12th ICCRTS: "Adapting C2 to the 21st Century"

Towards Commanding Unmanned Ground Vehicle Movement In Unfamiliar Environments Using Unconstrained English: Initial Research Results

(C2 Technologies and Systems, Cognitive and Social Issues, C2 Metrics and Assessment)

Robin Burk, Assistant Professor (POC)
Department of Electrical Engineering & Computer Science
U.S. Military Academy
Bldg. 601
West Point, NY 10996
845 938-3987
DSN 688-3987
Robin.Burk@usma.army.mil

Andrew Haas, Associate Professor
Department of Computer Science
College of Computing and Information
State University of New York, University at Albany
1400 Washington Ave. · Albany, NY 12222
518 442-9066
haas@cs.albany.edu

Frederick I. Moxley
Defense Information Systems Agency Fellow & Visiting Professor
Department of Electrical Engineering & Computer Science
U.S. Military Academy
Bldg. 601
West Point, NY 10996
845 938-5568
DSN 688-5568
Frederick.Moxley@usma.army.mil

Kevin Knuth, Assistant Professor
Physics Department and Information Science Program
State University of New York, University at Albany
1400 Washington Ave. · Albany, NY 12222
518 442-4653
kknuth@albany.edu

MAJ Benjamin Ring, Instructor
Department of Electrical Engineering & Computer Science
U.S. Military Academy
Bldg. 601
West Point, NY 10996
845 938-5571
DSN 688-5571
Ben.Ring@usma.army.mil

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2007		**00-00-2007 to 00-00-2007**

4. TITLE AND SUBTITLE
Towards Commanding Unmanned Ground Vehicle Movement In Unfamiliar Environments Using Unconstrained English: Initial Research Results

5a. CONTRACT NUMBER
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Military Academy,Department of Electrical Engineering & Computer Science,Bldg. 601,West Point,NY,10996

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Twelfth International Command and Control Research and Technology Symposium (12th ICCRTS), 19-21 June 2007, Newport, RI

14. ABSTRACT
Sensemaking in the 21st century C2 environment will be critical not only for soldiers but also for autonomous equipment. Sensemaking by humans entails understanding the meaning and import of information, often conveyed via natural human language, about events and objects in the battlespace. Analogous sensemaking in autonomous and semi-autonomous UGVs requires cognitive robotics, i.e. the ability to link human language and concepts to robot perception and object recognition. Advanced sensemaking in UGVs would allow soldiers to send such equipment through urban environments using the same verbal instructions they would give another soldier. A robust natural language-based sensemaking capability in UGVs could also contribute information about the battlespace to the Global Information Grid while requiring few or no services in return. Recent work by Haas and Shimizu has demonstrated the ability of a simulated robot to respond correctly and without additional guidance to naively-produced navigational commands (expressed in unconstrained English) with ~80% accuracy. Our current work extends this approach to natural language processing into physical robots, introducing uncertainties of sensor perception, object recognition and language-to-environment mapping. The goal of this research is to quantify accuracy for a simple indoor environment and then more complicated environments, characterizing sources of error and identifying strategies to reliably overcome them.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**36**	

ABSTRACT

Sensemaking in the 21st century C2 environment will be critical not only for soldiers but also for autonomous equipment. Sensemaking by humans entails understanding the meaning and import of information, often conveyed via natural human language, about events and objects in the battlespace. Analogous sensemaking in autonomous and semi-autonomous UGVs requires cognitive robotics, i.e. the ability to link human language and concepts to robot perception and object recognition. Advanced sensemaking in UGVs would allow soldiers to send such equipment through urban environments using the same verbal instructions they would give another soldier. A robust natural language-based sensemaking capability in UGVs could also contribute information about the battlespace to the Global Information Grid while requiring few or no services in return.

Recent work by Haas and Shimizu has demonstrated the ability of a simulated robot to respond correctly and without additional guidance to naively-produced navigational commands (expressed in unconstrained English) with ~80% accuracy. Our current work extends this approach to natural language processing into physical robots, introducing uncertainties of sensor perception, object recognition and language-to-environment mapping. The goal of this research is to quantify accuracy for a simple indoor environment and then more complicated environments, characterizing sources of error and identifying strategies to reliably overcome them.

Introduction

Network Centric Warfare is already a reality, in nascent form. A case study of Operation Iraqi Freedom performed at the U.S. Army War College [1] concluded that

> "the introduction of extended reach communications and networked information technologies significantly enhanced the ability of U.S. Army commanders to make faster decisions, more easily exploit tactical opportunities, conduct coordinated maneuver while advancing further and faster than at any previous time and more fully integrate and synchronize joint fires; all of which resulted in the rapid defeat of Iraqi military forces and the fall of the Ba'athist Regime in Baghdad."

The study describes an effective synergy between networked sensors (including Hunter UAVs, Predator UAVs and the Long Range Advanced Scout Surveillance System) and the Automated Deep Operations Coordination System which provided a common operational picture to commanders. Together with voice communications and enabled by the wideband TACSAT, the unmanned systems had significant tactical and operational level impacts [2].

As the Army's Future Combat Systems components mature, the number and nature of unmanned ground systems in the battlespace will evolve rapidly. Already, in recent years while the Hunter and Predator UAVs were bringing sensor data into the common operational picture in OIF, their ground equivalents were proving highly useful in the caves of Afghanistan. New UGVs continue to play an important role in the on-going mission in Iraq today.

