of these ratios is denoted the edge change ratio. A sudden spike in the edge change ratios is a detected as a hard cut, a gradual maximum is a cross-fade. [10] Simultaneous, for each pair of frames we compute the Horn-Schunck Optical Flow between them. [4] If practically all pixels have an optical flow value, then the camera is moving between those frames. Throughout a shot, if enough frames are moving, then the entire shot is classified as a moving camera shot. In order to pick out the most interesting of these shots, we decided to rearrange the shots in order of most motion activity as described by the MPEG-7 [6] Motion Analysis Descriptor. This descriptor measures the standard deviation of the variance of the magnitude of the motion vectors inherent in MPEG videos. The system outputs an MPlayer playlist of handheld camera shots sorted from most to least motion.

## V. Conclusion

In this project, a sort-first parallel graphics architecture is implemented to display a single seamless OpenGL application over 4 large plasma screens mounted horizontally on the wall at the Davidson Library. The user interface is a physically navigable visualization of our globe composed from NASA's blue marble satellite photography and video news segments from more than 70 different non-english speaking sources worldwide. A simple, non-intrusive, multi-camera position-tracking system allows for intuitive human interaction with the visualization application whereby the user/viewer's movement in the space acts as a multi-dimensional mouse pointer into the system.

## Acknowledgment

Realization of "Spheres of Influence" for the Davidson Library at UCSB, is supported by the Davidson UCSB library, the Alexandria Library project, the NSF IGERT doctoral program, and the Media Arts & Technology graduate program. Co-production by Larry Carver, director, Alexandria Library, George Legrady, faculty, MAT, B.S.Majunath, faculty, College of Engineering and Marko Peljhan, faculty, Department of Art, with graduate students from Art, CS, ECE and MAT, and technical staff (Justin Mathena & David Valentine) of the Alexandria Library project. Additional thanks goes to Dr. Mathias Kolsch for initial help with linux development.

## References

[1] Center for New Music and Audio Technologies. Osc. http://cnmat.cnmat.berkeley.edu/OpenSoundControl.

[2] I. Corporation. Intel opencv libraries. Website. `http://www.intel.com/technology/computing/opencv/index.htm`.

[3] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[4] B. K. Horn and B. G. Schunck. Determining optical flow. In *Artificial Intelligence*, 1981.

[5] G. Humphreys. Chromium. http://chromium.sourceforge.net/.

[6] B. Manjunath, P. Salembier, and T. Sikora. *Introduction to MPEG 7: Multimedia Content Description Interface*. Wiley: Berlin, New York, 2002.

[7] M. Piccardi. Background subtraction techniques: A review. In *Proceedings, IEEE International Conference on Systems, Man, and Cybernetics*, volume 4, pages 3099–3104, October 2004.

[8] SCOLA. Scola web address. http://scola.org/.

[9] Various. Linux1394.org. Website. `http://www.1394.org`.

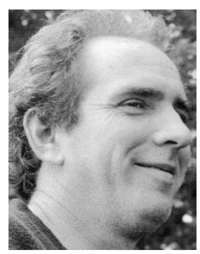
[10] R. Zabih, J. Miller, and K. Ma. A feature-based algorithm for detecting and classifying production effects. In *Multimedia Systems*, 1999.

**August Black** is an artist and FLOSS developer currently enrolled in the Media Arts & Technology program at UC Santa Barbara. His research interests lie in the very broad overlap of creativity and code, where computational methods apply to the production of radio, film, and performance works. http://aug.ment.org

**Thomas Kuo** is an intended Ph.D. student in the ECE Department at UC Santa Barbara. He received his B.S. in electrical engineering at UC Berkeley in 2004. His research interests are video and image processing and computer vision.

**Marko Peljhan** Studied theatre and radio directing at the University of Ljubljana and in 1992 founded the arts organisation Projekt Atol in the frame of which he works in the performance, visual arts, situation and communications fields. In 1995 he founded the technological branch of Projekt Atol PACT SYSTEMS and in 1999 he founded Projekt Atol Flight Operations. In 1995 he co-founded LJUDMILA (Ljubljana Digital Media Lab). He is coordinator of the international INSULAR TECHNOLOGIES initiative and the Makrolab project as well as flight director of the parabolic art and science flights with the Yuri Gagarin Cosmonaut Training Centre in Moscow within the MIR (Microgravity Interdisciplinary Research) initiative. He is currently associate professor in Art and Media Arts and Technology interdisciplinary studies at the University of California Santa Barbara.

**Michael J. Quinn** is currently a Ph.D. student in the ECE department of UC Santa Barbara. He received the B.S. and M.S. degrees in Electrical Engineering from the University of Missouri - Rolla in 1997 and 1999, respectively. His research interests include tracking, surveillance, camera sensor networks, and super resolution. His academic advisor is Professor B.S. Manjunath.



**Brian Springer** received his undergraduate degree in Media Studies at the State University of New York at Buffalo; in Buffalo worked with other artists to cofound a number of community organizations supporting the production of media art and public-access cable. Currently Brian is a graduate student in UCSB's Art Studio program.

**Jason Wither** is currently a Ph.D. student in the Computer Science department at UC Santa Barbara. He received his B.S. in Computer Science from Harvey Mudd College in 2002, and his M.S. also in Computer Science from UC Santa Barbara in 2005. His research interests include Augmented Reality, Human Computer Interaction, Computer Graphics, and Computer Vision. His academic advisor is Professor Tobias Hollerer.

# Multi-Domain GeoVisualization of News Stories

Julie Dillemuth
IGERT Trainee
Dept of Geography

Alex Villacorta
IGERT Trainee
Dept of Statistics

***Abstract*—The objective of this project is to analyze the geographical and temporal movement of information, specifically headline news stories in major US newspapers, from a central point of origin (the newspaper headquarters) to dissemination among the papers' readership, and relate this movement with the location of the news events in space and time. A mathematical model will be developed from a statistical analysis of newspaper data from a specific time period, and a dynamic visualization of the movement will be created to communicate the results, based on geovisualization theory and techniques.**

***Index Terms*—epidemiological modeling, geovisualization, newspaper stories, latent semantic indexing (LSI)**

## I. INTRODUCTION

HOW does information, specifically news stories, move around the globe from where the news happens to where the public reads or hears about it? Does it spread like disease, a story infecting the population as it is communicated through mass media or from one person to another, or can it be compared to point-source pollution, flowing outwards along particular vectors?

The goal of this project is to statistically analyze and visualize the movement of news stories over time as they are reported in U.S. newspapers. Traditionally, the idea of information flow pertains to the evolution of information from its inception to archival where its geographical path is of little significance. In this project we aim to consider both the temporal and geographical aspects of information flow.

Newspaper articles, unlike television, radio or Internet news sources, have been widely archived in digital databases (such as America's Newspapers by Newsbank, Inc.) and are published and circulated in specific geographical locations. By analyzing the statistical and geographical features of news stories from origin to archival and the dissemination of the papers to a readership it is hoped that a mathematical model may be created which helps to explain the spatial properties of communication channels in the US. Geovisualization theory and techniques provide the foundation for data exploration and communication through dynamic maps and graphics.

Four phases comprise this project, starting with a test-case single story tracked on a state level. Phase 2 moves to a nationwide comparison of front page news distribution among 15 papers with the largest circulation in the US, from a month-long sample of days over 3 different years. News spread will be modeled based on a modified epidemiological model. Phase 3 will then compare the front-page stories with all the stories available on the Associated Press news wire, to assess characteristics of the selection of news that appears on the front page of major papers. Finally, Phase 4 is concerned with automating our process, creating a broadly applicable tool for semantic, spatial, temporal data integration and analysis.

## II. Phase I

For Phase 1, articles on a particular story (the Terri Schiavo right-to-die case) were selected through an America's Newspapers database query (http://infoweb.newsbank.com) and parsed into a table with headline, newspaper, date, publication location, and word count. We used a geographic information system (GIS) to translate the table into a series of maps at different time steps, indicating the publication of stories at a city level. Macromedia Flash software animated the map sequence to convey the temporal characteristics of the news articles. This case-study analysis provided insight as to data availability, processing requirements, visualization design issues, and how to realize a visualization in Flash software, before we delved into larger and more complex datasets. The semantic layer, or the article content, was not integrated at this stage; any article "infected" by the words "Terri (or Terry) Schiavo" was mapped. Indeed, many articles were only tangentially or marginally related to the Schiavo story, making it clear that a semantic analysis would play an important role in filtering the articles for relevance.

