Computer Science and Artificial Intelligence Laboratory

The principal mission of the Computer Science and Artificial Intelligence Laboratory (CSAIL) is to invent new information technologies and discover new computing paradigms that improve lives of people everywhere. CSAIL is committed to discovering and developing innovative approaches and solutions to the way we understand and interact in the world. Our work addresses such critical areas as communications, health care, the environment and transportation.

CSAIL has a long history of technological innovation that has impacted how people interact and do business. Previous innovations from Laboratory members include time-shared computing, public key encryption, computer chess, web standards, GNU, TCP/IP, and ARPANet. Current research explores mobile computing, the next generation of laptops, and the application of sensor technology to traffic congestion, animal herding, medical monitoring, and climate observations. Robotic locomotion and human robotic interface are under investigation, as well as medical solutions with image-guided surgery and technical applications to aid clinical decisions. Advancements in biological research are also underway, including developments in the field of computational biology and the application of machine learning to the interpretation of complete genomes and understanding gene regulation.

CSAIL's research is administratively organized into three broad categories:

- *Systems* covers all aspects of the building of both hardware and software computational systems.
- *Theory* looks at the fundamental mathematical underpinnings of all aspects of computer science and artificial intelligence.
- *Artificial Intelligence* includes language, learning, vision, and graphics, which works on the sorts of things that all people manage to do effortlessly, both emulating those abilities and simulating their appearance; and physical, biological, and social systems or complex adaptive systems, covering work from robotics to molecular biology, and from semantic systems to computational models of politics.

Research at CSAIL is sponsored by a large number of US government agencies including the National Science Foundation, the Defense Advanced Research Project Agency, the National Institutes of Health, the Air Force Office of Scientific Research, the National Aeronautics and Space Administration, the Office of Naval Research, a wide spectrum of US and international companies including Quanta Computer, Inc., Nokia, Shell, Microsoft Corporation, Toyota, Nippon Telegraph and Telephone Corporation, The Boeing Company, DuPont, SAP, Cisco Systems, Inc, Pfizer, Inc., and other organizations including the Singapore-MIT Alliance, the Industrial Technology Research Institute, the Defense Science Technology Agency, the Commonwealth Scientific and Industrial Research Organization, the Delta Environmental and Educational Foundation, and the Epoch Foundation.

Research Projects

Within CSAIL we have many single- and multi-investigator projects as well as a number of virtual centers and large-scale mega projects. These large-scale projects and collaborations include:

T-Party

T-Party is a five-year, $20 million research project sponsored by Quanta Computer, Inc. The goal of the project is to (1) develop the next generation of platforms for computing and communication beyond personal computers, (2) create new systems for the development and seamless delivery of information services in a world of smart devices and sensors, and (3) move from a device-centric perspective to a human-centric one. Victor Zue, Srini Devadas, Jim Glass, Frans Kaashoek, Chris Terman, and Steve Ward are the lead principal investigators from CSAIL.

During this second year of the collaboration, 13 principal investigators pursued four projects. They are:

- Virtualized Computation Platform, focusing on the development of secure and reliable computation and storage
- T-Net, the development of a direct, secure, authorized, and authenticated access to (mobile) personal devices
- Just Play, distributed systems automatically constructed from ad-hoc collections of disaggregated devices
- Natural Interactions, emphasizing the use of human language as a central ingredient in a multimodal interface (combining speech and gesture) for naive users and in real-world environments where a traditional graphical user interface is not practical

Mobile Communications EcoSystems

We are halfway through this three-year, $7.5 million research project sponsored by Nokia. The goal of the project is to develop new modes of interaction in the use of mobile communication devices and to reduce the development cost of new services and energy-efficient phones. Researchers from Nokia Research Center Cambridge, located within a five minutes' walk from CSAIL, work closely with CSAIL researchers. So far four joint meetings involving 35 to 100 Nokia researchers have been held at MIT and at Nokia facilities in Helsinki and Tampere in Finland. Arvind, Rodney Brooks, and Victor Zue comprise the CSAIL committee that oversees this project.

