Massachusetts Institute of Technology
Proposal Summary Form

Activity Type	Research
Proposal Type	New
Class Code	E.01-Computer Sciences

Print Form

MIT WBS #

Title of Project Perceptual Priming for Language Learning

Investigator Data

	Department, Lab or Center (DLC)	DLC #	PI Status
PI Name: Patrick Winston	CSAIL	067900	Faculty
Co-PI 1			Faculty
Co-PI 2			Faculty
Co-PI 3			Faculty

Sponsor Data

Sponsor	**DARPA**
Contact	Phone
Address	

Deadline Date	Receipt
Submission Method *Paper Submissions: OSP does not mail proposals*	Other Electronic
Submission Type	Federal Solicitation
Notice of Opportunity *(Identify Program Number or Provide URL)*	DARPA IPTO BAA 08-34

Budget Data

	Initial Period	Total Project Period
Requested Start Date	Apr 1, 2010	Apr 1, 2010
Requested End Date	Mar 31, 2011	Mar 31, 2011
Total Direct Costs	247252	247252
Total F&A	132748	132748
Total Direct + F&A	380000	380000

Cost Sharing: None

Budget comments

F&A Base: MTDC

☐ X if Budgeted Subrecipients

On Campus Rate	68 %
Off Campus Rate	%

Under recovery of F&A (amount and source of funds)

Special Reviews

	Protocol Number	Application Date	Approval Date

☐ Check if any space change, renovation or additional infrastructure is required. This includes additional space, changes in space configuration, power and/or cooling to accommodate computers or other equipment.

Attach Additional Sheets as Necessary for Additional Budget Cost Sharing, Special Reviews and Space Change Explanations

Export Controls

○ Yes	◉ No	Will any equipment be exported by MIT in the course of this project?
○ Yes	◉ No	Will this project require any export controlled information to be received on campus? Contact Zachary Sweet at zsweet@ll.mit.edu if you have questions.

Embryonic Stem Cells

○ Yes	◉ No	Does this project involve the use of pre-existing human embryonic stem cells? (if yes, please submit approval letter from the MIT Committee on Assessment of Biohazards) or the **derivation** of human embryonic stem cells (if yes, submit approval letter from MIT Embryonic Stem Cell Research Oversight committee)? Contact the MIT Biosafety Program or Dr. Claudia Mickelson at 2-3477 if you have questions.

Conflict of Interest

Indicate Yes or No if there are any potential real or perceived conflict of interest as defined in MIT Policies and Procedures, 4.4?

	Yes	No
PI	○	◉
Co-PI/Co-I #1	○	○
Co-PI/Co-I #2	○	○
Co-PI/Co-I #3	○	○

Also, for NIH and NSF Proposals, indicate the date the required Financial Disclosure was made on-line in the space provided for all investigators.

PI Date

Co-PI /Co-I#1 Date

Co-PI /Co-I#2 Date

Co-PI /Co-I#3 Date

My Signature below confirms my review of the proposal. It also certifies that:

1) I am not presently debarred, suspended, proposed for debarment, declared ineligible, or voluntarily excluded from current transactions by a federal department or agency **(http://web.mit.edu/osp/www/debarmen.htm).**
2) I have not and will not lobby any federal agency on behalf of this award **(http://web.mit.edu/osp/www/fedlobrg.htm).**
3) I am familiar with the requirements of the Procurement Integrity Act [OFPP, Section 27 (1-3)] and will report any violations to the Office of Sponsored Programs; **(http://web.mit.edu/osp/www/Procuint.htm).**
4) I certify a) that the information submitted within this application is true, complete and accurate to the best of my knowledge; b) that any false, fictitious, or fraudulent statements or claims may subject me, as the PI/Co-PI/Co-I to criminal, civil or administrative penalties; and, c) that I agree to accept responsibility for the scientific conduct of the project and to provide the required progress reports if an award is made as a result of this application. [*This certification is being added at this time to meet a specific NIH requirement. It also reflects current federal regulations.*]
5) Applicable to the Investigator Only: For the principal investigator, I confirm that I have reviewed all subawards included in this proposal. All subaward direct costs have been reviewed and appear reasonable given the proposed statement of work. All fringe benefit and indirect cost rates have been verified with the subawardee organization as being current for the proposed duration (Verification may be in the form of a letter from an authorized official of the organization)

All Investigators Must Sign (attach additional sheets if necessary)

PI Signature _____ Date _____

Co-PI /Co-I #1 Signature _____ Date _____

Co-PI/Co-I #2 Signature _____ Date _____

Co-PI/Co-I #3 Signature _____ Date _____

Notice: Proposals in final format must reach OSP at least five working days prior to the Sponsor's deadline. Failure to meet the deadline may jeopardize the on-time submission of the proposal and may result in incomplete review by OSP. If subsequent review reveals that the proposal is incomplete or does not conform with Institute or Sponsor requirements, OSP may, on behalf of the Institute, withdraw the proposal from Sponsor consideration.

Institutional Approvals

Department or Laboratory Head

Signature	Date	Signature	Date
Signature	Date	Signature	Date

Other Approvals (Deans and/or VP Research, if required)

Note: Signature of VP for Research required for international programs and proposals

Signature	Date	Signature	Date
Signature	Date	Signature	Date

OSP Administrative Approval

Signature	Date	Signature	Date
Signature	Date	Signature	Date

OSP Use Only:

Sponsor Code

Prime Sponsor Code

Proposal Number

1. Cover Sheet

BAA: DARPA BAA-08-34 mod. 3

Proposal Title: Perceptual Priming for Language Learning

MIT

Technical POC: Professor Patrick H. Winston
32 Vassar St., Room 32-251
Cambridge, MA 02138
Phone: 617-253-6754
Fax: 617-258-8682
Email: phw@mit.edu