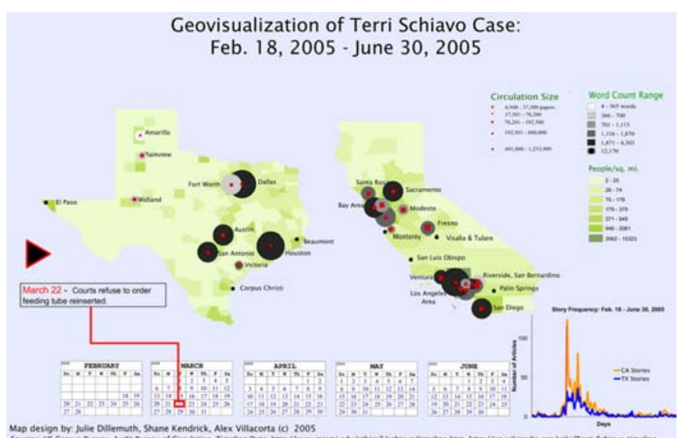



Fig. 1. Screenshot of interactive animated map

Figure 1 is a snapshot of one visualization, showing the "Terri Shiavo' articles for one day in California and Texas, with population density in the background and circulation size and word count represented in the symbology. Figure 2 is an alternate visualization, reducing the number of variables portrayed to focus on the spread of the story over 3-day windows. Visualizing the presence/absence of these articles through time with multiple representation types indicated that newspapers in the largest cities tend to pick up a story first, followed by smaller papers in nearby cities and towns. Also, comparing story activity with a time line of events in the Shiavo case revealed a clear point at which the story switched from local to national importance, and later when it no longer captured the attention of the press.

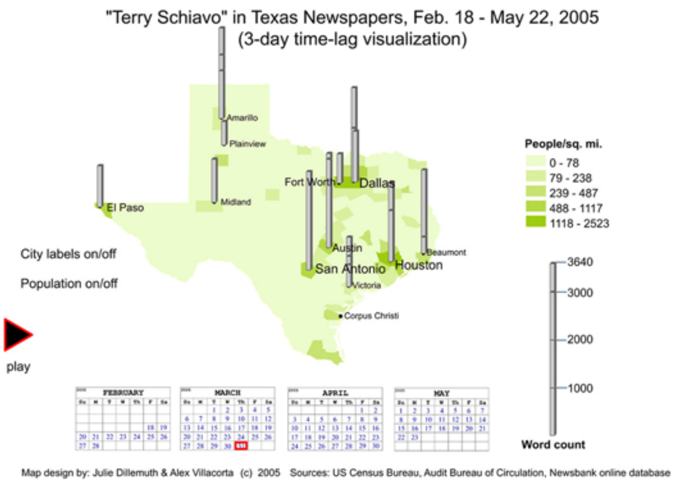



Fig. 2. Alternate design of interactive animated map

## III. Phase II

Phase 2 is the current focus of research, and concentrates on space, time and semantics for national-scale front-page news in major newspapers. Figure 3 shows the coverage of the 15 papers under consideration. Only those stories appearing in three or more papers, with five or more total articles, were tracked in the analysis. Space is considered primarily in terms of eastern versus western US papers, as the sample size is too small to warrant further regional subdivision. For the temporal domain, we consider which papers are first to report on a story and which papers are slower to respond. Story duration combined with the number of articles on that story and associated word count can be considered a measure of story 'importance'. For the story content, latent semantic indexing (LSI) was used to measure article similarity, with the objective of identifying how relevant a article is to a main topic and allowing us to track how a news story evolves over time. The dataset proved too complex for clean LSI results, so the data was classified by the researchers. As a consequence, this dataset can now be used as a ground truth to evaluate LSI processing and clustering methods as the parameters of the algorithms are tweaked to improve the performance.

Initial results from this analysis suggest that there are significant differences in the distribution of international stories in eastern versus western newspapers, but that there is no evidence of differences for domestic stories. Additionally we found that in our sample of papers with circulation sizes of $300,000 - 1,100,000$ there was no significant difference in the average number of articles carried on a given national-scale story. However, the total number of national-scale stories covered by each paper did correlate with the circulation size. For instance, in 2001 larger papers covered as many as 80% of all stories and the smaller papers covered as few as 15%.

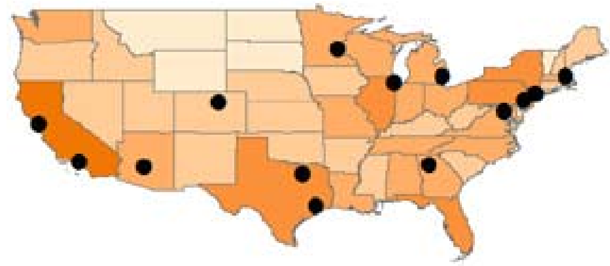


Fig. 3. Geographical Coverage of top 15 most circulated papers on a background of population density. Darker shades correspond to more densely populated states.

## IV. Epidemiology Models

In efforts to better explain the dynamics of news spread in the population of newspapers, we have adopted classical epidemiology models of disease spread and applied them to our national front page news data. Because of the discrete nature of printed newspaper articles, we have chosen the household binomial chain model, also known as Reed-Frost model. In this model, the population is assumed homogeneous and consists of households of equal size, $s$, with three possible member types: susceptible, infectious, and recovered. Further, it is assumed that every susceptible member in a household is equally likely to be infected from each infected individual during each unit of time $t$ with probability $p_t$. These probabilities of infection, $p_t \quad t = 1, \ldots, n$, characterize the model and are commonly taken to be constant, $p_t = p_{(t+1)}$ (Greenwood) or exponential, $p_t = p^t$ (Reed-Frost). Under this setup, for a given time period $t$ (taken as one day in our case), the probability that there are $x$ household members infected in time period (t+1), given that there are $S_t$ susceptible and $I_t$ infected is governed by a binomial distribution with parameter $p_t$. That is,

$$P(I_{(t+1)} = x | S_t = s, I_t = i) = \frac{s!}{x!(s-x)!} p_i^x q_i^{(s-x)}$$

where $x = 0, 1, \ldots, s$.

Finally, to model an outbreak of a particular disease, we focus on a sequence, or chain of infected individuals over time. The probability of this chain of infections is simply given by the product of the binomial terms for each of the corresponding time periods. For instance, suppose that for a particular household with 4 members including 1 infected individual, the sequence of infected members is $1 \to 1 \to 0$, meaning in time frame 1 (day 1) there was one member infected, on day 2 there was another member infected and day three there were no new infections. Using the binomial chain model, the probability of this event happening for a given $p_1$ ($q_1 = 1 - p_1$) is:

$$\begin{aligned} P(1 \to 1 \to 0) &= \left( \frac{3!}{1!(3-1)!} p_1^1 q_1^{(3-1)} \right) \left( \frac{2!}{0!2!} p_1^0 q_1^2 \right) \\ &= 3 p_1 q_1^4 \end{aligned}$$

In our current study we consider each newspaper as an individual household and the total number of articles available for front page news per day as the number of household members. With our observed data we have realizations of various types of chains corresponding to various news topics. The goal is to find estimates for all of the $p_i$ so that the spread of other news types may be predicted and analyzed. Clearly, specific news articles are not living entities which transmit or recover from diseases, nor do newspapers necessarily follow the assumptions of the chain binomial model. However, the overall presence and decay in the number of news articles of a particular story initially seem to very closely match those of an epidemic outbreak and warrant further investigation into this hypothesis.

## V. Conclusion and Outlook

While this research starts on a small scale of comparing the spread of an individual news stories within specific states, then moves to national-scale stories in papers across the country, the ultimate goal is to be able to analyze and visualize any dataset of news stories, from the scale of a city or urban area, to the nation as a whole, and potentially worldwide. In addition, there are several broader implications of this project. Contributing to a better understanding of the flow of information through a model will help to identify both strong and weak channels of information transfer. Such knowledge would be useful in preparing against attacks on communication systems within a region. Also, emergency response systems need to know which communication channels reach the most people in the shortest amount of time. Lastly, it is hoped that this project could lead to a determination of what effect, if any, the flow of information has on cross-cultural understanding and geopolitical relationships.