This initiative started in January 2006 and currently consists of the following projects:

- Simone, developing techniques to interact with applications on mobile phones primarily using spoken dialog
- ARMO, reducing the development cost of new energy-efficient, high-performance phones

- MyNet/UIA, making it dramatically easier to connect devices to share information and services in a secure manner
- Tra, developing techniques for secure information flow on phones running stock Operating Systems
- ConnectingMe, redefining personal information management
- StartMobile, developing Natural language queries via SMS
- Telepresence, providing high quality mobile medical telepresence

Defense Advanced Research Project Agency Urban Challenge

A team of MIT faculty, postdocs, graduate students and undergraduates spanning several departments and laboratories is competing in the 2006–2007 DARPA Urban Challenge, the goal of which is to develop a vehicle capable of safe, autonomous driving in city traffic at speeds up to 30 mph. The competition will be held on a closed course, probably a military base, and will not involve pedestrians.

The team uses two vehicles for testing, a Ford Escape and a Land Rover LR3. Each is equipped with a drive-by-wire system for computer control of the gas, brakes, steering, shifter, and turn signals. Each has a suite of exterior sensors—light detection and ranging, camera, and automotive radar—to estimate drivable road surface and track other vehicles, along with navigation sensors—GPS, inertial, and odometry—to estimate the vehicle's position and orientation with respect to the road network. A 40 CPU supercomputer in each vehicle's cargo area executes algorithms that plan a fast, collision-free path through the environment to accomplish the driving goals specified by DARPA observers. The supercomputer and sensors are powered by gas-fueled generators carried on board.

The team tests its vehicles extensively on the grounds of the former South Weymouth Naval Air Station. As of June 2007, the vehicles can perform basic navigation tasks such as road-following and U-turns, can pass stopped obstacles, can traverse intersections while respecting standard "rules of the road" for precedence, and can execute emergency stops. Over the next four months the vehicle will gradually exhibit more sophisticated behaviors such as merging and emergency collision avoidance.

Center for Information Security and Privacy

The goal of the Center for Information Security and Privacy (CISP) is to develop both the theoretical foundation for secure systems and to engineer practical systems. CISP addresses the full range of security concerns and technologies, from hardware and architectural support for secure computing to applications with significant societal impact such as electronic voting and email authentication. CISP has particular strength in cryptography, both theoretical and applied—indeed, many of the major advances and themes in cryptographic research originated in CISP. Much of the center's applied research focuses on problems emerging from the growth of the internet, as millions of embedded networked devices come online. CISP also develops techniques for

promoting individual privacy in the face of pervasive sensors and networks. CISP is headed by Ron Rivest. Srini Devadas and Frans Kaashoek act as associate heads. More information can be found online at http://cisp.csail.mit.edu/.

Center for Robotics

The CSAIL Robotics Center was formed in September 2005 to provide a mechanism for collaboration and for promoting robotics within CSAIL. The center is codirected by Tomas Lozano-Perez and Daniela Rus. The center has 12 princpal investigators from four different departments, 76 graduate students, 11 postdocs, and many MIT undergraduates participating in the Undergraduate Research Opportunities Program.

The CSAIL Robotics Center's mission is to conduct cutting-edge, long-term research and education in robotics. Our research addresses fundamental problems in designing more capable robots and controlling them to interact intelligently with people, the environment and each other. We also explore how increased capability and intelligence can enable new societal benefits through applications in homes, fields, oceans and outer space. Our basic and applied research inspires the development of novel course material aimed both at training students in robotics and at using robots to explore computation, sensing, and control.

World Wide Web Consortium

Hosted through a partnership between MIT, the European Research Consortium for Informatics and Mathematics, and Keio University, the World Wide Web Consortium (W3C) has standardized the foundation of today's web (including HTML and XML), and is developing the foundation for tomorrow's web.

The Consortium has experienced significant growth in technical work and participation. Members are drawn by new work transforming the current web of linked documents to an expanded web of data and services across a wide range of devices and that aims to enable everyone on the planet to collaborate and share information.

The W3C Rich Web Clients Activity is standardizing AJAX and other technologies to support improved client-side experience on the web. The Mobile Web Initiative is making web browsing on mobile phones and other small devices as useful and usable as it is on larger computers and screens. Future work will focus on web ubiquity, leveraging web technologies to support interoperable data exchange and multimodal interaction including graphical, voice, and gesture across devices as diverse as appliances, entertainment systems, printers, projectors, cars, planes, and specialized hand-held devices.

Foundational specifications for web services are developed at W3C, with work on policy and semantic annotations just recently underway. W3C's early leadership in semantic web research has resulted in completion of the basic semantic web standards, and is supporting current work on query and rules interchange languages and new work to address interoperability challenges within health care, life sciences, multimedia, and geospatial domains. New web content accessibility guidelines describe how to make the

web accessible to people with disabilities, which will also have increasing implications for the world's aging population. Though XML became a standard in 1998, important additions like XML Query, Voice XML and binary XML are under development.