Administrative POC: Laureen Horton
Office of Sponsored Programs
77 Massachusetts Avenue, Room E19-750
Cambridge, MA 02139
Phone: 617-253-3922
Fax: 617-258-4734
Email: laureena@mit.edu

Summary of Costs: $380,000 over a 12 month period of performance

Contractor Number: None

Contractor Type: Other Educational (MIT)

2. Table of Contents

1 Cover Sheet 1

2 Table of Contents 2

3 Innovative Claims for the Proposed Research 3

4 Proposal Roadmap 4

5 Detailed Research Objectives 6

5.1 Problem Description: 6

5.2 Research Goals: 6

5.3 Expected Impact: 7

6 Detailed Technical Approach 8

6.1 The system so far: supervised learning 8

6.2 The path forward: unsupervised learning 9

7 Experimentation Plans 11

8 Overall Statement of Work 12

9 Personnel, Qualifications, and Commitments 13

10 Facilities 16

11 Human use 17

12 Intellectual Property 17

13 Organizational Conflict of Interest Affirmations and Disclosure 17

3. Innovative Claims for the Proposed Research

We make 3 innovative claims:

1. **Learning perceptual event-fragments is a crucial first step for learning word meaning:** We propose that word-learning is a 2-stage process; perceptual processes organize and carve up perceptual experience into chunks in a data-driven, unsupervised manner. This is a crucial first step that paves the way for the next; supervised learning of word meaning. Our prototype system will demonstrate both stages. Contrary to other models, ours does not place the entire burden of learning on the few instances that a word is heard, but instead shifts the bulk of it to a preparatory step when *nameless* perceptual concepts are learned.

2. **A novel model of visual attention:** At the core of the visual processing is the notion of attentional state: a set of features that the system extracts while being blind to everything else. In our model the attentional state consists of 3 types of properties/features:

   - Properties of the object at the focus of attention (e.g size, direction of movement)
   - Global Spatial Context: the relative positions and distributions of salient regions with respect to the focus of attention
   - Local Spatial Context: Distribution of contact of the focus of attention region with other regions around it.

   Previous models of attention have simply viewed attention as a kind of filter for the tasks of visual search or object recognition. The model we propose is far richer because it makes visuospatial features explicit, thus enabling the construction of a wide variety of *visual routines* (not just search).

3. **Event signatures can be learned as finite-state-machines in attentional state space:** a time-series of attentional-states generated from a stream of video produces an attentional trace. This trace is a low-bandwidth stream of information that *reduces the dimensionality of the learning space while still preserving the information that distinguishes different events* (e.g. give vs take).

4. Proposal Roadmap

a. **Main Goals of the Proposed Research:** The main goal of this research is to demonstrate how perceptual processes can find structure in experience and use that structure as the target substrate for word meaning.

b. **Tangible Benefits to End Users:** If successful the research will make it possible for autonomous systems such as unmanned vehicles and robot scouts to map words to their perceptual experience - allowing them to understand orders and describe what they see.

c. **Critical Technical Barriers:** There are several challenging problems that we will face:

- Figure-ground-segmentation: Figure-ground segmentation is a well known problem in vision. We will circumvent it for now by using brightly colored objects whose color values are known in advance (see 6.1 for more details).
- Feature Extraction: Deciding which features to extract from video is the next problem. The core of our attentional state model deals with part of this problem by picking a set of elementary features. The attentional state that we're currently using has produced encouraging results. However, we may have to try other features as well and also decide what *combinations* of these elemental features (i.e. feature patterns) are useful for capturing regularities.
- Testing Feature Patterns: We expect the number of unsupervised patterns learned to be in the order of 30,000-50,000. We will need some criteria to decide which of these features are meaningful, and worth keeping, and which ones should be discarded.

d. **Main Elements of the Proposed Technical Approach:** The model of visual attention, and unsupervised extraction of a vast number of patterns from a time-series of attentional state are the main elements of the proposed approach for discovering perceptual patterns in experience.

e. **Basis of Confidence:**We have been developing and testing the visual attentional state model over the past 3 years. Encouraging results in supervised learning give us confidence to move on to the unsupervised learning which is at the heart of this project.

f. **Risk if Work is not Done:** If the proposed work is not done a crucial opportunity will be lost to build enabling technology that will allow artificial systems to communicate with their human operators in natural language.

g. **End Results to be Delivered to DARPA:** The end result will be a prototype system and a final report describing it's capabilities. The prototype system when given videos (of up to 3 interacting brightly-colored objects) sometimes accompanied with words describing the activity - will learn the visual meaning of those words. Subsequently allowing a human to ask it questions in terms of those words or allowing the system to describe a new event in terms of the words it has learned.

h. **Cost and Schedule of the Proposed Effort:** We propose to carry out this project over a 12 month period with demonstrable results in 9 months. The cost will be $380,000

i. **Criteria for Objectively Evaluating Progress:** The progress of this project will be evaluated by a review of a demonstration at the 9 month mark and a final report at the 12 month mark. The criteria for evaluating the demonstration will be

- The ability of the system to discover patterns in its visual experience and thereafter use those patterns to learn word meanings.
- The accuracy (precision and recall) of the concepts learned, evaluated against a test set of data (where ground-truth is labeled by humans).
- The relative merits of this 2-stage approach over supervised learning: Crucially does the 2-stage approach lead to faster learning or is it better at learning from large amounts of unlabeled data compared to a purely supervised approach?

5. Detailed Research Objectives

5.1 Problem Description:

Perceptual grounding is key to natural language understanding: Language makes human intelligence unique. However, we do not yet have systems that do highly accurate machine translation or process natural language. Current approaches are limited in scope to syntax, whereas a significant amount of implicit knowledge (semantics) is assumed in human communication. Having machines acquire this knowledge is important to improving their performance. This project attempts to research methods to enable machines to understand human language rather than just process it with no comprehension of the underlying meaning.