As the Network Centric Operations Conceptual Framework notes [3], collecting and sharing information does enable shared situational awareness. That is not, however, the broadest or deepest benefit to be gained. Shared awareness in turn enables other benefits, including shared sensemaking and the ability to substitute information for people and material. We might add to that a future ability to substitute autonomous systems for humans in some circumstances and for some purposes.

The U.S. Army's Future Combat Systems program envisions missions for UGVs that go beyond remotely-operated data collection. Intelligent munitions, robotic mules that carry soldiers' gear, autonomously navigating trucks that bring vital supplies and a variety of other robotic equipment will find their place on the networked battlefield. As these and other robotic systems proliferate it will be increasingly important to consider how they will be integrated into battlefield operations at the cognitive and social domains as well as at the level of the physical network and information gathering and sharing.

The War College's OIF case study notes that voice communications were key to developing shared situational awareness. Commanders benefited from the availability of real-time, extensive information collected by sensor platforms and fed into a common operational picture of ongoing events in the battlespace. But those commanders also *made sense* of the implications of that information in part through verbal communications. That's not surprising: speech is the most natural way for humans to share and interpret information.

The NCO CF posits that enhanced situational awareness and understanding of the information that is collected can and should, in turn, lead to more agile force elements and overall enhanced mission effectiveness [4]:

> What makes network-centric forces more effective? The answer that is emerging is twofold. First, mission effectiveness is greatly enhanced by *agility*: the ability to be quick and nimble; the ability to be adaptive and responsive to changing circumstances; and, the ability to innovatively solve problems. …. Second, agility is possible only if we accept that, *"Network Centric Operations is not about technology, it's about people!"* The most impressive gains in force effectiveness resulted from a synergy of investments across the lines of development (technologies plus training, leadership, organizational change, etc.).

This is a useful reminder when we think of unmanned systems in particular. Ultimately, it is not the significant technical challenges inherent in developing autonomous and semi-autonomous battlefield vehicles and their payloads that must dominate our attention, nor the complex work of integrating them technically as users and sources for the Global Information Grid. Rather, the more fundamental question regarding unmanned systems is how they will be designed and deployed to further facilitate agile, effective operations.

Soldier – UGV Interfaces: The case for natural language

We believe that an important aspect of useful unmanned systems is the interface through which soldiers must interact with and use them. The ideal UGV would be much like the ideal soldier: able to receive commands, interpret them intelligently, execute them reliably, ask questions when something is not clear and alert someone when unexpected or significant events occur. The more natural the means of communication, the less training required for soldiers and the easier it is for commanders to leverage UGVs as another element in the unit, as a truly organic capability.

In general, the most natural means with which to command an element on the battlefield is human speech – English, or whatever other language the unit members speak and understand well. There are, however, a range of useful ways to command UGVs that are simpler to implement than natural language understanding and which are appropriate for many valuable tasks.

For instance, voice recognition of a carefully constrained list of command words could be valuable in many systems, as would the ability of driverless vehicles to navigate through a route specified by GPS coordinate waypoints. Each of these is already a maturing capability. Voice recognition of constrained commands is now commonplace in the civilian world, although in some cases command execution is non-trivial (as with NASA's work towards voice commanded / voice output information lookup to support astronauts doing complex repairs in space). Although autonomous navigation is currently less mature than voice recognition, each successive DARPA Grand Challenge has demonstrated greater success as UGVs find their way towards that year's destination.

For many unmanned systems, simple voice recognition capabilities or the use of a console to input navigation waypoints will suit the mission well. However, these interfaces do impose some limitations. For instance, both constrained voice command recognition and waypoint-based navigation require prior planning before they can be used. Command lists must be drawn up and corresponding actions programmed into the equipment, producing a static set of actions which may be selected among. Waypoints must be mapped with precision if they are to guide navigation. These capabilities, then, will best fit unmanned vehicles intended for well-defined repeated tasks (voice commands) or for environments that are familiar to some degree (GPS waypoint navigation). In addition, personnel must be trained before they can operate systems through built-in touch screens or keyboards or a limited set of voice commands.

The constraining effect of these interfaces for unmanned systems is most obvious with regard to systems that will be tactical in nature and that most naturally are associated with small unit activities. If they are to support the tactical mission well, these UGVs require a human-to-machine interface that is more flexible, more powerful and that can be applied to a wide variety of tactical situations and environments while entailing a limited training burden. They must, in other words, be sufficiently intelligent to be extremely easy to deploy as part of small unit activities in a wide variety of circumstances.

Consider the value of being able to send small UGVs to perform tasks in response to the directions one would give a human soldier:

> "Go down this road to the first cross-street. Turn left, go two blocks and then turn right. Stop in front of the second building. Radio if you see any white vehicles parked along the streets as you patrol. Radio if you believe you may have found any unexploded ordinance or IEDs along the roadside."

> "Go to the second house ahead on the right. Enter the door, go up the stairs to the third floor."