Web Science Research Initiative

The Web Science Research Initiative is a joint endeavor between CSAIL and the School of Electronics and Computer Science at the University of Southampton. Its goal is to facilitate and produce the fundamental scientific advances necessary to inform the future design and use of the World Wide Web.

The initiative has four founding directors: Tim Berners-Lee, director of W3C, senior research scientist at MIT and professor at the University of Southampton; Wendy Hall, professor of computer science at the School of Electronics and Computer Science at the University of Southampton; Nigel Shadbolt, professor of artificial intelligence at the University of Southampton and director of the Advanced Knowledge Technologies Interdisciplinary Research Collaboration; and Daniel J. Weitzner, technology and society policy director of W3C and principal research scientist at MIT. Jim Hendler, professor of computer science at Rensselaer Polytechnic Institute serves as associate director.

Living the Future

A team of CSAIL researchers, along with colleagues from the Media Lab and from MIT Information Services and Technology, is exploring a new research initiative with the working name of "Living the Future" or LTF for short. The goal of LTF is to deploy various types of next-generation datastreams, infrastructure, services, and devices onto the MIT campus on an accelerated schedule, and encourage the MIT community to contribute to and develop creative applications of these new capabilities within an open-source model.

Some of the key novel capabilities envisioned within LTF include: fine-grained indoor location awareness, context awareness, and user activity awareness; user-controlled identity and privacy mechanisms; history logging and data mining; and complex social connectivity. The team hopes to engage a significant portion of MIT students, staff, and faculty in the adoption, extension, and application of these mechanisms to enable fundamentally new kinds of mobile experiences.

Research Highlights

In addition to the large-scale collaborative projects and center research, numerous individual and multi-investigator projects are under way. 306 research projects at CSAIL can be found online at http://publications.csail.mit.edu/abstracts/abstracts07. A sampling of the work is highlighted below:

Lecture Retrieval and Access System

In the past decade, there has been a dramatic increase in the availability of online academic lecture material. Low-cost media and fast networks have opened the door to new and exciting ways to disseminate knowledge in media formats, ranging from audio recordings to streaming video. These educational resources can potentially change the

way people learn—students with disabilities can enhance their educational experience, professionals can keep up with recent advancements in their field, and people of all ages can satisfy their thirst for knowledge. It is conspicuous, however, that in contrast to many other communicative activities, lecture processing has until now enjoyed little benefit from the development of human language technology.

The goal of CSAIL research in this area is to provide efficient access to large online multimodal collections of educational material through development of the Lecture Retrieval and Access system. The exploration began by analyzing the performance of commercial speech recognition engines on lecture recordings. An unexpectedly high word error rate (WER)—66.8 percent—motivated research on new recognition techniques tailored for the lecture genre. Preliminary investigations have improved the WER to 40.6 percent through acoustic and language model adaptation. The new method depends on effective use of a domain-specific written corpus (e.g. textbooks) to accurately predict lecture vocabulary. In parallel, methods that can accurately recognize lecture fragments, which are crucial for successful navigation, are being explored. Specifically, an unsupervised learning method that can automatically identify technical terms that are spoken multiple times in an audio-only source (i.e. no text) is being developed. These technical terms can subsequently be used for browsing and for retrieval of lecture material. Future plans include exploration of synergies between text-based and audio-based keyword discovery mechanisms.

As a first step toward automatic summarization and content retrieval, we focused on the problem of topical segmentation of lecture material and developed a segmentation algorithm based on spectral clustering methods and random walks that has been successfully applied to image segmentation. These methods were able to pinpoint the most salient topic boundaries in spontaneous speech transcripts, but at a finer granularity the performance degraded. Currently we are exploring ways to modify existing spectral clustering methods to take into account the constraints imposed by lecture discourse.

Finally, we developed a tool for automatic terminology extraction from a corpus of lecture transcripts. We are currently in the process of customizing this tool to the needs of the MIT OpenCourseWare Initiative, where the tool will be used to automatically create metadata for a large repository of MIT course materials.

This research is headed by associate professor Regina Barzilay.

Robot Locomotion in Fluids

The CSAIL Robot Locomotion Group has already demonstrated that machine learning can design good control solutions for bipedal walking robots where traditional control techniques fail. Now we are working on building robotic birds that you can see flying around Killian Court and Briggs Field. Although there are many technical challenges, the primary scientific challenge in this enterprise is controlling the delicate interaction between our robot and the very complicated, unsteady airflow. This interaction is the

only mechanism by which the bird can produce lift, thrust, and maneuvering forces. Based on some initial experimental success, we are once again investigating model-free machine learning approaches to control.