Julie Dillemuth is working towards a PhD in Geography, after receiving her Masters degree at UCSB in June 2005. Her faculty advisor is Dr. Keith C. Clarke and her primary research interest is cartography for small, mobile devices.

Alex Villacorta is a PhD student in the Department of Statistics (expected graduation 2006). His faculty advisor is Dr. S. Rao Jammalamadaka and his research interests are in the areas of multimedia data mining and modeling, and technology for education.

# Interaction Studies with Autonomic Systems

Peter Khooshabeh[1, 2], Eser Kandogan[2], Christopher Campbell[2], Paul Maglio[2]
IGERT Fellow[1, 2]
Department of Psychology[1], IBM Almaden Research Center[2]

***Abstract*— As IT systems become increasingly more complex, there need to be new interaction paradigms in order for human operators to stay in control. The IBM Autonomic Computing vision aims at making systems manage themselves given high-level objectives from administrators in the form of policies. Policy-based interfaces are a means of keeping the human in the loop in the face of this increased automation. This paper reports on a controlled experimental study where three factors were manipulated. Preliminary results and lessons learned are discussed.**

***Index Terms*— automation, human-computer interaction, mental representation**

## I. Introduction

COMPUTING systems have grown in complexity and scale. As large networks of systems continue to grow, the control bottleneck resides in software used to manage systems. Current tools do not accommodate system administrator work practices and mental representations. Human operators need new software tools and automation in order to stay in control.

Most research in the field of automation and human control has been in the areas of nuclear power plant (NPP) control and aviation (air traffic control) [1]. However, IT systems pose different constraints on human operators. NPP operators deal with continuous variables like fuel and temperature. IT system administrators control discrete variables such as port numbers to establish network connections, which can be either right or wrong. IT systems need to adapt to constant change. Whereas NPP are built once for a specific context, IT systems grow as business demands change. A start-up company with five clients must drastically scale its IT infrastructure when it gets 5,000 clients. The next section describes the motivation for this study and a model of a simulated IT system called Sim Sys.

## II. Motivation

Previous work used a model of a city simulation to study policy-based interaction [2]. Participants in the previous study were told to take on the role of a city mayor and maximize city health, similar to the popular Sim City game. The advantage of using a city model was that there was no prior knowledge required since most people have a good understanding of how of to run a city. However, the city simulation is not completely isomorphic to an IT system. Therefore, the Sim Sys model was developed to simulate a webstore IT infrastructure. Participants were told to take on the role of a Chief Information Officer (CIO) and match the configuration of their system in response to customer load on the webstore. The scenario in the study was that different shoppers would visit the webstore to get gifts for Mother's Day, and the participants' task was to get high profits for the webstore just as a CIO aligns business strategy with technical resources. Customers came in three different types: Window Shoppers, Direct Shoppers, and Mixed Shoppers. Window shoppers did not lead to more sales because they only browsed. However, they would hold up a spot in the queue that would prevent Direct Shoppers from making purchases. Mixed Shoppers behaved like either Direct Shoppers or Windows Shoppers; that is, some Mixed Shoppers did make

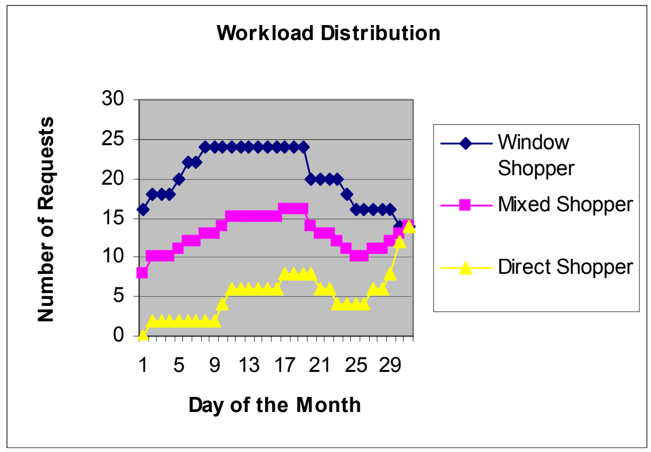



Fig. 1. The distribution of shoppers over the 30 minute duration of the study. Each minute represents a day leading to Mother's Day on the 31st.

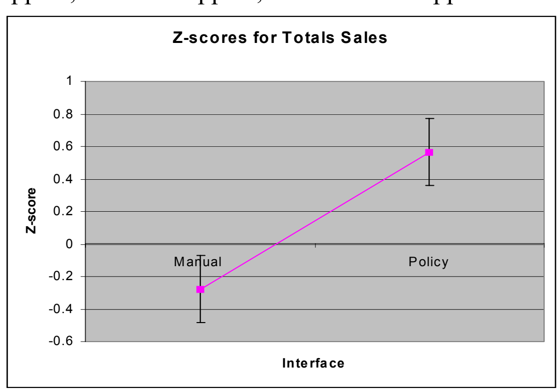



Fig. 2. Participants accrue more sales using the policy-based interface.

purchases. Each shopper request needed to be processed within a certain amount of time. For example, Direct Shoppers needed to be processed within 4.5 seconds or the webstore would miss a sale.

The Sim Sys model consists of Processes that have a list of properties, can be connected to other process, and can contain sub-processes. The processes also have a steady-state behavior.

## III. METHOD

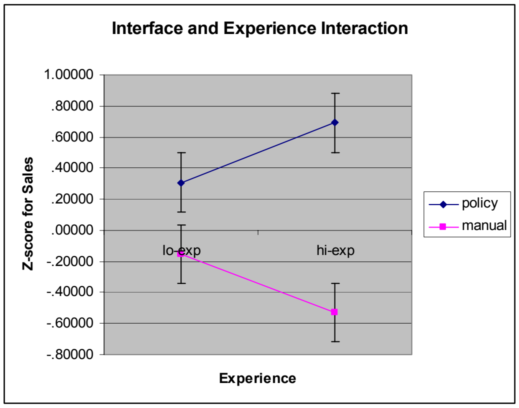



Fig. 3. Participants that use the policy-based interface perform worse when switching to manual interface.

Seven participants were recruited from the IBM staff (4 Females). The experiment consisted of four sessions that lasted an hour each. Most participants performed one session per day, participating in the experiment on four separate days. One participant chose to participate in two sessions separated by a break on the same day. A pre-study questionnaire was given to gauge the level of general technical experience as well as domain knowledge in the fields of system administration and web architectures.

The three independent variables that were manipulated were Interface, Experience, and Policy Representation. Manual and Policy were the two types of Interface. Experience was manipulated by having participants perform the study in multiple sessions, thereby gaining experience by playing the Sim Sys game. The Policy Representation could be either High Detailed Representation or Low Abstract Representation. An example of a highly detailed policy is, “Over the last 5 days when average request latency is above 10 sec increase server CPU and Disk speed by 10 percent a day.” The same policy represented with less detail is, “There is a lot of request latency in the servers so increase server capacity.” The latter policy is more abstract. All the policies were categorized with a label that indicated their triggering event. The previous policies were under the Latency Increase category.

Upon completing the four experimental sessions, the three high scorers were invited to play a “bonus” round. During the final bonus session, the selected participants controlled the Sim Sys game with the highly detailed policies but were able to edit these policies. This allowed for mixed-initiative control where both the system and the user make decisions [3].

The experimenter encouraged participants to speak their thoughts similar to a verbal protocol study.

## IV. PRELIMINARY RESULTS

An ANOVA was computed to compare the total sales accrued using manual or policy interfaces to manage the webstore[1]. A significant difference was found between the manual and policy-based interface ($F(1,23)=4.843$, $p < .05$) (see Figure 2).

## V. DISCUSSION AND FUTURE WORK

The order of Interface was counterbalanced for each participant. Four of the participants started the experiment with the manual interface and the other three started with the policy-based interface. Since there were two policy sessions (low and high representation), participants also used the manual interface in two sessions. Figure 3 indicates that when participants start with the policy-based interface their performance drops when they switch to the manual interface. After using the policy based interface and gaining expertise with it, switching to a completely manual interface hurts their performance because the interaction model is different.