Scouting through a hostile neighborhood. Delivering ammunition, food or medical supplies to soldiers under fire. These are tasks that have historically been executed by soldiers but that might well be assigned to autonomous UGVs at some point in the foreseeable future. The ability to send UGVs to perform these kinds of tasks using natural language will enhance the agility of the units they serve, allowing them to quickly respond to changing circumstances and facilitating creative responses to problems as they are encountered. Moreover, in stressful combat and near-combat situations it is a significant advantage if soldiers need not remember artificial means of using their equipment, but rather can fall back on the linguistic capability they have used for most of their lives.

UGVs with these capabilities would indeed "empower the edge". There are, however, significant hurdles to overcome before they can be deployed.

The challenges: natural language, cognitive robotics and GIG interface

Before squad leaders can send their UGVs off with a few terse directions to do autonomous reconnaissance, progress must be made in three areas.

First, we must be able to construct software that can interpret directions given in unconstrained English (or other natural language). This is the natural language processing challenge

Second, we must be able to construct robotic equipment that can recognize objects in the environment and we must be able to link that recognition to the object attributes that humans are likely to reference when giving directions. This is the cognitive robotics challenge.

And third, we must consider the degree to which it is necessary or desirable for UGVs with natural language interfaces to interact with the Global Information Grid.

On the one hand, it would be ideal if small-unit tactical UGVs were capable of processing and responding to command sequences using their own computational power most or all of the time. This would allow deployment of many such systems in a battlespace without over-burdening the communication network and the Global Information Grid repeatedly during unit operations.

On the other hand, a UGV that can interpret and execute spoken navigation directions of the sort listed above is more than a sensor platform – it is a significantly intelligent application in its own right. Of necessity it must be capable, not only of fusing information from its own sensors in order to navigate, but also of analyzing and interpreting that information in order to recognize objects and their attributes as described by human language. In other words, it must *make sense* of the sensor data it is collecting, understanding the import of that information for execting the mission it has been assigned. And in the course of sensemaking for its own purposes, it may well be generating information and understanding that would be of use to other automated systems and to humans.

The natural language processing challenge

Natural language processing (NLP) has been a goal of artificial intelligence research for decades [5]. Results have been slow coming, however. There are several reasons for this.

First, natural languages are complex, with large vocabularies and variable syntax. Moreover, people often cut grammatical corners when they speak, making spoken language even harder to parse than written texts.

Second, language is often ambiguous, metaphorical or idiomatic, making semantic interpretation a difficult task for literal-minded software. "Run that by the commander." "Hang a right at the corner." "We're going to slow roll this one." "I am, like, sooooo dead when Sarge finds out …." Plus, as any parent of teenagers knows, languages like English add idioms and metaphors easily, baffling the uninitiated.

These characteristics have presented significant barriers to full syntactic and semantic analysis of natural language by software. Although many computational linguists continue to chip away at this problem, major breakthroughs do not seem to be on the immediate horizon.

Ambiguity and often errors in spoken language aren't problems for software alone. In one recent NLP effort, Macmahon and his colleagues set up a simple virtual indoor environment and asked experimental subjects to write directions for a trip from a given starting place to a given destination. Out of 786 examples collected from 6 subjects, other human beings could only reach the correct destination by using the directions 69% of the time [6]. In this experiment, the majority of failures were due to clear-cut errors in the directions: saying 'left' where 'right' was intended, for instance. As we will see below, other researchers have found more fundamental sources of potential error and ambiguity in navigation direction giving.

Constraining language helps somewhat, but doesn't remove the problem. Although the services invest considerable training time teaching specialized vocabularies relating to military matters, and structure communications in predictable formats such as op orders, anecdotal evidence suggests that ambiguity in natural language persists even in the context of military operations and must be overcome by verbal interaction (verifying correct understanding of information or orders, asking for clarification) or through maps and other visual aids. Although we are not familiar with any rigorous studies of the issue, it is likely to be the case that requests for clarification occur as part of sensemaking and not primarily due to difficulties with linguistic processing per se. Native speakers of a language generally are fluent at parsing grammar and have extensive vocabularies – their difficulties arise due either to ambiguous wording or to a perceived mismatch between the other speaker's statement and the assumptions and information that the hearer had previously acquired about the topic at hand. Thus we are not surprised that the War College case study of OIF found voice communications to be a critical component of the network centric operations in Iraq. Clarifying and verifying information in this way would have enabled more rapid and more confident sensemaking, resulting in more rapid and confident decisions and execution.

All is not hopeless with regard to software and natural language processing, however. There are several areas of progress in this regard. For instance, considerable success has been achieved in text summarization, query answering and document retrieval by constructing indices that relate terms and phrases statistically. The most familiar such system for most people is probably a Web search engines such as Google. National security and intelligence community members will be familiar with other examples as well.

Information retrieval (IR) and text summarization approaches are powerful ways to find information of interest but as with most technologies they have limitations. First, they require large corpuses of reference documents from which to establish statistical correlations. Second, as the name implies, information retrieval software doesn't so much interpret natural language as it characterizes and retrieves documents containing it. Indeed, some attempts to improve information retrieval by augmenting queries using semantic databases such as Princeton University's WordNet have resulted in degraded performance for many search techniques rather than in the hoped-for improvement.

Information retrieval approaches function best with a (potentially considerable) degree of interactivity between the software and the human user. Search engines suggest different search terms, allow users to identify more- and less-relevant results from initial searches and otherwise use feedback to refine the program's ability to find the desired text.