To control fluid systems, there is a strong claim that it is be easier to design a controller than to understand the fluid dynamics (birds probably don't know Navier-Stokes). Our initial experiments included three simplified versions of the bird problem. First, we developed a control system for a fixed-wing model aircraft that already allows the plane to land on a perch by executing a high angle-of-attack (stall) maneuver. Second, in collaboration with an experimental fluid dynamics lab at New York University Courant Institute, we are optimizing the cost of transport of a flapping-plate (a good candidate for the simplest model of flapping flight). Third, in collaboration with Northeastern University, we have built a small water tunnel in CSAIL to conduct control experiments with real-time flow visualization. These projects have produced new collaborations across MIT (including Aero-Astro, Mechanical and Ocean Engineering, and Applied Math) and indicate great strength in fluid dynamics and particular fluid locomotion. Ultimately, we believe this technology will have broad impacts in flow control and may revolutionize techniques as diverse as stroke (blood clot) prevention or energy conservation through more efficient and responsive building heating and cooling.

This research is directed by assistant professor Russ Tedrake.

StreamIt: A Language and Compiler for Multicore Architectures

StreamIt is a programming language and a compilation infrastructure specifically engineered for modern streaming systems. It is designed to facilitate the programming of large streaming applications. By breaking the von Neumann bottleneck, StreamIt provides better modularity, composability, malleability, and portability for the streaming applications than sequential languages such as C. With StreamIt, parallelism is abstracted away from architectural details, thus providing an efficient and effective mapping to a wide variety of target architectures including commercial off-the-shelf uniprocessors, multicore architectures, and clusters of workstations. StreamIt compiler has demonstrated that it can obtain sustainable speedups on multicores (5x to 19x on the 16 core Raw), enabling increased abstraction without sacrificing performance.

This research is directed by associate professor Saman Amarasinghe.

Programmable Microfluidics

Microfluidic devices are emerging as an attractive technology for automatically orchestrating complex biology protocols, with applications ranging from biomedical research to biological computing. The device technology in microfluidics has been advancing faster than Moore's Law, with the number of features (valves) per chip doubling every four months. However, current microfluidic chips remain very inflexible, designed and fabricated for a narrow application domain.

We are developing new hardware and software abstractions to support fully programmable, general-purpose microfluidic chips. Our approach uses a digital design: fluidic samples are discretized into unit volumes, isolated from one another, and

manipulated independently during operation. Biologists use a high-level programming language called BioStream to map their own unique experiment to our general-purpose chips. In this way, a biology experiment becomes a "program" that can be seamlessly mapped across successive generations of microfluidic chips.

This work is done jointly between Saman Amarasinghe and assistant professor Todd Thorsen's group in Mechanical Engineering.

Model Based Views for Sensor Data

Database management systems (DBMSs) have traditionally been used to store and query business data. For example, many e-commerce web sites use a DBMS to store information about customers, products, and sales. A DBMS provides a convenient and efficient programming model for accessing and manipulating data, as well as recovery and transactional features to prevent data from being left in an inconsistent state by system crashes or interrupted processing. In the FunctionDB project, we focus on extending the benefits of these tools beyond the confines of business applications to a variety of scientific domains.

In particular, FunctionDB provides tools to allow scientists to store and manage measurement data from field sensors that measure various numerical parameters (e.g., temperature, air quality, etc.). Unlike business transaction data, sensor data is typically used as input to a mathematical model of the measured phenomena, and this model is then used to predict or explain the behavior of those phenomena. This modeling process is done because the raw data is noisy and incomplete and so of little direct utility. Modeling is used to smooth out these inconsistencies. To help support modeling, FunctionDB provides direct support for fitting, constructing, and querying a broad class of models created via interpolation and/or regression directly inside of a database system. It allows users to define so-called "model-based views" that fit a model to a collection of raw data and then issue queries over those models. For example, if a user has a collection of temperature readings over time, FunctionDB can fit a series of line segments to those readings using regression and then allow users to ask queries about times when the temperature went or is expected to go above (or below) some threshold. To support this approach, FunctionDB includes an unusual query execution system that evaluates queries over models by solving systems of equations rather than iterating over tables of records as in a traditional database system. Our initial prototype of the system demonstrates the feasibility of the idea of allowing users to store and manage their models in a database system and provides the aforementioned benefits of an efficient high-level query language and a recovery system that can help ensure the integrity of stored data.