Understanding even a single word like *give* requires that a system be able to visually recognize the action, have access to the first person experience of the giver or the recipient, guess the intent involved in the action, recognize the default social context/conventions surrounding the action (e.g. giving medicine versus giving a present), and imagine or predict the consequences of the action (e.g. that the giver no longer has the object at the end).

Most of these facets of understanding can be tested for and found in a one-and-a-half year-old child. Parents do not systematically enumerate these facets of meaning; indeed, most aren't even aware of them. It is significant that children have already learned an enormous amount about the event *give* well before they may have a word for it. This suggests that *if we are to build a system that understands language we must build representations and processes that enable pre-linguistic conceptual development.*

Pre-linguistic conceptual development happens in multiple modalities (e.g. vision, somatosensory, motor, auditory, ...) For this project we will restrict ourselves to just one: vision. The core problem that the project will focus on is **how to build representations and processes that enable a system to learn a vast library of visuospatial concepts from its visual experience such that it is primed to quickly learn word meanings later on with just a few examples.**

5.2 Research Goals:

We would like to build a system that can understand language, and use that understanding to translate speech or process text. Our approach ignores syntax and focuses instead on learning the meaning of individual words.

We have recently shown a system that can learn perceptual signatures associated with 16 common verbs given just 3 example videos of each. This pushes the state-of-the-art in perceptual-grounding by going well beyond the prior work on learning nouns and adjectives. Verbs are much harder to learn than nouns by virtue of being more abstract and having a temporal nature.

Two long-standing problems in learning word meaning via perceptual grounding are

- *Sparseness of stimulus*: How is it that children learn word meanings from such few instances of hearing a word?

- *Ambiguity of reference or Binding problem*: When they hear *Red Ball* how do they know to not associate it with distractor objects, say the tree in the background?

We are exploring a promising line of work that addresses both issues. The key-ideas are

- *Unsupervised perceptual learning goes on continuously and results in the learning of a vast number of perceptual patterns from experience.*
- *When words come along, certain pre-existing perceptual patterns (the most frequent ones that are triggered by what's currently happening) form the target candidates for the meaning of the word. I.e. word-learning is reduced to attaching a label to one of a set of pre-existing concepts.*

In other words, our hypothesis is that learning word meaning is a two-stage process. First there is unsupervised pre-linguistic conceptual development that carves up perceptual experience into chunks, then there is a supervised learning stage that simply attaches a label (the word) to one or more of these chunks. We call this hypothesis *perceptual priming of language learning*. **Our research goal is to demonstrate both stages in a prototype system over the next year.**

5.3 Expected Impact:

The project will impact the performance of text or speech processing systems by giving them access to the underlying meaning of words. It will also directly impact the ability of unmanned vehicles or command and control systems to have a perceptually grounded vocabulary with which to interact with a human operator.

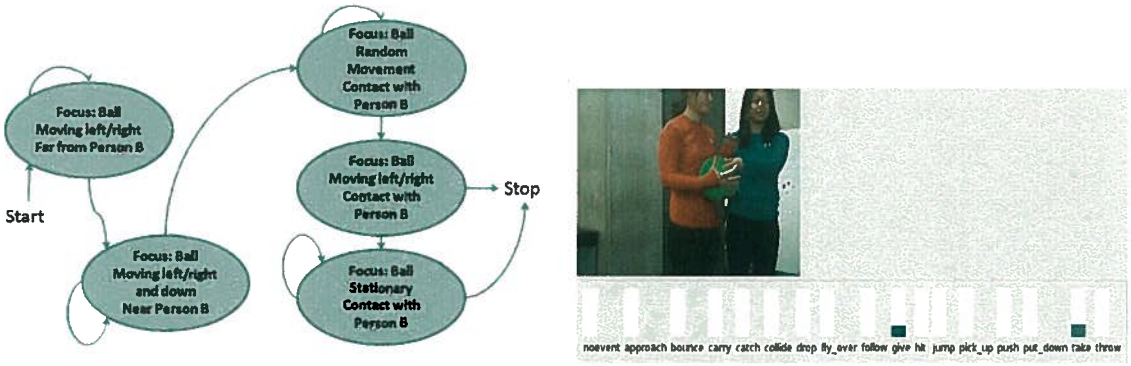
6. Detailed Technical Approach

6.1 The system so far: supervised learning

The results described in this subsection were obtained by one of the investigators (Rao) in a DARPA funded seedling on *Visuospatial Reasoning* at CSAIL MIT.

(a) Give (b) Collide (c) Drop (d) Fly over

Figure 1: Examples of video clips shown to the system - from which it learns visuospatial patterns



(a) Learned model for catch

(b) The system correctly identifies a give/take action using the learned models

Figure 2: Sixteen verb models were learned from training data. Left: one such model - an FSM - for catch is shown here. Right: All learned models compete to explain a new test clip - the models for give and take are detected here.

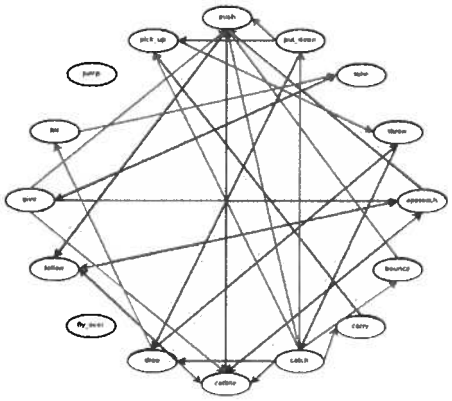
Event perception: We use brightly colored, easily identifiable objects to simplify object-segmentation and tracking. At any moment the system attends to the fastest moving object (a simple, but biologically sound, criterion). At the focus of attention it extracts three types of properties which together constitute the attentional state: (i) Figure properties of the object attended to (e.g., it's size, and direction of motion) (ii) Global context - the relative positions of the other colored blobs of interest and (iii) Local context: a polar histogram of contact relations with the other blobs. As it perceives an event the system generates a time-series of attentional states called an *attentional event-trace*. The main idea is that these event-traces are all that the system retains from an event, however they contain sufficient information about important changes that can be used to characterize the event.