Future work will include further analysis of interaction data with respect to type of interface and experience. Preliminary results indicate that the top performers reach a personal record when playing Sim Sys with the ability to edit policies. A next step in this direction is to study negotiation and grounding with policies [4].

## REFERENCES

[1] E. Hollnagel, Woods, D., *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. Boca Rotan: CRC Press - Taylor and Francis, 2005.

[2] C. S. Campbell, Kandogan, E., November, A., Barrett, R., Maglio, P.P., "Policity: An Experimental Evaluation of Policy-based Administration in a City Simulation," presented at IEEE POLICY, Stockholm, 2005.

[3] M. Hearst, "Mixed-Initiative Interaction," *IEEE Intelligent Systems*, pp. 14-24, 1999.

[4] S. E. Brennan, "The Grounding Problem in Conversations With and Through Computers," in *Social and cognitive psychological approaches to interpersonal communication*, S. R. Fussell, Kreuz, R.J., Ed. Hillsdale, NJ: Lawrence Erlbaum, 1998, pp. 201-225.

**Peter Khooshabeh** is currently in the Psychology department working towards a PhD in Cognitive and Perceptual Science (expected graduation in 2008). His faculty advisor is Dr. Mary Hegarty and his research interests are in the broad areas of human-computer interaction and spatial cognition. He holds a Bachelor’s degree from UC Berkeley in Cognitive Science with emphases in Cognitive Psychology and Computational Modeling.

---

[i] Total sales were used as the dependent measure instead of profits. This seems to be a better measure of performance. Other measures of performance, such as throughput, are currently being analyzed.

# Sound Putty: Developing an Augmented Reality Experience for the FogScreen

Stephen DiVerdi
IGERT Fellow
Computer Science

MarkDavid Hosale
IGERT Associate
Media Art and Technology

***Abstract*—The goal of the project is to exploit the potential of the FogScreen through the intuitive interaction with virtual dynamic physical phenomena. Coupled with various tracking technologies, the FogScreen facilitates interaction with physical phenomena through direct interaction. With the FogScreen users can manipulate content simply by placing their hands on the screen. This differs from current interaction techniques where the users must place their hands in front of the screen (or somewhere else entirely). The FogScreen has a better potential for emerging a user in a virtual environment by removing the spatial disconnect commonly found in more conventional computer interfaces.**

**Content developed for the FogScreen, ranging from the literal to the abstract, will explore its potential as a new medium. Content should be intuitive to navigate, but, as with any new medium, may require some initial trials before a user can successfully navigate a particular space. In the long term content for the FogScreen will be able to provide a meaningful interface for virtual exploration, performance and composition.**

***Index Terms*—new medium, direct interaction, sonification, isosurface rendering, particle system, physical simulation**

## I. The Interactive FogScreen

INTRODUCED in 2003, the FogScreen [1] is a novel display technology that creates a stable sheet of thin fog in mid-air which can be used as a screen for a standard projection system. Unlike traditional screens, the FogScreen's stable, immaterial nature allows users to penetrate and even completely walk through its surface, making a variety of new applications possible. Aesthetically, the FogScreen is a captivating visual experience that has captured the attention of every type of audience.

The Interactive FogScreen [2] is a logical next step in the evolution of the display technology, adding an interaction component to the large format display, allowing users to directly manipulate applications designed for display on the FogScreen. At the beginning of this summer, we experimented with a number of different input technologies to determine what would best compliment the FogScreen's particular strengths and weaknesses. The final determination was to use a WorldViz Precision Position Tracker [3], an unobtrusive, infrared vision-based tracking system. The resulting interactive environment includes a FogScreen placed in the center of a large room, to accommodate a user viewing the FogScreen from both sides. Four infrared cameras are mounted around the perimeter of this space to track user-worn infrared LEDs for tracking. Eight loudspeakers are also placed around the space, four at the ceiling and four on the floor, to spatialize sound in the environment.

Designing an interface for use within this environment proved to be another challenge. Screens are typically associated with passive experiences. Whether in a movie theater or watching television at home, a user is not typically inclined to participate directly with the content on the screen. With the FogScreen a user is compelled to make direct contact with the screen, either by disrupting its surface with a hand or by walking through the FogScreen itself. This creates an expectation that the screen will respond directly to the user. The most common desire is to want to grab an object and manipulate it on the screen. This need for interactive content on the FogSceen is what drives the content we have developed.

Based on these observations several criteria were established for the FogScreen content. As stated before, the FogScreen implies direct interaction, therefore the content should be more physically intuitive and less abstract. If the interaction with the content is too ambiguous the user tends to think that there is no interaction, or that the system is malfunctioning. The level of intuitiveness needs to be balanced with keeping the content compelling – overly intuitive content can easily become uninteresting because it is too superficial and lacks depth. Finally, the content should hold a user's attention even though they may not understand exactly how the interaction works. This is important to emphasize because, as with any new medium, the FogScreen experience will involve a short learning curve as the user becomes accustomed to the interface.

## II. Sound Putty

The result of our exploration of the FogScreen as an artistic and interactive medium is our project, called "Sound Putty". Sound Putty is a virtual blobby fluid that simulates the dynamics of a putty-like material, which the

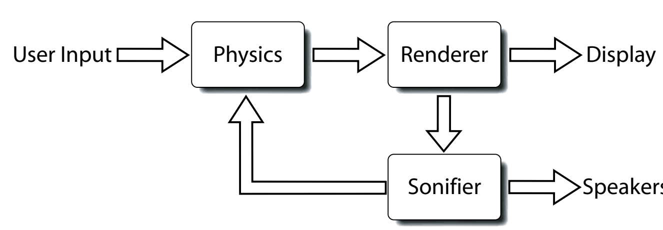



Fig. 1. Overview of the Sound Putty system architecture. User input is given to a physics engine to compute the fluid's configuration. The renderer creates a surface to visualize the fluid and displays it on the screen. The surface is also passed to the sonifier for interpretation as audio on the system speakers. The audio is sent back to the physics engine to complete the feedback loop.

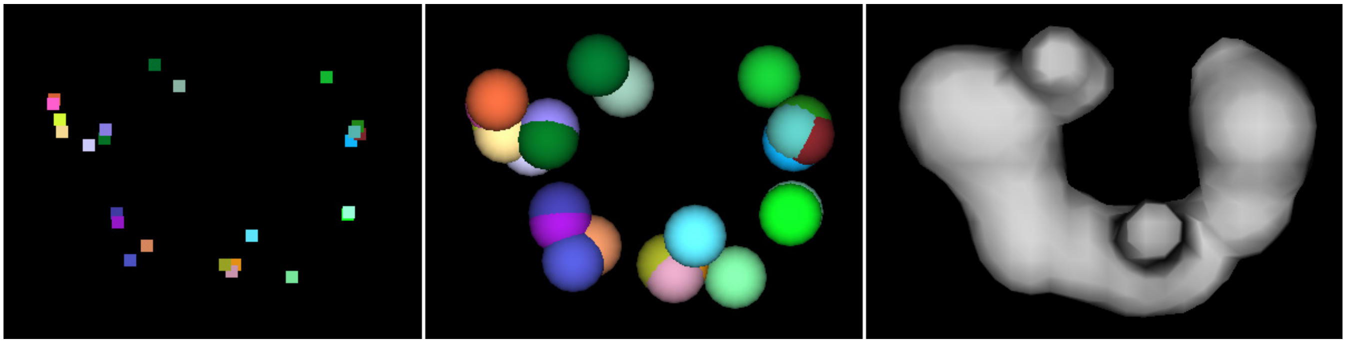


Fig. 2. Various renderings of the Sound Putty. *Left to right:* a) The 3D particles rendered as points in space. b) The same particles rendered as spheres showing their individual extents. This shows the discrete nature of the data, which lacks the cohesion of a fluid. c) The isosurface rendering of the same particles. The surface is determined by the weighted contributions of each particle, creating a smooth, continuous result.

user can stretch, squeeze, and manipulate in 3D. The shape and position of the Sound Putty is interpreted as part of an abstract audio environment, controlling the parameters of a sound composition system.