This research is under the direction of associate professor Samuel Madden.

Applied Visual Perception: Using Thumbnails to Search the Web

Recognizing that the visual system processes pictures better than words, our research addresses how we can make web searching easier by making use of human vision. We developed enhanced thumbnails; using pictures of web pages enhanced with text, instead of text alone, to summarize search results, thereby providing the benefits of

graphical summaries (visual genre, page layout, images, and a quick interpretation of summary), along with the advantages of text summaries (making useful text readable and being able to graphically depict search terms relative to other parts of a document).

The design starts with a plain thumbnail, enlarges useful text to make it readable, and creates text "popouts" to show location in page of text relevant to query. We have found that web searches are 29 percent faster with enhanced thumbnails than with text and 22 percent faster with enhanced thumbnails than with plain thumbnails.

This research was performed by principal research scientist Ruth Rosenholtz, in collaboration with Allison Woodruff, Andrew Faulring, Julie Morrison, and Peter Pirolli.

Managing Visual Clutter

Visual clutter is a practical problem because it can interfere with searching for an important item in a display or scene. It can also interfere with the easy understanding of a complex information visualization such as a scatter plot, a web page, or a map. A computational measurement of clutter could help optimize display clutter in situations in which there is no human designer in the loop, e.g. dynamic displays. Or it could help by providing system alerts when clutter may impair task performance, e.g. road clutter in driving, clutter in x-ray inspection. Previous measures of clutter often relied on information about the number of objects in the display, or were image-based but overly simple.

We have suggested several measures of visual clutter. One, which we call the "feature congestion measure," is based on the intuition that the more cluttered someone's desk is, the harder it is to add a note that will reliably draw their attention. We have previously proposed the statistical saliency model for visual search, which can tell us how difficult it would be to add an attention-grabbing item to an existing display. According to the statistical saliency model, the difficulty in adding a target that draws attention is related to feature variability. When a display is cluttered, it is (locally) difficult to find features that draw attention because the feature space is already "congested," thus the name of this clutter measure. The measure can output either clutter "maps," showing where a display is most cluttered, or a single value indicating the clutter for the display overall. Considerable testing suggests that this is a promising candidate for a measure of visual clutter.

This research is directed by Ruth Rosenholtz, along with Yuanzhen Li and Lisa Nakano.

A Novel Approach to Representing and Analyzing fMRI Imaging of Brain Activity

In recent work we have demonstrated a novel approach to representing and analyzing spatial activation patterns that emerge in functional magnetic resonance imaging (fMRI) of brain activity. In contrast to traditional methods that employ expert-defined hypotheses to identify all brain locations that co-activate with the predefined "seed" region, the new method simultaneously estimates interesting seed time courses and associates voxels with the respective networks. The resulting network patterns exhibit a highly hierarchical nature, leading to a notion of functional hierarchy. Just as anatomical

hierarchies represent the structure of the brain as a tree of increasingly simple systems, the functional description of the brain that emerges from such an analysis yields a description of the spatial decomposition of the cortex into systems at an increasingly fine resolution. Moreover, a hierarchy can be naturally interpreted by a human observer, providing an intuitive and meaningful representation that can be easily integrated with neuroanatomy and other sources of information. When coupled with a model of interaction across brain areas, the proposed representation yields comprehensive, high-level descriptions of brain activity. The experimental results demonstrate that the functional hierarchy provides a robust and anatomically meaningful model for spatial patterns of co-activation in fMRI. The hierarchical representation leads to insights into the structure of the functional networks that are not immediately apparent from flat representations that segment the brain into a large number of small regions. In addition, subject-specific hierarchies tend to share a common tree structure, further confirming the validity of this representation for modeling group-wise patterns of co-activation.

This research is directed by assistant professor Polina Golland.

Machine Learning for Natural Language Processing

Natural language processing (NLP), or computational linguistics, deals with the application of computational methods to problems involving linguistic data. Examples of application areas within NLP include automatic (machine) translation between languages; dialogue systems which allow a human to interact with a machine using natural language; and information extraction, where the goal is to transform unstructured text into structured (database) representations that can be searched and browsed in flexible ways. In terms of technical and scientific challenges, computational linguistics involves fundamental questions concerning how to structure formal models (e.g. statistical models) of natural language phenomena, and the design of algorithms that implement these models.