The experiment: The system was shown 3 examples each for 16 common verbs: *push, drop, pickup, put-down, give, fly-over, take, carry, throw, catch, jump, bounce, approach, follow, collide,*

hit along with the rough frame-interval when the event was actually happening (e.g. Carry: frames 81-125)

	Approach	Bounce	Carry	Catch	Collide	Drop	Fly Over	Follow	Give	Hit	Jump	Pick Up	Push	Put Down	Take	Throw
Approach	3				3			2					3			
Bounce	1	2			1	3		1	1	2		1		1	1	
Carry		1	8	3	2					3	2	2	3	3		
Catch		2	2	3	2	3				3	1	2	2	3		1
Collide	2				3			1					3			2
Drop		2			1	3				2		1	1	2		2
Fly Over					2		2			2		1				
Follow	3				3			3			1		3			2
Give	2				2			2	3	1	1	1	2	2	3	1
Hit				1	1	1	1		2	1		1		2	1	1
Jump											3					
Pick Up		1	1							1		2	2	2		
Push	2		1	1	3			1		2	1	1	3	2		1
Put Down			2			3				3		2	3	3		
Take						1			3					2	3	
Throw		2			2		1			3		1	3	3		3

(a) Testing results: 43/48 clips classified correctly, 126 correct and 96 marginal classifications



(b) Discovered mutual associations between the 16 verbs: green is correct, and red is incorrect

Figure 3: Test results

Learning patterns from the attentional-traces: The attentional traces from the 48 examples labeled with the ⟨word, frame-interval⟩ as described above was given as input to a HMM. Sixteen HMMs were learned - one for each verb - a sample HMM is shown in figure 2(a)

Testing: To test the learned models we recorded another 48 clips for the same 16 verbs using different actors and a different background (so that the learning algorithm does not pick on some spurious regularities). Each of these clips was tested against all sixteen models. 43 of 48 clips were correctly classified. Event clips do not map to unique verbs - a test clip for *give* is likely to have an *approach* or a *take* in it as well. Accordingly we manually went through all the categorizations made by the learned models and detected 127 correct classifications, and 96 marginal or incorrect ones (false positives).

For only 3 training examples per verb the supervised learning results so far have been encouraging and proof-of-concept that attentional state representations can indeed capture implicit knowledge about events.

6.2 The path forward: unsupervised learning

Going forward, we will keep the idea of event models being FSMs in attentional state, however we will abandon the HMM framework altogether and defer the labeling of events. Instead, we will let the system discover recurring patterns of various sizes in its attentional state and build up a vast number of patterns (a kind of visual alphabet) with which to interpret new events.

The approach that we expect to take for discovering unsupervised event patterns is:

1. Make *transitions* in attentional state explicit. See Table 1 for a sample set of transition

variables

2. Build up a dictionary of attentional states and the transitions that occur from those states. In order to accomplish this we must find a way to match states and transitions.

3. Evaluate the efficacy of any pattern by how frequently it is detected or its predictive power - i.e. how often it correctly predicts the next few states. In practice such probabilities will be built into the FSM and transitions learned.

We will very likely have to try a few different choices of attentional state *and iterate through these 3 steps* for every choice to find the attentional state choice that gives the best set of learned patterns - that is, with the most predictive power.

variable	description
F	Current focus. -1 is none, or object number otherwise.
E_i	1 if object i is in scene, 0 otherwise.
M_i	1 if object i is moving, 0 otherwise.
X_i	1 if object i is going to the right, -1 if to the left, 0 if not moving horizontally.
Y_i	1 if object i is moving up, -1 if moving down, 0 if not moving vertically.
S_i	1 if object i is getting bigger, -1 if shrinking, 0 otherwise.
C_{ij}	1 if objects i and j are in contact, 0 if not.
R_{ij}	1 if the distance between objects i and j is increasing, -1 if decreasing, 0 otherwise.
W_{ij}	1 if objects i and j are rotating around each other, 0 otherwise.

Table 1: Transition variables: making changes in attentional state variables explicit

We have conducted some initial tests of the unsupervised learning approach described above and learned more than 30,000 visual event fragments. But what do these thousands of learned patterns mean? and are they at all useful? To answer this we chose 4 patterns at random, verified that they do occur across the events they were detected in, and most importantly, verified that they describe valid visuospatial concepts that can be described using a few words as shown below.

- $R_{01} = 0 : -1 \quad X_0 = 0 : 1 \quad X_1 = 0 : -1$: two objects start approaching, one moving to the right, and one moving to the left. This can be described as *walking towards each other.*

- $X_0 = 0 : 1 \quad M_0 = 1 : 0 \quad C_{01} = 0 : 1 \quad C_{01} = 1 : 0$: an object moves to the right, stops, makes contact and breaks contact. This can be described as *hit from the left.*

- $C_{01} = 0 : 1 \quad R_{01} = 0 : -1 \quad Y_0 = 1 : 0 \quad R_{01} = -1 : 0$: two objects make contact, start approaching the other, one stops moving up, and then the distance stops increasing. This can be described as *nudge from the bottom.*

- $R_{01} = 0 : 1 \quad C_{01} = 1 : 0 \quad Y_0 = 0 : -1 \quad M_1 = 0 : 1$: two objects get distant from each other, break contact, one starts moving down, and the other starts moving. This can be described as *drop and go.*

7. Experimentation Plans

We will perform the following experiments:

Experiment 1: Learning a visual alphabet: Recorded videos are used as unlabeled input and the system identifies recurring patterns of various sizes (FSMs with increasing numbers of states and transitions) in its attentional state. A preliminary exploration shows that the system finds more than 30,000 patterns of size 4 - i.e. patterns which have 4 transitions occurring in attentional state variables - each of these patterns is unique, and occurs in at least 3 different videos.