IR approaches are unpromising for natural language processing in UGVs and other autonomous and semi-autonomous systems for several reasons. First, they require truly huge collections of texts and large indices, imposing very large hardware requirements. Second, at heart they are suited not to understanding language so much as to finding relevant pre-written language in response to user queries. And third, they accomplish these tasks primarily through statisitical correlation that is generally void of most (or any) semantic understanding of the language involved.

Other current attempts at artificial intelligence for NLP utilize formal semantic frameworks such as ontologies both to describe linguistic mechanisms and also to guide automated translation and summarization of documents. These techniques show some promise for those applications, but again are unpromising for NLP as the interface for UGVs on the battlefield.

The most promising approach for our purposes emerges, not from traditional linguistic study of syntax (in particular) nor from the large corpus-oriented world of search engines and text summarization. Instead, it begins with the simple observation that we want UGVs to interpret natural language *in order to do some important task as a result of that language*. In other words, our primary goal in UGV NLP is to connect words to objects and actions in the real world. In linguistics, this is the sub-discipline called pragmatics.

Focusing our attention on pragmatics simplifies the NLP task in several ways. We don't need to interpret all possible constructs in English, only those likely to be produced as imperative sentences in a particular context. We will, however, also be interested in cognitive linguistics, i.e. in how a speaker's sentences reflect his assumptions and understanding of the world.

Initial research results for direction following

The potential utility of adopting a pragmatics focus on NLP for unmanned vehicles is suggested by the initial success of one of us (Haas [7]) and his doctoral student Shimizu [8]. The setup for both approaches was similar. Experimental subjects were presented with a simple layout of a building interior, marked with icons indicating "the robot is here" and "destination". They generated written directions for the simulated robot to follow in order to reach the destination. Shimizu applied rule-based heuristics and machine learning techniques to interpret directions. Haas, on the other hand, limited his language processing to extracting a limited number of relations expressed in the directions.

Haas' results are noteworthy for the significant accuracy achieved with a very simple pass through the direction sets. Out of 865 sets of directions written by 89 subjects, and tested against 218 sets of directions written by 22 new subjects, the program correctly interpreted the directions and reached the destination 79% of the time. (The experimenter also tried to follow the directions and only those which he or another native English speaker agrees are adequate were counted among the successes.)

These results were reached without any secondary requests for clarification and without reference to syntactic or semantic models other than the basic language and world familiarity that identified the key relations to be extracted. Each step can be characterized in terms of:

- The type of destination for this step (doorway, side hall, end of hallway)
- Direction (left, right, forward)
- An ordinal characteristic for the destination (first, second, third, last)
- At-end (true if the destination for this step is the end of the hallway ahead of the agent as it begins the step)
- The action required for this step (advance, advance and turn, or do nothing until the next step)

Despite the wide range of potential grammatical constructs available in English for the purpose, the subjects tended to use a limited number of constructs when giving directions. However, the software needed to keep track of some meaning beyond individual phrases due to the narrative flow of some direction sets. For instance, subjects sometimes will say "Go forward until you reach the second door on the right. Turn right." This illustrates the need to fill in implicit references, in this case that "turn right" means "turn right at the second door on the right". Similarly, "You will see a door on your left. Go inside." means "Go in the door on your left."

Both Haas' approach to interpreting the directions and the machine-learning approach of Shimizu are built on the fact that successive steps in the directions depend on the agent's position and orientation as the step begins. This matches the extensive literature on first-person orientation in navigation by humans.

The most common problem in directions resulted from ambiguity about ordinals. For instance, Figure 1 appears to be straightforward:

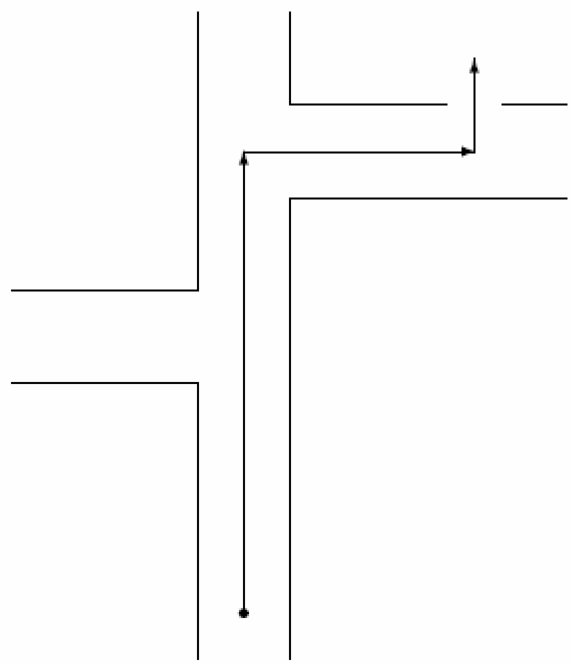


Figure 1: "Make a right at the second hallway"

Some subjects, however, say things like "Make a right into the first hallway" in order to accomplish this movement. The desired hall isn't the first encountered by the simulated robot, but it is the first one on the right.

How should the program treat these two commands? If it insists on linking the direction of the intended turn with the count of halls, then the initial example will not execute correctly – the desired turn is not into the second hallway on the right, but rather the first hallway on the right. On the other hand, if the program ignores this issue, it will correctly interpret the first example but not the second.