Our research focus has been on machine-learning approaches to NLP problems. A first sub-area of research is the development of a theory and algorithms underlying machine-learning approaches to *structured* problems—problems that involve complex, discrete structures such as strings, labeled sequences, or trees. Structured problems of this kind are extremely prevalent in NLP as well as in related research areas such as computational biology, speech recognition, and machine vision. One contribution over the last year has been the development of a new family of algorithms for discriminative parameter estimation. A second contribution has been the development of a new formalism—case-factor diagrams—that subsume two central frameworks in probabilistic models for structured problems: Markov random fields and weighted context-free grammars.

A second sub-area of research has concentrated on applying machine-learning methods to NLP problems such as machine translation, dialogue systems, and speech recognition. One approach learns to map sentences to logical forms that represent their meaning. For example, in building a natural language interface to a database of geographical information, the goal would be to map strings such as “what states border Texas?” to logical forms that can be used to query the database. This research has developed a

radically new approach to the problem, one that simultaneously learns the syntactic and semantic representations required for the mapping. A second application is machine translation. Recent work has shown how syntactic information can be used to improve the quality of machine translation systems through a linguistically motivated account of systematic differences in word order between languages such as German and English.

In the past year, this research has resulted in best paper awards at three major conferences in the field: Empirical Methods in Natural Language Processing 2004 and Uncertainty in Artificial Intelligence 2004 and 2005.

The research is headed by associate professor Michael Collins.

Realistic Graphics and Computational Photography

In the areas of realistic graphics and computational photography, a new theoretical framework has been developed that studies light transport from a signal-processing perspective. The frequency spectrum of the light array is modeled and analyzed to discover how it is modified by phenomena such as propagation in free space, occlusion, or glossy reflection. This approach has important potential applications in light simulation and inverse problems. New techniques have been developed to model the appearance of materials—that is, their texture and the way they reflect light. Unique high-resolution measurements of reflectance functions have been performed and made available to stimulate research in this area. In the domain of computational photography and video, a video processing technique has been introduced that analyzes and magnifies small motions. A hardware and software solution has been designed that can extract a foreground object from a video sequence based on defocus effects.

This research is headed by associate professor Frédo Durand.

Computational Modeling of Biological Shape

The study of biological shape and its variability has always played an important role in the life sciences. In human brain mapping, recent advances in magnetic resonance imaging have enabled noninvasive studies of human neuroanatomy at high resolution, yielding important insights into normal development and the pathologies of anatomical structures. Our research focuses on computational shape modeling and analysis that provide detailed, accurate characterization of shape variations and their interpretation in biologically meaningful terms of development and deformations.

In spite of the recent advances in computational approaches to modeling shape, many biological structures—including the exceedingly complex folding patterns of the cortical surface—still present a significant challenge for statistical shape analysis. We are developing novel shape representations that will lead to robust and efficient learning of shape variability from images. This research will enable statistical analysis of complex shapes in many different areas of biomedical imaging.

Together with our collaborators from the Harvard Medical School, we are validating novel methods for shape analysis in the neuroimaging studies of normal aging and diseases such as Alzheimer's and schizophrenia. In a more recent collaboration with a

group from the Whitehead Institute, we are building on our techniques to characterize the shape and appearance of cells from high-throughput imaging data in the gene knockdown experiments. This project aims to identify the relevant pathways and the functionality of previously uncatalogued genes based on the changes in the cellular phenotype caused by the gene knockdown. Beyond biomedical imaging, our research offers improvements in modeling shape in many other fields that require understanding and characterization of 3D shape from a set of examples, including computer vision, geology, and biochemistry.

This research is headed by Polina Golland.

Databases

The research focus of this group is on sensor-related database applications, with a particular focus on building interfaces that facilitate collection of large amounts of data from widely distributed networks of sensors, particularly those connected via wireless and intermittent network connections. For example, in the CarTel project, we are building a distributed wireless platform for automotive applications. One primary research focus is on providing a secure, easy-to-program environment that allows a wide range of automotive diagnostic and traffic-related applications to be rapidly built. Another research thrust involves integrating probabilistic and statistical models into databases of sensor information to detect unusual sensory events and predict future behavior. In CarTel, such models can be used to predict traffic at particular times on particular routes using historical data.

This research is headed by Samuel Madden.

Laboratory Sponsored Activities

CSAIL Outreach

Imara

The overall goal of Imara is to find and implement long-term, sustainable techniques for making educational technology and resources available to domestic and international communities.