Experiment 2: Learning word meanings: If our hypothesis is correct, then the learned visual alphabet should make it easier to learn word meanings in a supervised manner (by attaching words to one or more elements of the visual alphabet - the technical challenge is to make this association with just a few examples). We will present a new, small training set of videos labeled with words or simple sentences and have the system learn word meanings.

Experiment 3: Testing classification: We will measure recall and precision of the learned word meanings on a new test set of videos.

Experiment 4: Learning part-based models of humans: A convenient assumption that we've made (and will continue to do for this project) is that there are at most 3 brightly colored objects in the scene whose color values are known in advance - thus freeing us from the problem of figure-ground segmentation. While we use this simplifying assumption in experiments 1-3, we would like to be free of this assumption in the long-run. Accordingly we have a completely independent experiment that tries to learn human part-based models from *arbitrary videos of people in motion.* As a result we expect to learn a code-book of static and dynamic human part templates at various scales and view-points. This is a hard problem - which is why it is an independent thread. Lack of progress in this experiment will not affect the results from Experiments 1-3 at all. However even moderate success would tremendously amplify the impact of this project as it would mean that Experiments 1-3 can be conducted on more realistic events without requiring the humans to wear brightly colored apparel.

8. Overall Statement of Work

The work can be divided into 7 tasks:

1. **Data recording for event learning:** We use videos of up to 3 interacting brightly colored objects as examples for 16 verbs (see section 6.1 for the list of verbs). For this project we will use the existing video, but will need to record more such data for both training and testing. We expect this task to consume **3%** of the total effort.

2. **Data annotation:** Sequester a part of the data as test data and have ground-truth annotated by humans. Frame sequences of each video will be annotated with one of 16 labels or *none*. The events can overlap (e.g. Video #25: *approach* frames[100-145], *give* frames[130-160], etc). We expect this task to consume **2%** of the total effort.

3. **Attentional State and Event Representation:** As described earlier, the choice of attentional state features and event patterns (currently defined as Finite State Machines defined over attentional states) is central to our approach. It's possible that we might have to experiment with the feature set to get better performance. We expect this task to consume **10%** of the total effort.

4. **Unsupervised learning of visual patterns:** Developing algorithms for extracting event patterns. This task is at the center of the project and will occupy the majority of our time. The output will be a large number of visual patterns (FSMs of various sizes), We expect this task to consume **35%** of the total effort.

5. **Mapping word meaning to learned unsupervised patterns:** Develop algorithms for associating words with previously learned patterns given just a few examples of the word. The **deliverable** for this task is a mapping from each word to one or more of the unsupervised patterns. We expect this task to consume **20%** of the total effort.

6. **Testing classification:** Recall and precision of the learned word meanings using the test data. The **deliverable** associated with this is a table with the recall and precision results. We expect this task to consume **10%** of the total effort.

7. **Part-based human detection:** Develop algorithms for unsupervised learning of human appearance and subsequently detecting humans in arbitrary video. As described in experiment 4 - this is a completely independent thread with a view towards tackling figure-ground segmentation head-on and making the project results applicable to a broad class of video streams. If successful - the **deliverable** is a table with recall and precision figures for detecting humans in a set of test videos. We expect this task to consume **20%** of the total effort.

9. Personnel, Qualifications, and Commitments

Professor Patrick H. Winston, MIT PI

Professor Winston has over forty years of experience in Artificial Intelligence and Computer Science. His research group at MIT studies how vision, language, and motor faculties account for intelligence, integrating work from several allied fields, including not only Artificial Intelligence, but also Computer Science, Systems Neuroscience, Cognitive Science, and Linguistics. He brings extensive management and technical expertise.

Education

B.S. Electrical Engineering MIT Cambridge, MA 1965 M.S. Electrical Engineering MIT Cambridge, MA 1967 Ph.D. Electrical Engineering MIT Cambridge, MA 1970

Experience

Professor Winston received his Ph.D. in Electrical Engineering from MIT. His doctoral dissertation introduced ideas on the subject of computer learning from examples and near misses. He has been a Professor at MIT ever since. He was a founding member of the MIT Artificial Intelligence Laboratory, and for 25 years, from 1972 to 1997, he was its director. Professor Winston's publications include 17 books, comprising major textbooks on Artificial Intelligence and various programming languages, several edited collections of key MIT research papers, and an edited collection of papers about the application of Artificial Intelligence to business. Professor Winston has over fifteen years of service (three terms) on the Naval Research Advisory Committee (NRAC), serving from 1985-1990, 1994-2000, and 2003 to the present. From 1997 to 2000 he was chair of the committee. During his service on NRAC he chaired several studies, including a study of how the Navy can best exploit the next generation of computer resources and a study of technology for reduced ship manning. He is a member of the Massport Security Advisory Committee, has served as a member of Defense Intelligence Agency Advisory Board, and is a past president of the American Association for Artificial Intelligence. Professor Winston is chairman and co-founder of Ascent Technology, Inc., a company that develops products that solve complex resource-planning, resource-scheduling, resource allocation, and situation-assessment problems.

Select Publications

- Finlayson, M. A. and P. H. Winston (2006). *Analogical Retrieval via Intermediate Features: The Goldilocks Hypothesis*. Cambridge, MA: MIT CSAIL Tech Reports: #2006-071.
- Winston, P.H. and Narasimhan, S. (2001). *On to Java (3rd edition)*. Boston: Addison Wesley.
- Winston, P.H. (1992) *Artificial Intelligence (3rd edition)*. Boston: Addison Wesley.
- Winston, P.H. and Shellard, S. A. (1990) *Artificial Intelligence at MIT: Expanding Frontiers (2 volumes)*. Cambridge, MA: MIT Press.
- Winston, P. H. (1982). *Learning new principles from precedents and exercises.* Artificial Intelligence 19: pp. 321-350.12

Dr. Sajit Rao, MIT Research Scientist

Dr. Rao has over twenty-three years of experience in Artificial Intelligence; both in basic research as well as developing applications.