Thus even this simple experiment illustrates the need for context and clarification, beyond basic relation extraction, to correctly interpret directions in all cases. In the results for this stage of the research, the two potential errors are about equally common so for the program assumed that we are counting all hallways when we say "second" or "first" hallway in a step.

Ordinal numbers occur frequently in this experiment and other ambiguities arise with ordinals as well. For instance, although the program interprets ordinal numbers as in Figure 2 below, with the arrow representing the robot agent, there is inherent ambiguity regarding paths like the one shown in Figure 3. Eleven subjects consider B to be "the first door on the left", but three others considered it to be the second such door. One subject made an extensive attempt at disambiguation:

> Turn left into the second room on your left side (the first room being the one at the corner of the hall where you just turned right)

Correctly interpreting phrases like "the hall where you just turned right" requires a more sophisticated mapping of the building space and interior objects than was implemented in the simulation for this first set of experiments, so unfortunately this careful exegesis was utterly ignored during program execution.

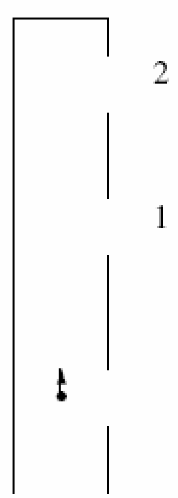


Figure 2: Semantics of Ordinal Numbers

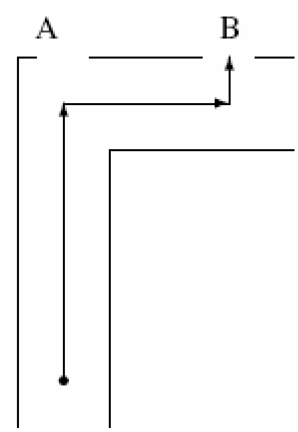


Figure 3: First Door or Second?

The experiment results were matched against a test corpus of 218 sets of directions. A native English speaker attempted to follow the directions and found the correct destination in 134 out of 137 attempts. By breaking down the directions into individual steps and manually extracting the relations listed above, Haas was able to establish an overall success rate for the system of 79%, with a 91% success rate for the back end portion of the program which executes the extracted steps. Haas concludes that a simple parser was needed to eliminate overly-simplistic extraction of relevant phrases, a conclusion that agrees with other literature on relation extraction. By way of comparison, Macmahon's program produced significantly poorer results (61%) despite the fact that he preprocessed all instructions in an attempt to clarify their syntax

Moving into a real robot: the cognitive robotics challenge

Haas' initial results for processing natural language navigation directions intentionally used a simple simulated environment, so as to establish a baseline for language-only performance. The next step we are taking is to replicate his experiments in a real robot traversing a physical environment of similar layout and complexity.

The simulated robot perfectly knows the location of objects (doors, hallways) and moves to them flawlessly based on the language interpretation. Real robots introduce variability and imperfection in sensing, in object recognition and in moving to desired locations. This is true no matter which of several possible algorithms are used for vision processing and object recognition. Thus, if we make no changes to Haas' NLP code we can expect that the accuracy of the same approach within a physical robot is likely to be less than that the same code executed via the simulation.

Nonetheless, the choice of reasoning approach used to identify objects is important. Many robots achieve object recognition by training artificial neural nets or other software with large training datasets. A robot presented with several hundred, or thousand, images labeled 'door' extracts patterns by means of which it classifies new images as 'door' or 'not door'.

If our primary purpose were door recognition for its own sake (in, say, a dedicated security robot used to verify that all interior doors are fully shut at night), this would be an effective approach. Our problem is somewhat different, however. We want to identify doors *as they are described by humans when giving navigation or other task directions*. That is to say, we want our door identification to be based on the attributes that the direction giver chooses as salient for the context. In keeping with our overall intent, the robot will not learn or map its environment prior to being given directions – it must construct its map as it draws conclusions about the presence of objects it encounters along the way.

This is the challenge of cognitive robotics: to make a connection between a conceptual description of objects and sensory perception. In many cases, artificial neural nets that are trained to recognize faces or do similar tasks do so on the basis of often-complex relationships they've extracted that do not make intuitive sense to humans. Cognitive robotics, on the other hand, builds on the theory of dual channels in human cognition: a symbolic, 'logical' channel that reasons about objects and a probabilistic 'connectionist' channel that processes vision rapidly and unconsciously.

Our current work emphasizes the logical side more than behavior-oriented robotics, but we are also very concerned with the impact of sensory processing. We address the problem of object recognition by mimicking the probabilistic approach that neuroscience has begun to suggest characterizes human reasoning as well [10]. Bayesian belief nets are being constructed to map robot perception patterns to objects and their attributes. Beliefs about the identify and location of objects are updated in response to new information perceived as the robot moves through the environment.

As is true in humans, vision processing plays a large role in object recognition for robots. Color interacts extensively with shape in human visual processing [11]; we will be interested to see how much our new experimental subjects refer to color when giving directions. (Color was not an attribute of objects in the simulated environment.)

Thus we are collecting two sets of metrics in the robot in the current phase of research: one set that characterizes the variability of robot perception and movement in response to software controls and another set that establishes the likelihood of a given object being the cause of specific sensor input values that have been received. Our aim is to identify the nature and degree to which various causes contribute to failure by the robot to reach destinations in response to the same sets of directions produced in Haas' initial results. We will also collect the results of new directions given by subjects who have no experience with the simulation and who work only with the physical robot.