CommuniTech

CommuniTech is a project that has recently come under the umbrella of the Imara organization, founded in 2000 by a former CSAIL graduate student and managed by the Public Service Center. The program seeks to bridge the digital divide in our local communities by providing adults in low-income families with the tools they need to gain access to valuable information they can use to better their lives and the lives of their families.

The Families Accessing Computer Technology (FACT) portion of the program provides training in basic computer applications in a six-week course taught by MIT student volunteers.

The second prong is the Used Computer Factory (UCF), which refurbishes used computers donated from the community and distributes them to the FACT participants at the end of the six-week course. An alliance with a local internet provider gives a discounted rate to participants in the program.

The combination of skills participants gain through FACT, and their ability to refine those skills on the computers provided by UCF, facilitates greater computer proficiency and marketability to prospective employers.

Lacotec Laare, Kenya

The Laare Community Technology Centre has been making great progress since its inception in 2005. This past January Aisha Walcott, an electrical engineering and computer science graduate student, and Shawntel Hines made a trip to Laare to help set up more computers in the center. They also launched a new primary school outreach program and carried out a very successful community-wide information technology awareness event. At least eight primary schools are equipped with laptops running Edubuntu software.

Due to the overwhelming popularity and success of this program, there are plans to expand the center. Eric Mibuari, an MIT alumni and the center's founder, has secured four acres of public land donated by the community to build a new, larger center. The plan is to improve the center's learning facilities, expand its curriculum, access online resources, and increase its capacity to serve more people in the community.

Navajo Nation

This summer, Jonathan Proulx, a senior systems administrator at CSAIL, cotaught a unique computer science class with Dr. David Broderick of Diné College. Students learned practical hardware and software skills while preparing used PCs for distribution to the local Navajo community. The final project in the class was the distribution of the refurbished computers and training local families in their use. While refurbishing computers for use in underserved communities is nothing new, having a college-level course focused on it taught to local students is a unique opportunity to build a sustainable local technology community.

This project is the first fruit of a partnership between Imara and Diné College which began in February, 2007 when Francesca Sheik, director of Information Technology at Diné College, emailed Jack Costanza, her counterpart at CSAIL, about the possibility of Imara working with Diné College and the Navajo Nation. This course represents the first step in that partnership.

Fijian Islands

In April 2006, CSAIL's assistant director for infrastructure, Jack Costanza, and Professor Daniela Rus took several brand new laptops to Taveuni Island, Fiji. The laptops, which were purchased with funds graciously donated by Google, were installed with the Kubuntu GNU/Linux operating system and the Edubuntu educational software

package. Elementary school teachers from many of the Fijian islands traveled to the Maravu Plantation in Taveuni, where Costanza and Rus provided introductory computer classes and hands-on training with the laptops.

The laptops were evenly divided between the various school districts, and typically were located in the homes of school principals, since they were often the only ones with electrical power to charge the battery. Professor Rus has made return visits to refresh training and to answer new questions about the computing systems.

Middle East Education through Technology

CSAIL continued to support Middle East Education through Technology (MEET) this past year. MEET is an innovative educational initiative aimed at creating a common professional language between Israeli and Palestinian young leaders. MEET enables its participants to acquire advanced technological and leadership tools while empowering them to create positive social change within their own communities. Many of our students volunteer to teach MEET summer courses at the Hebrew University in Jerusalem. CSAIL continues to host http://meet.csail.mit.edu/ and provide technical support to the MEET program.

Professional Development

CSAIL hosted the Center for Women and Enterprise/Springboard Enterprise Bootcamp event in November 2006. This workshop was designed to provide entrepreneurs with the tools they need to successfully develop and execute a venture capital presentation. Established in 1995, the Center for Women and Enterprise is a nonprofit organization dedicated to helping women start and grow their own businesses. Springboard Enterprise forums are designed to increase investment opportunities for women-led firms, and to help women entrepreneurs navigate the equity markets.