His primary research goal is to understand the role of visual processes in cognition and build human-like intelligence in embodied systems. His work is heavily influenced by the fields of Developmental Cognition and Systems Neuroscience.

Education

Massachusetts Institute of Technology

Ph.D degree in Artificial Intelligence, Feb 1998.
M.S. degree in Electrical Engineering & Computer Science, Feb 1991.
B.S. degree in Mathematics, February 1990.
B.S. degree in Brain and Cognitive Science, June 1992.

Key Accomplishments

1. Member of the European Union panel of experts in Artificial Intelligence, and called upon yearly to review E.U. grant proposals, interview teams, evaluate their execution plan, and finally decide which teams get funded (total fund of more than $100 million) every year
2. Inventor on six issued U.S patents in the area of vision.
3. Selected as a Kauffman fellow by the society of Kauffman fellows - a network of venture capitalists present in 18 countries across 140 venture firms that are collectively investing $40B in capital.
4. Won the highly-competitive MIT $50K Business Entrepreneurship competition and launched a company.
5. Built a vision application that processes patient records for the biggest hospital in the Boston area. The system is currently being tested. When deployed it will be used by doctors in the hospital to visually interpret and file patient reports.
6. Designed and launched the flagship product of an educational-software startup where he served as Vice-President of new product development.
7. Successfully executed a DARPA seedling for his research on Visuospatial Reasoning.
8. Wrote a successful 2 million Euro grant application for five research labs across Europe.

Select Publications

- *Learning To Act On Objects* Lorenzo Natale, Satyajit Rao, and Giulio Sandini, in Proceedings of Biologically Motivated Computer Vision (BMCV) November 2002. Springer-Verlag.

- *Learning about Objects through Action: Initial steps towards Artificial Cognition*, P. Fitzpatrick, G. Metta, L. Natale, S. Rao and G. Sandini. In Proc. IEEE International Conference on Robotics and Automation (ICRA 2003,)

- *Development of the mirror system: a computational model.* G. Metta, L. Natale, S. Rao, G. Sandini. In Conference on Brain Development and Cognition in Human Infants. Emergence of Social Communication: Hands, Eyes, Ears, Mouths. Acquafredda di Maratea - Napoli. June 7-12, 2002

- *Repairing Learned Knowledge using Experience* - chapter 14 of the book "*Artificial Intelligence at M.I.T Expanding Frontiers*", Patrick Henry Winston and Satyajit Rao, Volume 1, MIT Press 1990.

10. Facilities

The research described here will be carried out entirely using the facilities described below. No Government Furnished Property is required for conduct of the proposed research, other than the participation of our DARPA Program Manager, Dr. Joseph Olive.

MIT CSAIL Facilities

MIT is a private research university located in Cambridge, Massachusetts. Founded in 1861, it employs over 1,000 faculty members (7 current Nobel-prize winners), and educates over 10,000 students (including over 6,000 graduate students) on its 168-acre campus. The MIT Computer Science and Artificial Intelligence Laboratory (CSAIL) has long been a leader in the fields of Artificial Intelligence, Cognitive Science, and Computer Science. CSAIL is located in the Stata Center on MITs campus, and is home to 95 principal investigators, who include both MIT faculty and research staff; their numbers include five current or former MacArthur fellows, and five Turing award winners. CSAIL consistently ranks near or at the top of undergraduate and graduate Computer Science programs in the world.

CSAIL is equipped with a variety of computers and computational equipment, ranging from PCs and Macs to workstations, Silicon Graphics systems, and a number of machines of our own design (e.g., Lisp machines), along with access to a 400+ node computing cluster within the building. Large, backed-up file servers support the labs large network of local workstations; Internet access is available through the MIT spine, and wireless networking is available throughout the building. CSAIL and MIT provide software licenses for all standard office and productivity software as well as development environments for numerous programming such as Lisp, Scheme, Java, and all Microsoft-supported languages and operating systems,. CSAIL provides an environment of excellent students interested in advanced systems, networking, HCI technologies, knowledge-representation, reasoning, planning, decision-theoretic techniques, and inference.

11. Human use

NONE

12. Intellectual Property

NONE

13. Organizational Conflict of Interest Affirmations and Disclosure

NONE

2 Cost Proposal

BAA: DARPA BAA-08-34 mod. 2

Technical Area: Cognitive Systems

Lead Organization: Massachusetts Institute of Technology

Type of Business: OTHER EDUCATIONAL

Contractor's Ref #: N/A

Other Team Members: N/A

Proposal Title: Perceptual Priming for Language Learning

MIT

Technical POC:
Professor Patrick H. Winston
32 Vassar St., Room 32-251
Cambridge, MA 02138
Phone: 617-253-6754
Fax: 617-258-8682
Email: phw@mit.edu

Administrative POC:
Laureen Horton
Office of Sponsored Programs
77 Massachusetts Avenue, Room E19-750
Cambridge, MA 02139
Phone: 617-253-3922
Fax: 617-258-4734
Email: laureena@mit.edu

Award Instrument Requested: Cost Reimbursement

Place(s): Massachusetts Institute of Technology, Cambridge, MA

Period of Performance: 04/01/10 – 03/31/11

Total Proposed Cost: $380,000

DCAA:
Linda Shipp
Office of Naval Research
703-696-8559
shippl@onr.navy.mil

Date Proposal Prepared: 02/11/10

DUNS:	00-142-5594
TIN:	04-210-3594
CAGE Code:	80230
Proposal Validity Period:	02/11/10 – 7/31/10

PI Name: Patrick Winston
Sponsor: DARPA
Title: Perceptual Priming for Language Learning
Period: 04/01/10 - 03/31/11