Figure 1 shows a high-level view of the Sound Putty system architecture. The user's input from the 3D tracking system is fed into a physical simulation engine which determines the motion of the fluid. The resulting shape of the fluid is sent to a visualization engine that generates a 3D surface of the fluid based on the physical simulation. This 3D surface is sent as polygon data to the FogScreen for rendering, while metadata (shape, position, etc.) about the surface is sent to a sonification engine. The sonifier generates sound based on the fluid metadata, which is rendered to the eight speakers around the FogScreen. The output audio is also fed back to the physical simulation engine, where it is used to further manipulate the fluid, creating a closed loop.

## III. Visualization

To create a realistic and interactive physical simulation, a particle system is used to simulate the Sound Putty's fluid dynamics. Fluid attraction and repulsion forces are simulated per particle pair, creating a reasonable approximation of real particle interaction at very low computational cost. User interaction is then enabled by allowing the user to manipulate attractive and repulsive forces in the environment. These forces do not directly control the particles' positions, but instead allow the user to coax the fluid into different shapes by adjusting the shape of the resting state the fluid is attracted to.

The particles themselves do not create a compelling visual, however, as they are separate, discrete entities, while a fluid is a continuous substance. To create the appearance of a smooth, blobby surface out of the particles, each particle is treated as a point source of a smoothly decaying field in 3D, akin to an electrical charge. The set of particles creates a complex distribution of charge within the volume occupied by the particles, which can be rendered with an isosurface – a surface within the volume where every point on the surface has the same value in the field. This is the 3D equivalent of a contour plot, where each line represents a set of points of equal value over a 2D function. To construct the isosurface for our fluid, we use the standard marching cubes algorithm [4], accelerated with an octree data structure [5] to improve interactive performance. See Figure 2 for a comparison of various rendering techniques.

## IV. Sonification

As mentioned before, the sonification of this world is based on two models. Musically speaking the relationship between these two models is analogous to that of a melody and accompaniment.

The first is, acoustically, a direct representation of the graphical representation on the screen. As the user interacts with the Sound Putty, they change its shape, while simultaneously controlling a sound object that represents the equivalent acoustical state of the graphical representation. The mapping between the fluid properties and its acoustical equivalent is basically one-to-one. Mappings are directly connected to the features of the virtual fluid, such as particle density, object position and size. For example, if the object is stretched out, then the result will be a warping of the sound's timbre. If the fluid's position changes on the screen, the resulting sound will have its panning altered. If the fluid's density is altered then the resulting sound may change in pitch.

The second model is the sonification of an invisible plane, parallel to the surface of the FogScreen, that the putty rests on. The behavior of the plane is similar to that of the surface of water, but instead, it is made out of a mesh of sound nodes. As the putty moves around the FogScreen it creates ripples in the plane, exciting the nodes in the mesh. The nodes then create sound ripples that represent the background sound of the work.

Because the fluid is stimulated by the sound it makes and indirectly causes the sound mesh to activate, this environment can be thought of as a self-actuated system. By changing the state of the fluid in the system, the user is having an indirect effect on the environment as a whole.

## V. Future Work

The future of the Interactive FogScreen includes exploration of its potential as a pseudo-3D display, creating an immersive volumetric experience. This would clearly benefit Sound Putty, giving users a much more profound sense of directly interacting with the virtual object in physical space. We'd also like to work on enabling multi-user collaboration in a visual and aural environment, allowing multiple simultaneous users to work with multiple Sound Putties and their surrounding environment. This work will present unique challenges in tracking, real-time rendering, and visual-aural mapping.

## References

[1] Ismo Rakkolainen and Kari Palovuori, "A walk-thru screen," in *SPIE Electronic Imaging 2002, Proceedings of Projection Displays VIII Conference*, January 2002.
[2] Ismo Rakkolainen and Kari Palovuori, "An immaterial, interactive screen," in *Society for Information Displays*, May 2005, pp. 102–105.
[3] WorldViz, Semptember 2005, http://www.worldviz.com/.
[4] W. E. Lorensen and H. E. Cline, "Marching cubes: a high resolution 3D surface reconstruction algorithm," *Computer Graphics*, vol. 21, no. 4, pp. 163–169, 1987.
[5] H. Samet, *The Design and Analysis of Spatial Data Structures*, Addison-Wesley, New York, 1989.

**Stephen DiVerdi** actively researches in the fields of real-time graphics and augmented reality. He received his B.S. in Computer Science from Harvey Mudd College in 2002, and is currently pursuing a Ph.D. in Computer Science at University of California in Santa Barbara. His IGERT advisors are Tobias Höllerer and George Legrady.

**MarkDavid Hosale's** interests are interdisciplinary, but the connecting tissue comes from thinking of art and music as a narrative. Particularly, the kinds of narrative that are structured using non-linear representations of information, time and space. He received his B.A. in Music from the College of Creative Studies at UCSB in 1999, and a Masters of Music in Music Composition and Theory from the University of Illinois at Urbana-Champaign in 2001. Currently, he is pursuing a Ph.D. in Media Art and Technology at University of California in Santa Barbara. His IGERT advisors are Marcos Novak and Stephen Travis Pope.

# Investigating Matching Pursuit Decompositions of Non-Noisy Speech SignalsUsing Several Dictionaries

Bob L. Sturm
IGERT Fellow
Electrical and Computer Engineering

***Abstract*—Matching pursuit (MP) provides a method for expanding arbitrary signals in terms of any usually over-complete set of time-limited functions, or atoms, together called a dictionary. The signal can thus be represented sparsely by functions that are well-localized in time and frequency, or are more highly correlated to the signal than other basis functions. It has been shown that care must be taken when designing the dictionary, but little work has studied in detail the effects of the dictionary on decomposition. Below we investigate the effects of the dictionary on the decomposition of speech signals using non-orthogonal MP. Decompositions of four speech signals using five different dictionaries are compared in their ability of efficiently and accurately representing signal energy.**

***Index Terms*—Matching Pursuit, Atomic Decomposition, Speech Analysis**

## I. INTRODUCTION

RELATIVELY recent work has focused on expanding digital signals in terms of functions that are more correlative than other basis functions such as complex exponentials. One such method, matching pursuit (MP), iteratively finds best matches of a signal to atoms in a usually highly redundant and over-complete set of functions, called a dictionary [1, 2]. The selection is based on maximizing the similarity of the signal and dictionary vectors. The signal can then be represented in a very flexible way as a linear combination of $N$ scaled dictionary functions and a residual:

$$f(t) = \sum_{n=0}^{N-1} \langle R^n f, g_{\gamma_n} \rangle g_{\gamma_n} + R^N f \qquad (1)$$

where $g_{\gamma_n}(t)$ is a unit-norm dictionary vector, indexed by $\gamma_n$. After each match, the process repeats on the residual signal until its energy passes a specified minimum or a required number of atoms have been found.

Selected atoms are recorded in a "book" containing parameters that characterize each one, such as center time, frequency, and inner product magnitude. The resulting representations are more sparse and flexible than other expansions, but at the price of efficiency, convergence, and increased computation.

MP decompositions (MPD) can be used to, among other things, code and reconstruct a signal, aid in feature extraction and signal analysis [3, 4], classification [5], and modification [6]. Furthermore one can achieve time-frequency distributions (TFD) of signals that are more precise than the short-term Fourier transform by a superposition of Wigner-Ville transforms of individual atoms. In this way one also removes the interference terms created by a Wigner-Ville transform of the entire signal [1].

Mallat and Zhang [1], and Davis [2], have studied the theoretical behaviors of the decomposition process, and the importance of the dictionary. Others have researched optimal dictionary sizes for a decomposition [7], and have suggested that particular atoms are more efficient for some signal types, such as speech [8]. Despite this growing research, there still exists little exploration in the actual effects of different dictionaries on signal decomposition. Given a monophonic, non-noisy speech signal at 8 kHz, how efficiently can it be represented by a given dictionary? In this paper we present our results in decomposing several speech files with five different dictionaries, two of them hybrid.