Seminar Series

Four distinguished speakers gave presentations during this year's Dertouzos Lecture Series. They were:

- Professor Robert J. Full, University of California at Berkeley: "Bipedal Bugs, Galloping Ghosts and Gripping Geckos: Neuromechanical Systems Biology"
- Professor Andrew Yao, Tsinghua University, Beijing: "A Modern Theory of Trust-but-Verify"
- Professor Ed Lazowska, Bill and Melinda Gates Chair in Computer Science and Engineering, University of Washington: "Computer Science: Past, Present and Future"

Organizational Changes

The leadership of CSAIL was under transition this year. Director Rodney Brooks announced his resignation effective June 30, 2007. This spring it was announced that codirector Victor Zue would take over the top leadership role starting July 1, 2007. The director determines policies, examines promotion cases, and discusses strategies

designed to keep CSAIL growing and evolving. Jack Costanza was promoted to the assistant director for infrastructure, overseeing IT infrastructure, IT user support, operations, and communications. Karen Shirer took on the role of assistant director for administration, overseeing finance and human resources. CSAIL's research agenda was steered by four research directors. Steve Ward oversaw the systems research directorate; Madhu Sudan oversaw the theory research directorate; Leslie Kaelbling oversaw the language, learning, vision, and graphics directorate; and Randall Davis oversaw the physical, biological, and social systems directorate.

Awards and Honors

Our faculty and staff won many awards this year, including the following:

Arvind: Association for Computing Machinery (ACM) Fellow

Krste Asanovic: Institute of Electrical and Electronics Engineers, Inc. (IEEE) Computer Architecture Conferences Top Picks, and Design Automation Conference DAC/ISSCC Student Design Contest Winner

Hari Balakrishnan: ACM MobiCom Conference Best Paper Award, and ACM SenSys Conference Best Demo Award

Tim Berners-Lee: National Academy of Engineering (NAE) Foreign Associate; NAE 2007 Charles Stark Draper Prize; UK, Queen's Order of Merit; British Computer Society Lovelace Medal; Massachusetts Innovation and Technology Exchange Leadership Award; Academy of Achievement Golden Plate Award; D&AD President's Award for Outstanding Contribution to Creativity

Rodney Brooks: Australian Academy of Technological Sciences and Engineering Foreign Fellow, and American Academy of Arts and Sciences Fellow

Erik Demaine: Honorary Doctorate, Dalhousie University

Michael Ernst: IBM Faculty Research Award; Automated Software Engineering 2006, paper selected for expedited journal publication; Sun Microsystems Most Innovative JSR of the Year Award; 29th International Conference on Software Engineering 2007, ACM Distinguished Paper Award

Michel Goemans: John Simon Guggenheim Memorial Foundation Fellow

Shafi Goldwasser: International Association for Cryptologic Research (IACR) Fellow, and Israeli Math Union Conference keynote speaker

Frédo Durand: 2007 Spira Award for Distinguished Teaching

Polina Golland: National Science Foundation Career Award

John Guttag: ACM Fellow

Manolis Kellis: *Technology Review's* "Top 35 Innovators Under the Age of 35"; *Genome Technology's* "Top Young PIs"

Bradley Kuszmaul: Jim Gray's Sorting Benchmark Contest winner in five categories; 2006 High-Performance Computing Challenge winner in Best Overall Productivity

Charles Leiserson: ACM Fellow, and MIT EECS MacVicar Faculty Fellow

John Leonard: IEEE Robotics and Automation Society 2006 King-Sun Fu Memorial Best Transactions on Robotics Paper Award

Nancy Lynch: ACM Knuth Prize; and Principles of Distributed Computing Best Student Paper

Samuel Madden: ACM MobiCom Conference Best Paper Award; ACM SenSys Conference Best Demo Award; Sloan Research Fellow; IBM Faculty Research Award

Silvio Micali: NAE Member, and IACR Member

Robert Miller: MIT EECS Louis D. Smullin Award for Excellence in Undergraduate Teaching

Victoria Palay: School of Engineering Infinite Mile Award for Sustained Excellence

Peter Shor: California Institute of Technology Distinguished Alumni Award

Peter Szolovits: Harvard Medical School, Center for Biomedical Informatics Innovation in Personally Controlled Health Record Infrastructure

Brian Williams: Association for the Advancement of Artificial Intelligence Outstanding Senior Program Committee Member

Key Statistics for Academic Year 2007

CSAIL members numbered 784 in AY2007:

Faculty	78 (13% women)
Research Staff	41 (17% women)
Admin, Tech, and Support Staff	77 (60% women)
Postdocs	29 (31% women)
Visiting Faculty	90 (14% women)
Paid UROPs	49 (22% women)
MEng Students	49 (24% women)
Graduate Students	371 (21% women)

Rodney Brooks
Director (2003-2007)
Panasonic Professor of Robotics

Victor Zue
Director
Delta Electronics Professor of Electrical Engineering and Computer Science

More information about the Computer Science and Artificial Intelligence Laboratory can be found at http://www.csail.mit.edu/.