	# OF MONTH	Effort %	GRAND TOTAL
PERSONNEL			
Patrick Winston	1 MOS	100.00%	16,270
Sajit Rao	12 MOS	100.00%	100,883
RA PhD(1.5)	12 MOS	100.00%	34,545
Total Salaries & Wages			151,699
Technical and Admin Support - Not MTDC Base			18,298
Employee Benefits			27,531
Employee Benefits - Not MTDC Base			4,300
Vacation Accrual			8,575
Vacation Accrual - Not MTDC Base			1,555
Sub-Total of Fringe Benefits			41,962
TOTAL PERSONNEL COSTS			211,958
OPERATING EXPENSES			
Domestic Travel			2,163
M & S			0
Network Services			5,250
Equipment - Not MTDC Base			7,000
RA Tuition - Not MTDC Base			20,881
TOTAL OPERATING EXPENSES			35,294
TOTAL DIRECT COSTS			247,252
OVERHEAD (F&A)			132,748
TOTAL PROPOSAL COSTS			380,000

MTDC Base	195,218
Allocation Base	159,112

PI Name: Patrick Winston
Sponsor: DARPA
Title: Perceptual Priming for Language Learning
Period: 04/01/10 - 03/31/11

	# OF MONTH	Effort %	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	GRAND TOTAL
PERSONNEL										
Patrick Winston	1 MOS	100.00%	488	325	1,627	5,695	3,254	1,627	3,254	16,270
Sajit Rao	12 MOS	100.00%	3,027	2,018	10,088	35,309	20,177	10,088	20,177	100,883
RA PhD(1.5)	12 MOS	100.00%	1,036	691	3,455	12,091	6,909	3,455	6,909	34,545
Total Salaries & Wages			4,551	3,034	15,170	53,095	30,340	15,170	30,340	151,699
Technical and Admin Support - Not MTDC Base			549	366	1,830	6,404	3,660	1,830	3,660	18,298
Employee Benefits			826	551	2,753	9,636	5,506	2,753	5,506	27,531
Employee Benefits - Not MTDC Base			129	86	430	1,505	860	430	860	4,300
Vacation Accrual			257	172	858	3,001	1,715	858	1,715	8,575
Vacation Accrual - Not MTDC Base			47	31	156	544	311	156	311	1,555
Sub-Total of Fringe Benefits			1,259	839	4,196	14,687	8,392	4,196	8,392	41,962
TOTAL PERSONNEL COSTS			6,359	4,239	21,196	74,185	42,392	21,196	42,392	211,958
OPERATING EXPENSES										
Domestic Travel			65	43	216	757	433	216	433	2,163
M & S			0	0	0	0	0	0	0	0
Network Services			158	105	525	1,838	1,050	525	1,050	5,250
Equipment - Not MTDC Base			210	140	700	2,450	1,400	700	1,400	7,000
RA Tuition - Not MTDC Base			626	418	2,088	7,308	4,176	2,088	4,176	20,881
TOTAL OPERATING EXPENSES			1,059	706	3,529	12,353	7,059	3,529	7,059	35,294
TOTAL DIRECT COSTS			7,418	4,945	24,725	86,538	49,450	24,725	49,450	247,251
OVERHEAD (F&A)			3,982	2,655	13,275	46,462	26,550	13,275	26,550	132,748
TOTAL PROPOSAL COSTS			11,400	7,600	38,000	133,000	76,000	38,000	76,000	379,999
MTDC Base			5,857	3,904	19,522	68,326	39,044	19,522	39,044	195,218
Allocation Base			4,773	3,182	15,911	55,689	31,822	15,911	31,822	159,112
			3%	2%	10%	35%	20%	10%	20%	

MIT/Computer Science and Artificial Intelligence Laboratory
Budget Justification for Cost Proposal

A. Key Personnel:

Prof. Patrick Winston has committed 1 summer month to the project.

MIT fully supports the academic year salaries of professors, associate professors, and assistant professors, but makes no specific commitment of time or salary to any individual research project.

Dr. Sajit Rao is a Research Scientist. He has committed 100% effort (12 months) to the project. A 4% raise is applied each year in January.

B. Other Personnel:

1. Research Assistants (RA)
   1.5 RA's have been budgeted to the project for the period 4/1/10 – 1/15/11, for a total of 14.25 person-months. The RA stipend is not subject to employee benefits. Stipend for 6/1/09-5/31/10 is $2,350/mo for a PhD student. A 4% raise is applied each year in June.

2. Other (Technical & Admin Support)
   The Computer Science Artificial Intelligence Laboratory (CSAIL) provides administrative services for all principal investigators who submit proposals through CSAIL. These administrative services are run by the Headquarter Staff and include Fiscal, Personnel, Facilities and other CSAIL operations.

   These services are supported by an Allocated Project Level Cost, which is assessed against all contracts and grants. The Salary Allocation rate for FY 10 is 11.5%. The Allocation Base is $159,112.

C. Fringe Benefits

(a) Employee benefits are calculated at the rate of 22.0 % and are applied to total salary expenses, less Research Assistants. As of 7/1/2010, the rate will increase to 24.0%.

(b) Vacation accruals are calculated at the rate of 8.5% and are applied to total salary expenses, less Faculty and Research Assistants.

D. Travel

1. Domestic Travel

Two trips have been budgeted for travel to project-related meetings at DARPA for two people per trip. Each person/trip is anticipated to be same-day Boston-DC, and includes shuttle airfare, taxi & food, budgeted at $540.75/person/trip. The total for both trips and all travelers is $2,163. The following has been budgeted per person/trip:

Item	Factor	Type	$/type	**Total**
Airfare	1	Round trip	395.75	**395.75**
Cab	2	Trips	35.00	**70.00**
Food	1	Days	75.00	**75.00**
Grand Total per person/trip		**540.75**		

E. Other Direct Costs:

1. Equipment
   $7,000 has been budgeted for the purchase of two Dell Precision T3500 computer systems. Please reference the attached equipment quote.