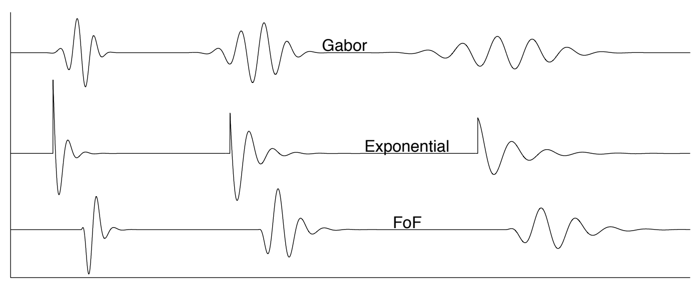



Fig. 1. Examples of different dictionary atoms.

## II. RESEARCH METHOD

To investigate the effects of dictionaries on non-orthogonal MPD of speech, we have used the LastWave (LW) software package, with the MP implementation by Gribonval, Bacry, and Abadia [9]. There are numerous atom types available in this software, including Gabor functions (modulated Gaussian windows), modulated exponential windows, and FOFs ("Fonction d'onde Formantique") [10]. Figure 1 shows examples of these atoms at different frequencies and scales.

Four short speech signals from different speakers were selected for this study, taken from "Books on CD." Each signal was reduced to one channel and resampled to 8 kHz, with 16-bit words. Non-orthogonal decompositions of each speech signal were found using three homogeneous, and three hybrid dictionaries: Gabor functions, modulated exponentials, FOF, and unions of those three in pairs. These atoms were selected because the inner products can be computed analytically. The entire signal was decomposed at once, as opposed to using windowed approach.

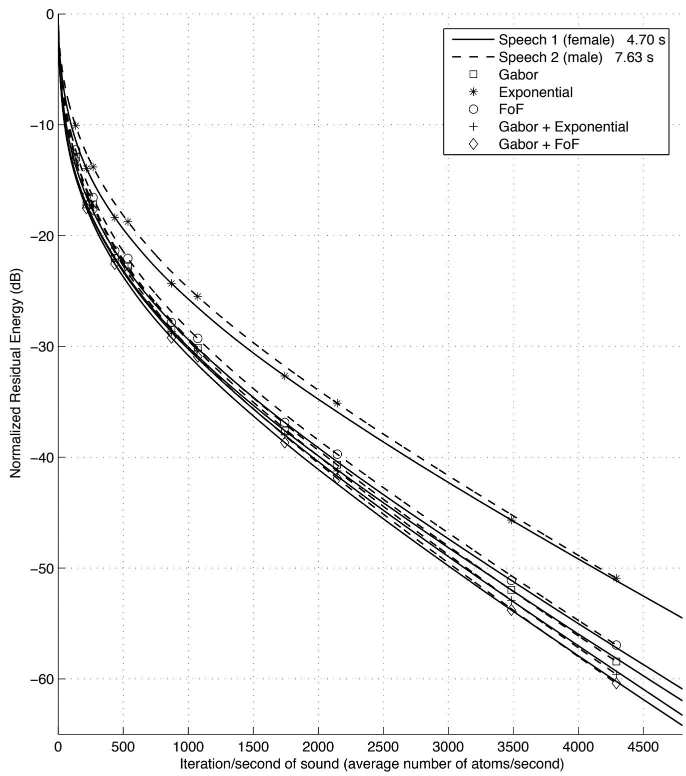



Fig. 2. Residual energy of two speech signals as a function of average atom rate for the five dictionaries.

## III. RESULTS

Figure 2 plots the decay of residual energy for two speech signal decompositions for each dictionary, as a function of average number of atoms per second (aps). Each decomposition follows a common exponential descent. Though non-orthogonal MP is not convergent in a finite number of steps [2], for these 16-bit quantized signals a decomposition can be considered to have converged once the residual energy is lower than -96 dB, or in the worst case when the number of atoms exceeds the number of samples $N$.

It is clear here that the dictionary of modulated exponentials is the least efficient out of the six at representing the signals. For speech 1, at a residual level of -50 dB, over 1000 more aps are needed for the exponential than for the Gabor dictionary. Both hybrid dictionaries perform better than the three homogeneous ones, which is no surprise since increasing the size of a dictionary will give a more sparse decomposition. The dictionary that combines Gabor and FOF atoms performs the best, though minimally so compared with the homogeneous Gabor dictionary. At a residual level of -50 dB, this difference is about 250 aps.

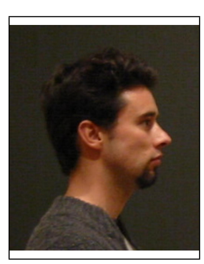
When using a hybrid dictionary, MP has a choice between the different types of atoms. We have found in these cases that Gabor atoms are selected more often than exponential or FOF atoms. Furthermore the number of FOF atoms selected in each 6 dB reduction of the residual is about twice the number of exponential atoms selected; and the ratio of FOF and exponential atoms to Gabor atoms increases each time the residual energy is halved. This implies that as the decomposition advances, the residual begins to correlate more with those atom types than before. The probability of selecting a particular atom type becomes more uniform.

## IV. CONCLUSIONS

We have experimentally compared the performance of five different dictionaries for non-orthogonal MP decomposition of non-noisy speech signals at 8 kHz. It has been shown that the performance of dictionaries containing Gabor atoms work sufficiently well for representing signal energy and in reconstructing the signals with a minimum amount of distortions. Though average atom rates for near-perfect reconstructions of these speech signals are around 3000 - 4000 aps, we believe that this number could be reduced by removing atoms that have little perceptual significance, as well as using a windowed approach to the decomposition. However, in non-orthogonal MP one can not be sure if a particular atom corresponds to the signal, or to an error introduced by previously selected atoms. Thus removing a perceptually insignificant atom could result in a perceptible distortion it was used to correct. Future work will aim to distinguish between these possibilities.

## References

[1] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Sig. Proc.*, vol. 41, no. 12, pp. 3397–3415, 1993.
[2] G. Davis, *Adaptive Nonlinear Approximations*, Ph.D. thesis, New York University, 1994.
[3] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Sig. Proc.*, vol. 51, no. 1, pp. 101 – 111, 2003.
[4] S. Jaggi, W. Carl, S. Mallat, and A. Willsky, "High resolution pursuit for feature extraction," *Technical report, MIT, November*, 1995.
[5] K. Umapathy, S. Krishnan, V. Parsa, and D. G. Jamieson, "Discrimination of pathological voices using a time-frequency approach," *IEEE Trans. Biomedical Eng.*, vol. 52, no. 3, pp. 421–430, 2005.
[6] G. Kling and C. Roads, "Audio analysis, visualization, and transformation with matching pursuits," in *Proc. of the COST G-6 Conference on Digital Audio Effects (DAFX-04)*, 2004.
[7] Q. Liu, Q. Wang, and L. Wu, "Size of the dictionary in matching pursuit algorithm," *IEEE Trans. Sig. Proc.*, vol. 52, no. 12, pp. 3403–3408, 2004.
[8] M. Goodwin and M. Vetterli, "Atomic decompositions of audio signals," in *Proc. IEEE ASSP Workshop Appl. Signal Process. Audio and Acoust.*, 1997.
[9] E. Bacry, "Lastwave software [online]," .
[10] X. Rodet, "Time-domain formant-wave function synthesis," in *Spoken Language Generation and Understanding*, J. C. Simon, Ed., pp. 429–441. D. Reidel, New York, 1980.



**Bob L. Sturm** has presented his research and music all around the world. In 1999 he earned an M.A. from Stanford University in Computer Music Technology, and in 2004 he earned an M.S. in Multimedia Engineering from UCSB. He is currently working on a PhD in Electrical and Computer Engineering at UCSB in digital signal processing with applications to music. His IGERT advisors are Jerry Gibson and Curtis Roads.