Scaling up to more complex environments

The first step in this research program was Haas' and Shimizu's work with unconstrained English directions for commanding a simulated robot through a simple indoor environment.

The second step is under way as this paper is being written, namely to replicate Haas's approach in physical robots using a Bayesian approach both in terms of characterizing variability of sensor perception and vision processing from a Bayesian statistical data analysis perspective and also in terms of nets of Bayesian inferences for object identification. We chose the Bayesian approach because it best fits how we understand humans to draw conclusions about the likely identity of objects we perceive: we adjust our belief as new information is received (as we grow closer to the object, for instance).

The third step in our research efforts will be to introduce more complicated environments which will require more complex references, vocabulary and relations in the directions required to command the robots to their desired destinations. Here our question is one of scale up. How much additional complexity is required in the relations extracted from the directions as a result of more classes of objects, more attributes used to identify those objects and the presence of landmarks which, while not destinations of their own, are likely to factor into many subjects' direction giving?[9] We are also interested in measuring code size and computational load changes.

A fourth area of interest has to do with variations in typical directiongiving on the part of those for whom English is not a native language. Istvan Kecskes, a linguist who merges pragmatics and a cognitive approach, points out that people who are learning a second language often go through an extended stage in which they understand basic syntax and semantics but don't "think" in the new language yet. [12] As a result they make typical mistakes in generating sentences and have an imperfect grasp of metaphors and idioms commonly used in that language. Kecskes believes that the influence of language and thought is mutual: language reflects the speaker's concepts and assumptions, but learning a new way of speaking can reshape those concepts to the point that there are noticeable changes in how he uses his *original* (native) language.

Thus people learning a second language must develop sensitivity to the implications of word and phrase choice in that new tongue. This phenomenon is a familiar one to military trainers, whose work consists in part in conveying new terms and concepts and then teaching soldiers to think using them.

Sensemaking in a new language is in many ways similar to sensemaking on the battlefield. Therefore results with non-native speaking directiongivers may shed light on the issues that can arise during joint operations among allied militaries in a networked battlespace, whether or not it includes autonomous equipment.

Finally, while we have no plans at the moment to pursue this issue in our own efforts, we note that research documents gender differences regarding navigation strategies in virtual environments [13]. These gender differences mirror differing facility with spatial orientation vs. verbal fluencies. As unmanned vehicle designs mature, it may be prudent to test both natural language and other interfaces against a diverse user base before proceeding to implementation.

Conclusions

Although we are not yet capable of producing UGVs that can correctly interpret and follow complex sets of unconstrained natural language navigation directions, we believe that the ability to field such equipment would enable the kind of the agility at the small unit echelons that contribute to enhanced mission effectiveness. Moreover, fielding such equipment may be possible without placing significant demands on battlefield communications networks or Global Information Grid services. Indeed, intelligent UGVs capable of sophisticated object recognition and probabilistic reasoning may contribute useful information to other GIG users.

For these reasons, and in response to the significant results achieved by Haas and Shimizu in their experiments based on simulated robots, we believe that our research program is of use to network centric operations in two ways. First, it will produce detailed metrics regarding the reliability of Haas' natural language processing approach when implemented in a physical robot which must be commanded through increasingly complex environments, identifying sources of error for which cost-effective responses (such as limited parsing or requests for clarification) can be added to the baseline system. And second, this research has the potential to shed light on language- and sense-making issues that may emerge as the result of joint operations among allied militaries as they bring different linguistic experiences, doctrinal assumptions and personal fluencies to the networked battlefield.

REFERENCES

[1] Tisserand, et. al. *Network Centric Warfare Case Study. U.S. V Corps and Third Infantry Division during Operation Iraqi Freedom Combat Operations Combat Operations (Mar-Apr 2003). Volume III: Network Centric Warfare Insights*. U.S. Army War College, Carlisle Barracks, PA. August 2006. Pg. viii.

[2] ibid, pg. 10

[3] Garstka, J. and Alberts, D. *Network Centric Operations Conceptual Framework Version 2.0.*

[4] ibid, pg. 5.

[5] Winograd, T. *Understanding Natural Language*. New York: Academic Press, 1972.

[6] Macmahon, M. et. al. "Walk the Talk: Connecting Language, Knowledge and Action in Route Instructions". *Proceedings of the 21st National Conference on Artificial Intelligence*. 2006.

[7] Haas, A. tbd (journal articles in review at time of this paper).

[8] Shimizu. Noboyuki. *Understanding Navigational Route Instructions.* Doctoral dissertation, University at Albany, State University of New York, 1996.

[9] Burnett, G. et al. "Supporting the Navigation Task: Characteristics of 'good' landmarks." *Contemporary Ergonomics*, 2001.

[10] Rao, Olshausen and Lewicki, eds. *Probabilistic Models of the Brain: Perception and Neural Function.* MIT Press, 2002.

[11] Barnes, N. and Liu, Z. "Knowledge-Based Shape from Shading" *International Journal of Pattern Recognition and Artificial Intelligence*. Vol 3, No 1, 1999. pp 1-24.