2. Computer Services
   MIT/CSAIL has a centralized network services function. The costs for this function are calculated at $175/person/month. The base number of people used for this calculation was one.

3. Tuition
   RA tuition – For academic year '09-'10, MIT 9-month tuition is $37,510. A 4% annual inflator is applied each year. MIT will subsidize 50% of tuition, leaving 50% to be charged to the project. During the summer, MIT has waived tuition.

F. Indirect Costs (Facilities & Administrative Costs):
Effective 7/01/09, F&A Costs are calculated by applying the negotiated rate of 68% to the Modified Total Direct Cost (MTDC) base. The MTDC base includes all direct costs, except Tuition, Subcontracts after the initial $25,000, Network Facilities Charges, Equipment, and Salary Allocation (and associated benefits).

Laureen Horton
Manager, Contracts and Grants

Massachusetts Institute of Technology
77 Massachusetts Avenue, Building E19-750
Cambridge, Massachusetts 02139-4307

Office of Sponsored Programs

Phone 617.253.3922
Fax 617.253.4734
Email laureena@mit.edu
http://web.mit.edu/osp/www/

David Frey CONTR-IPTO
ATTN -BAA-08-34
David.Frey.ctr@darpa.mil

2 September, 2009

RE: Proposal Entitled "Perceptual Priming for Language Learning" under DARPA BAA 08-34

Dear Mr. Frey:

Massachusetts Institute of Technology (MIT) submits herewith a request for capital equipment for the above referenced proposal. The Institute hereby certifies that it is unwilling and or unable to use Institute funds to acquire any of the equipment specified in the proposal. The equipment budget for the proposal is comprised of the following pieces:

- 2 Dell Precision Workstation T3500 at $3,473 each

Please see attached quotes for references and specific components of each particular piece of equipment. Please contact Prof. Patrick Winston at phw@mit.edu regarding any technical aspects of this request. Questions of an administrative/contractual nature, please contact the undersigned.

Sincerely,

Laureen Horton

cc: Prof. Patrick Winston
Ms. Karen Shirer

Windows® . Life without Walls™ . Dell recommends Windows 7.

Windows® . Life without Walls™ . Dell recommends Windows 7.

Print Summary

Precision T3500 64bit

Starting Price	$3,714
Instant Savings	$241
Subtotal	**$3,473**

Quickly create your own professional-looking custom forms, such as customer estimates, invoices and reports by using any of the over 100 included templates. Upgrade to Quickbooks Software Now!

As low as **$87**/mo.

Dell Business Credit | Apply

Discount Details

Preliminary Ship Date: **2/19/2010**

My Selections **All Options**

- **Precision T3500 64bit**

Date	2/10/2010 7:43:29 PM Central Standard Time
Catalog Number	4 Retail 04

Catalog Number / Description	Product Code	Qty	SKU	Id
Dell Precision T3500: Dell Precision T3500, CMT, Standard Power Supply	T3500	1	[224-4422]	1
Operating System: Genuine Windows® 7 Professional Bonus 64- Windows XP Professional downgrade	PW7PXD6	1	[421-1993] [468-4322]	11
Energy Efficiency Option: No Energy Star	NOESTAR	1	[330-3201]	25
Processor: Quad Core Intel® Xeon® W3570 3.20GHz, 8M L3, 6.4GT/s Turbo	W3570	1	[317-0127]	2
chassis configuration: Mini-Tower Chassis Configuration	MT	1	[311-7463]	15
Memory: 12GB, 1066MHz, DDR3 SDRAM, ECC (6 DIMMS)	12G3E66	1	[317-0111]	3

Hardware Support Services: 3 Year Basic Limited Warranty and 3 Year NBD On-Site Service	U3YOS	1	[992-9022] [993-3120] [993-9008] [993-9017]	29
Graphics: 256MB ATI FireMV® 2260, 2MON, 2 DP w/ 1 DP to DVI Adapter	ATI2260	1	[320-1551]	6
Security Software: Trend Micro Internet Security 30-day trial, English	TREN30E	1	[410-2395]	37
Hard Drive Configuration: C1 All SATA drives, No RAID for 1 Hard Drive	SATA1	1	[341-8562]	9
Hard Drive Controller: Integrated Intel chipset SATA 3.0Gb/s controller	NSASCTL	1	[341-9289]	24
Boot Hard Drive: 1TB SATA 3.0Gb/s,7200 RPM Hard Drive with 16MB DataBurst Cache™	1TBST	1	[341-8997]	8
CD-ROM, DVD and Read-Write Devices: 16X DVD-ROM with Cyberlink Power DVD™	DVD16	1	[313-7458] [421-0536]	16
Monitor: No Monitor	NMN	1	[320-3316]	5
Floppy Drive and Media Card Reader Options: No Floppy Drive and No Media Card Reader	NFD	1	[341-5255]	10
Resource DVD: Resource DVD - contains Diagnostics and Drivers	RDVD	1	[330-4025]	27
Quick Reference Guide: Quick Reference Guide, English	REFE	1	[330-4020]	39
Shipping Packaging Options: Shipping Material for System	SHIP	1	[330-3209]	40
Windows 7 Upgrade Program Info: Windows 7 Upgrade Web Site	WIN7UP	1	[468-3168]	461
Speakers: No Speaker option	NSPKR	1	[313-2663]	18
Keyboard: Dell QuietKey Keyboard	QUSB	1	[330-3203]	4
Mouse: Dell USB 2 Button Optical Mouse	USBO	1	[330-3945]	12
Documentation: Documentation, English, with 125V Power Cord	DOCENG	1	[330-3156] [330-3157]	21

Print

*Offers subject to change. Taxes, shipping, handling and other fees apply. U.S. Dell Small Business new purchases only. LIMIT 5 DISCOUNTED OR PROMOTIONAL ITEMS PER CUSTOMER. LIMIT 5 VOSTRO UNITS PER CUSTOMER. Dell reserves right to cancel orders arising from pricing or other errors.

snFG10