# InVision: A Tool for Interactive Visualization of High Dimensional Data

Dan Overholt
IGERT Fellow
Media Arts & Technology

Wesley Smith
IGERT Fellow
Media Arts & Technology

Nhat Vu
IGERT Fellow
Electrical & Computer Engineering

***Abstract*—Visualization of high dimensional data enables the user to gain an intuitive understanding of the data's behavior and greater interaction with the data. In this paper, we present an interactive visualization tool developed based on the kernel Principal Component Analysis (kPCA) framework. Given a set of feature vectors in high dimension, the visualization tool projects the data onto two principal components at a time and provides variation isocontours and data proximity maps that guide the user's interpretation of and interaction with the data.**

## I. Introduction

IN many applications ranging from biological image processing to interactive audiovisual composition and performance, high dimensional data or features are generated and analyzed to make predictions and decisions. However, to gain visual insight into the behavior of the data or to interactively navigate the feature space, the original data must be reduced down to two or three dimensions for visualization. Often dimensionality reduction is a tradeoff between how much information is preserved versus the simplicity of visual representations. Consequently, the reduced data may not wholly depict the behavior of the original dataset.

Motivated by a common need to gain intuitive understanding of the data's behavior and greater interaction with the datasets from our individual research areas, we developed a versatile software tool that allows interactive visualization of high dimensional data.

## II. Dimensionality Reduction

### A. Principal Component Analysis

We investigate a widely used data reduction technique known as Principal Component Analysis (PCA). PCA projects the original high dimensional data onto a smaller subset of dimensions that capture most of the variations or energy in the data. The advantages of PCA are its ease of implementation and its computational efficiency for reasonable size datasets. However, PCA assumes the data to have a multivariate Gaussian distribution, which is not always true in practical settings. Consequently using only the principal axes as a guide when visualizing or synthesizing new data points may yield misleading behaviors. See Fig. 1.

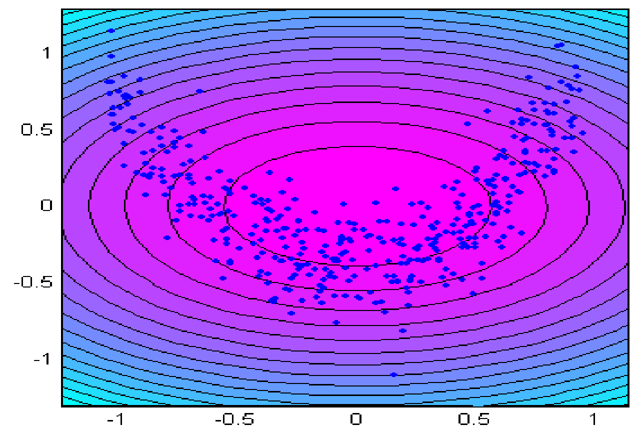



Fig. 1. Non-Gaussian distributed test data points projected onto the first two principal dimensions. The contour lines indicate the level of variations from the mean at (0,0). PCA suggests that data far away from the mean is less likely to be observed (lighter area), but there are clearly data points present at those locations.

### B. Kernel Principal Component Analysis (kPCA)

To supplement the PCA framework, we use kernel PCA (kPCA) to enhance the visualization. In kPCA, the original high dimensional data is nonlinearly mapped onto a possibly higher dimensional space where the mapped data is approximately Gaussian distributed. Taking advantage of the Gaussian nature of the mapped data, PCA can then be used to determine the major axes of variations in the data. These

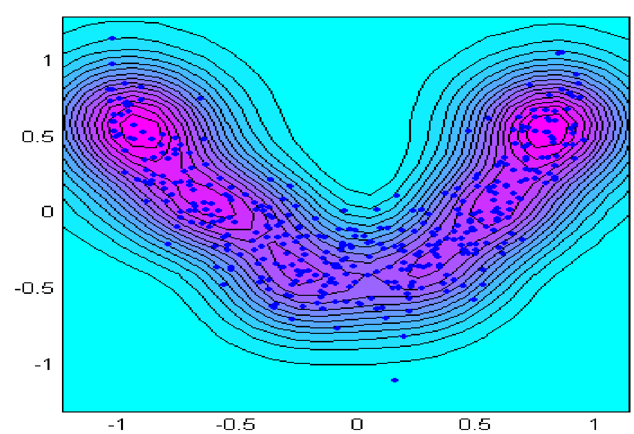



Fig. 2. The same data as in Fig. 1 along with variation isocontours and data proximity map (shading) calculated using kPCA. The distribution of the data is more accurately characterized compared to using PCA alone.

principal components can more accurately capture the spread or behavior of the data. We also calculate the data proximity function, which is a measure of closeness to the data points in kPCA space, and project it onto PCA space. The data proximity function is shown as a grayscale map in Fig. 2 and as a 3D surface in Fig. 3.

## III. Interactive Visualization Tool

### A. Functionality

We wish to take advantage the PCA framework because it is easy to implement and can be used to synthesize new data points interactively. However, to provide the user with greater intuition about the behavior of the data, we also use information from kPCA space to supplement the visual display. The visualization is performed by the following steps:

1. Perform PCA on the set of training data.
2. Use kPCA to compute the data proximity map.
3. Display the result as an interactive 3D surface.

The user can then interact with the training data using two main modes.

1. Synthesize new data by moving the mouse cursor over the surface and visualize the result.
2. Use the 3D surface to monitor incoming user input in real-time from devices such as the FiberEye discussed below.

### B. Implementation

We implemented InVision using the Jitter media programming environment. Jitter is a very flexible, real-time data processing system that can readily connect with sensors, video streams, audio, and many other types of data. It is also widely used in the New Media Arts community for performances and installations. In our implementation, we also added some matrix analysis extensions and visualized the data using OpenGL.

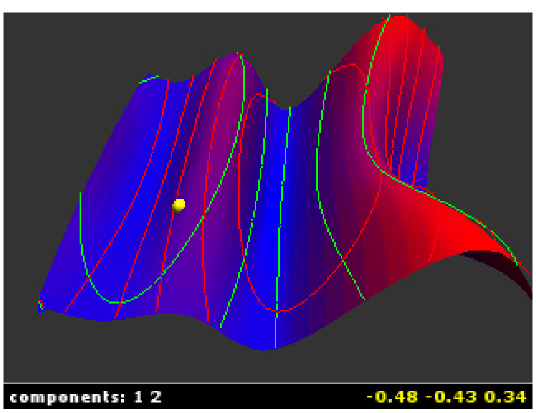



Fig. 3. From InVision: a surface showing the kPCA data proximity with variation isocontours 1 and 2. The yellow sphere is the current location of the input data vector.

## IV. Preliminary Datasets

### A. FiberEye

The FiberEye is an interactive device that uses a bundle of fiber optics to sense hand positions. A camera located at the base of the bundle records the amount of light traveling through the fibers. We divide the bundle image into 64 equal area sections and use the mean pixel intensity of each section as our 64 dimensional data vector.

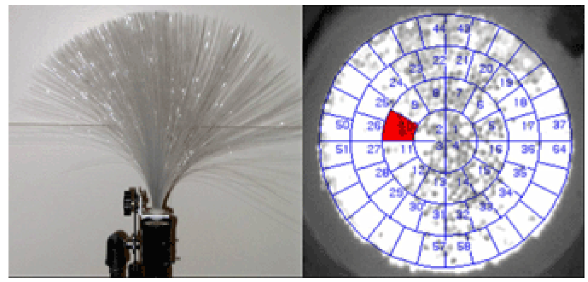


Fig. 4. The FiberEye (left) and the captured image (right) with the section grid overlay and section 10 highlighted.

In order to recognize hand positions, we collect training vectors from a set of predetermined poses and then process this data using our visualization tool. Once the data proximity surface is computed, we can navigate the surface by using 'live' data from the FiberEye. The kPCA framework could also be use to determine the pose of a current hand position. The visualization of this process allows us to more precisely map input vectors to control parameters for live audiovisual performances, as well as to gain better control of the interface as a human-computer interface.

### B. Bioimage Analysis

Working with Dr. Steven Fisher in the Molecular Cellular and Developmental Biology (MCDB) department, we seek to quantify changes observed in retinal images during detachment experiments. The features extracted from these images include both cone shape data and the statistics describing the spatial patterns of nuclei in the outer nuclear layer. The visualization tool can potentially reveal underlying biological processes throughout detachment.