[12] Kecskes, I.. et al.. *Foreign Language and Mother Tongue*. Mahwah, NJ: Lawrence Erlbaum. 2000. See also Kecskes, I. Editor in Chief. *Intercultural Pragmatics*. Mouton de Gruyter.

[13] Cutmore, T. et. al. "Cognitive and gender factors influencing navigation in a virtual environment." *International Journal of Human-Computer Studies (2000),* No 53, pp. 223-249.

AUTHORS

Robin Burk joined the faculty of the U.S. Military Academy in 2001 and has taught in both the Electrical Engineering and Computer Science department and the Systems Engineering department. Prior to coming to USMA she spent 25 years in industry as a software engineer, project manager, program manager, and senior executive in early- and mid-stage high technology companies serving both defense and commercial markets. She holds a bachelor's degree in physics from St. John's College, an MBA from the University of North Carolina at Chapel Hill (finance and operations) and is currently completing a Ph.D. in Information Science at the University at Albany, State University of New York. Her research interests include artificial intelligence for natural language understanding, cognitive robotics, knowledge representation and management, and metrics for characterizing complex intelligent software systems.

Andrew Haas is an associate professor in the Computer Science department at the University at Albany, State University of New York. He holds a Ph.D. in computer science from Rochester University. His research interests are in several branches of artificial intelligence, including syntax, semantics and parsing of natural language, knowledge representation, and planning.

Frederick Moxley is the first Defense Information Systems Agency Fellow & Visiting Professor appointed to serve at the United States Military Academy. Previous assignments include Technical Director of the Intelligent Software Engineering Environment Laboratory initiative, Deputy Director of the Ada Joint Program Office, DoD Intelligence Information Systems Liaison Officer, and Chief of the Center for Joint & Coalition Interoperability. He holds a Ph.D. in information systems and science from Nova University and is a Ph.D. candidate in biodefense (biomedical/engineering) at George Mason University. His academic interests include systems interoperability, software architectures, net-centricity, and molecular and microbial biodefense related research.

Kevin Knuth is an assistant professor in the Physics department at the University at Albany, State University of New York, where he also serves as faculty for the Information Science PhD program. Prior to joining SUNY, he was a research scientist at NASA Ames Research Center in the Intelligent Systems Division, where he developed machine learning techniques for autonomous systems and robots. He holds a Ph.D. in physics from the University of Minnesota with a minor in mathematics and held postdoctoral appointments in neuroscience His research interests include information physics, signal processing and source selection, intelligent instruments and robotics, and the underlying mathematical disciplines that support them.

Ben Ring is a major in the U.S. Army and currently instructs in the Electrical Engineering and Computer Science department at the United States Military Academy. An armor officer, he has deployed in support of Operations Joint Endeavor (Bosnia, 1997), Enduring Freedom (Kosovo, 1999), and Iraqi Freedom (2004). He holds a B.S. degree in computer science from USMA and an M.A. degree in computer science from Boston University. His research interests include graphics, image & video processing and computer vision. In addition to his academic responsibilities, Ben is head coach and officer-in-charge of the USMA judo team (currently the three-time Eastern Collegiate Champions).

This abstract, the full paper and the research they describe are unclassified.

Towards Commanding UGV Movement In Unlearned Environments Using Unconstrained English

Initial Research Results

Robin Kowalchuk Burk
U.S. Military Academy

Contributors

U.S. Military Academy (Dept. of EE&CS):

Robin Burk

Frederick Moxley

MAJ Benjamin Ring

State University of NY (Albany):

Andrew Haas (Dept. of CS)

Kevin Knuth (Dept. of Physics)

Motivation

- Robotic systems such as UGVs will play a key role in Network Centric Warfare
- How best to command them?
- How best to network them?
- **How can we predict their impact? (network science question)**

Agenda

- Situational Awareness and Sensemaking: key activities
- Sensemaking in Autonomous Vehicles
- Unconstrained English Navigation Commands for UGVs
  - Why it's valuable
  - Challenges to implementing it
  - Research results to date
  - Next steps

U.S. Army War College case study:

OIF = nascent Network Centric Warfare

- *Extended reach communications*
- *Integrated information flows*
- *Synchronized joint fires*
- *Networked sensors & platforms (UAVs)*
- *Common operational picture for commanders*

Situational Awareness & Sensemaking

- Shared situational awareness in OIF
  - Real-time sensor data (incl. from UAVs)
  - Integrated into common operational picture
- Sensemaking in OIF
  - Implications of situation
  - Validated through verbal communications
- Outcomes: agility, people-centric

Situational Awareness & Sensemaking

- Integrated information flows facilitate shared situational awareness
- Shared situational awareness benefits
  - Sensemaking
  - Ability to substitute information and material for personnel
  - **Ability to substitute unmanned systems for personnel under some conditions**

Sensemaking in UGVs

- Future Combat Systems UGV roles
  - Driverless trucks
  - Robotic mules (soldier, squad aid)
  - Intelligent munitions
  - And more!
- Some degree of autonomy required for these roles
- **Autonomy requires sensemaking, not just information exchange**.

Natural Language for UGVs

- Ideal UGV would
  - Be **easy to integrate** into operations
  - Require **minimal operator training**
  - **Resemble a good soldier**
    - › Accept commands
    - › Interpret them intelligently
    - › Execute them reliably
    - › Ask questions when something is unclear
    - › Alert when significant events occur

Natural Language for UGVs

- NL communicates **intent** (commands) and **the sense we make of situations** (features we key on)

- NL is the most natural way for humans to interact
  - Adaptable rich interface
  - Suited to a wide range of situations
  - **Language skills persist under stress**

Natural Language and UGVs

"Go down this road to the first cross-street. Turn left and go two blocks. Stop in front of the second building on the right.

Radio if you see any white Toyotas parked along the streets."

Challenges

- Natural language challenge
- Cognitive robotics challenge
- Integration with the Global Information Grid
  - Desireable to minimize load on GIG (autonomy for most decisions)
  - Desireable to contribute useful information to GIG (intelligence)

Sensemaking in UGVs

Autonomous UGVs must:

- *Fuse sensor inputs*
- *Recognize objects in the environment*
- *Interpret events in light of task*
- *Adjust task execution as required*

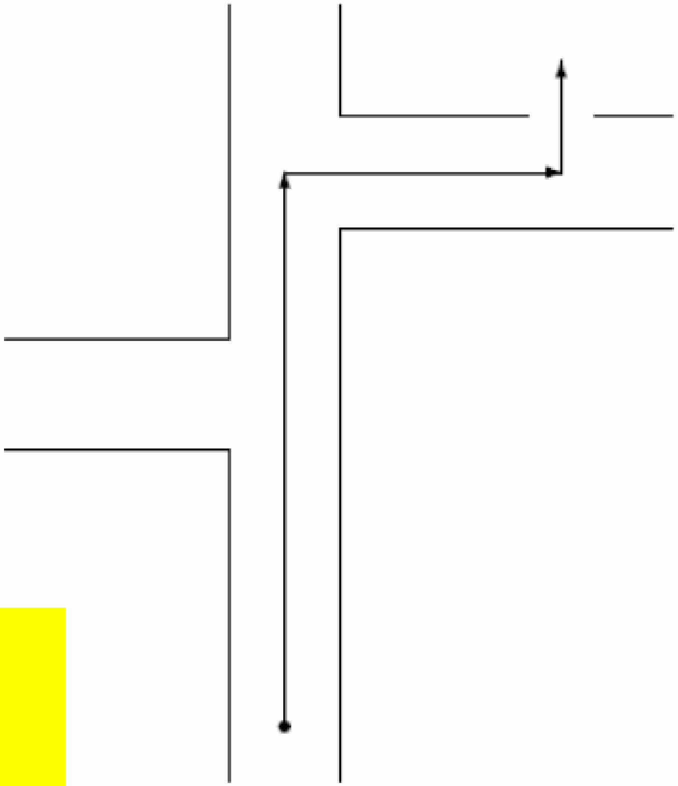
This is sensemaking!!!

Research Results to Date

- Executed in a simulated environment
- Approach = find minimum of NL features and interpretation needed to understand & execute navigational commands
- Results: 80% accuracy
  - Better than human interpretation in some experiments
  - Possible due to ***pragmatics*** approach, i.e. language as a tool to accomplish a task

Ambiguities in Language

Turn right at the ???? hallway



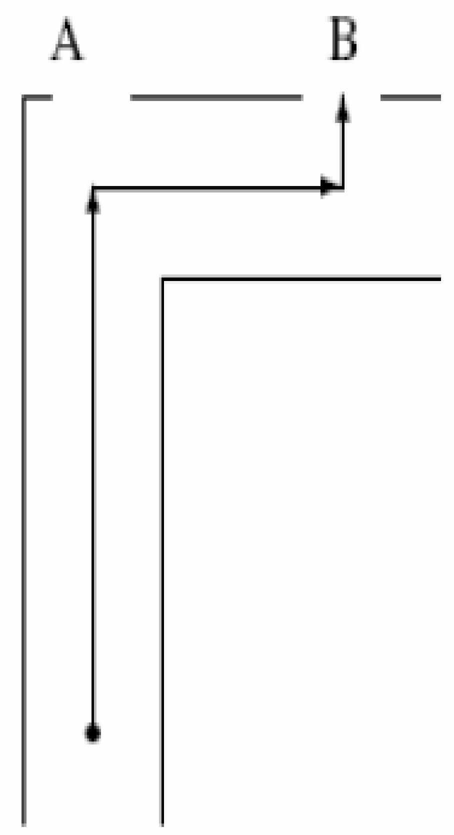
1st or 2nd?

Ambiguities in Language

To get to B:

Turn right and stop at the 1st door in the hallway?

Or the 2nd?



Cognitive Robotics

- Mapping perception (sensor input) to concepts (attributes described in language)
- Cognitive object recognition is key to a rich language capability in robots
- Probabilistic!!! (for both humans and robots)

Next Steps

- Replicate simulation experiments in physical robots
- Quantitatively characterize robustness and computational requirements in more complex environments
- Expand language interpretation strategies

Other Considerations

- Non-native English speakers
  - Cognitive structures differ between languages
  - Under stress, reversion to native structures
  - Clarifying questions can bring learned (English) context back to the foreground

- Gender differences in spatial perception and language

Summary

- Sensemaking will be a valuable attribute of unmanned systems in NC
  - Not the same thing as shared situational awareness / information sharing
- Sensemaking in unmanned systems requires many of the same abilities as natural language processing
- Natural language for commanding UGVs would be valuable and may be possible

Questions?