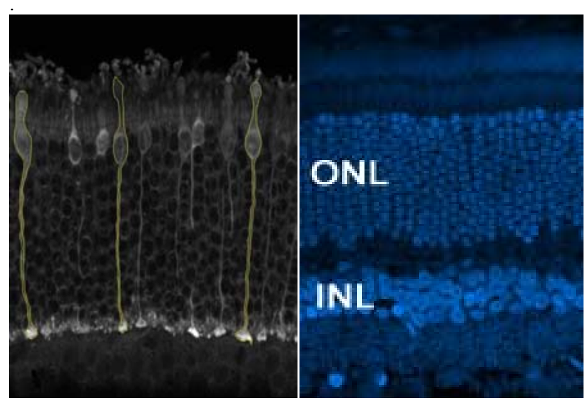



Fig. 5. Cone contours (left) and stained nuclei (right).

## V. Conclusion

A new visualization framework has been developed to facilitate the scientific analysis and interactive control of high dimensional data. We have used this kPCA-based system for both feature analysis of retinal detachment images and interactive human-computer interfaces for the control and generation of artistic audiovisual material. Future work will involve extending InVision to describe kPCA space in more detail and applying our system to more computer vision and media arts related applications.

## Acknowledgment

We would like to thank our advisors George Legrady from Arts and Media Arts and Technology (MAT), B. S. Manjunath from Electrical and Computer Enginerring (ECE), Stephen Pope from MAT, and Steven Fisher from MCDB for their guidance and our colleagues in the IGERT program for their constructive feedbacks. We would also like to thank the National Science Foundation for its financial support in the form of NSF IGERT Grant# DGE-0221713.

**Dan Overholt** received a B.S. in Electronics Engineering and a B.A. in Music from California State University, Chico in 1998 and the M.S. from the MIT Media Laboratory in 2000. He is currently working towards a Ph.D. in Music Technology at UCSB.

**Wesley Smith** received his B.S. in Electrical Engineering and B.A. in French from Johns Hopkins in 2002. He is currently working toward a Masters in Visual and Spatial arts in MAT at UCSB.

**Nhat Vu** received his B.S. in Electrical Engineering from Washington University in St. Louis in 2003 and the M.S. in Electrical and Computer Engineering from UCSB in 2005. He is currently working towards a Ph.D. in Electrical and Computer Engineering.

# Appendix: IDM IGERT Publications to date (2003-2005)


## Journal Publications

1. Sturm, B. “Pulse of an Ocean: Sonification of Ocean Buoy Data.” *Leonardo*, Vol. 38.2, pp. 143-149, 2005.

## Conference Publications, Posters, and Presentations

1. Sturm, B. and J. Gibson. “Signals and Systems Using MATLAB: An Integrated Suite of Applications for Exploring and Teaching Media Signal Processing.” 2005 Frontiers in Education Conference. Indianapolis, IN. October 2005.
2. Khooshabeh, P., E. Smith, and J., Thompson. “Gestural Music and Improvisation Programming.” IEEE Symposium on Visual Languages and Human-Centric Computing. Dallas, TX. September 2005.
3. Zhang, H., K. Almeroth, and M. Bulger. “An Activity Monitoring System to Support Classroom Research.” Proceedings of ED-Media 2005 Conference on Educational Multimedia and Hypermedia. Montreal, Canada, pp. 1444-1449. June 2005.
4. Keehner, M., and P. Khooshabeh. “Computerized Representations of 3D Structure: How Patterns of Interactivity Differ Among Learners.” Paper presented at the American Association of Artificial Intelligence, Stanford, CA. March, 2005.
5. Dillemuth, J. “Map Design for Mobile Display.” AutoCarto 2005, Las Vegas, NV. March, 2005.
6. Khooshabeh, P., M. Keehner, C. Cohen, M. Hegarty, D. Montello. “How Learners Comprehend and Interact with 3D Computerized Represenatations of Anatomy-Like Structures.” Gaming and Simluation Based Learning: Applications for Medicine, Stanford University Medical Media & Information Technologies SIMWorkshop, Long Beach, CA. January, 2005.
7. Cohen, C. “The Influence of Spatial Ability on the Use of Dynamic, Interactive Animation in a Spatial Problem-solving Task.” Gaming and Simluation Based Learning: Applications for Medicine, Stanford University Medical Media & Information Technologies SIMWorkshop, Long Beach, CA. January, 2005.
8. Overholt, D. “The Sonic Scanner and the Graphonic Interface.” International Computer Music Conference, Miami, FL. November, 2004.
9. Sturm, B. “MATConcat: An Application for Exploring Concatenative Sound Synthesis Using MATLAB.” International Computer Music Conference, Miami, FL. November, 2004.
10. Black, A. “Userradio,” Piksel Festival, Bergen, Norway. November, 2004.
11. S. DiVerdi, T. Höllerer and R. Schreyer, “Level of detail interfaces,” In Proc. ISMAR 2004 (IEEE/ACM Intl. Symp. on Mixed and Augmented Reality), Arlington, VA, November, 2004.
12. Dillemuth, J. “Cartography for Mibile GIS.” GIScience 2004, Aldephi, MD, October 2004. **Finalist** in the AAG (Assoc. of American Geographers) GIS Specialty Group Student Paper Competition.
13. Goldsberry, K., Fabrikant, S. I., Kyriakidis, P. “The Influence of Classification Choice on Animated Choropleth Maps.” Annual Meeting of the North American Cartographic Society, Portland, ME, October, 2004.
14. Dillemuth, J. “Generalization of Navigational Maps for Mobile GIS.” GIScience 2004, Aldephi, MD, October 2004.

15. Overholt, D. “Visually Controlled Synthesis using the Sonic Scanner and the Graphonic Interface.” 117th Audio Engineering Society Convention, San Francisco, CA, October, 2004
16. Boughman, T. and S. Fabrikant. “Realism and Perceptions of Quality in Computer-Displayed Maps.” GIScience 2004, Aldephi, MD, October 2004.
17. Black, A. “Userradio,” ACM Multimedia Conference, New York, NY. October, 2004.
18. Sturm, B. “MATConcat: An Application for Exploring Concatenative Sound Synthesis Using MATLAB.” 7th International Conference on Digital Audio Effects (DAFx'04), Naples, Italy. October, 2004.
19. Hosale, M. and J. Thompson. “DEFENDEX-ESPGX.” SIGGRAPH 2004, the 31st International Conference on Computer Graphics and Interactive Techniques. Los Angeles, CA. August, 2004.
20. Drescher, B., “Prior Expectations of Context and Saccadic Decisions in Natural Scenes,” Western Psychological Association Annual Convention, Phoenix, AZ. April 2004. **Winner**: Psi Chi Graduate Research Award.
21. Drescher, B.“Prior Expectations of Context and Saccadic Decisions in Natural Scenes.” Vision Sciences Society Annual Meeting, Sarasota, FL, April 2004.
22. S. DiVerdi, D. Nurmi, T. Höllerer. “ARWin - A Desktop Augmented Reality Window Manager.” In ISMAR '03, October, 2003.
23. S. DiVerdi, D. Nurmi, T. Höllerer. “A Framework for Generic Inter-Application Interaction for 3D AR Environments.” In The Second IEEE International Augmented Reality Toolkit Workshop, October, 2003.
24. Kaplan, E. “Algorithmic Ontology and Polyfurcation of Self.” 2003 Interactive Media Forum, Oxford, OH. October, 2003.

## Performances

1. Black, A. and B. L. Sturm. “Pacific Pulse.” Los Angeles So
2. Overholt, D. “Duet for Violin + Violinist.” Dutch Electronic Art Festival (DEAF'04), Rotterdam, Holland, November 2004.
3. Black, A. “DataDada.” Piksel Festival, Bergen, Norway, October 2004– December 2004.

## Installations and Exhibitions

1. Hosale, M. and J. Thompson. “DEFENDEX-ESPGX.” SIGGRAPH 2005, the 32nd International Conference on Computer Graphics and Interactive Techniques. Los Angeles, CA. August, 2005.
2. Black, A. “PropaganDada.” Evolving Perception: Time, Space, and the Human Body, Los Angeles, CA. November 2004 – December 2004.
3. Black, A. and B. L. Sturm. “Pacific Pulse.” Evolving Perception: Time, Space, and the Human Body, Los Angeles, CA. August 2004 – December 2004.

# Notes

# Notes

# Notes