From soup to nuts: Using AI and robotics to improve Human Robot Interaction

Greg Trafton
Ed Lawson
Anthony Harrison
Laura Hiatt
Cody Narber
Alan Schultz

From spark to fire: Building interactive systems, even when people think they aren't needed

Greg Trafton
Ed Lawson
Anthony Harrison
Laura Hiatt
Cody Narber
Alan Schultz

Research in AI

- What the heck do researchers in AI do?
- Most of the time, we need funding to work on a specific problem.
  - Sometimes we have a lot of control over what kind of problem we work on, sometimes the funding has certain constraints
- This is going to be a story talk: Our path through though those constraints
- Our target problem: shipboard firefighting

US Navy shipboard fires

- *USS Conyngham* (DDG 17) May 8, 1990 suffers major fuel leak and fire in forward fire room. Fire burned for 23 hours. *Conyngham* decommissioned 6 months after incident.

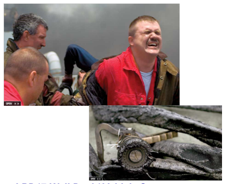
USS *Conyngham*

- *USS George Washington* (CVN 73) May 22, 2008 suffers fire in auxiliary boiler exhaust and supply space. Fire burned for 12 hours. Fire spreads to 80 compartments involving 9 decks. Cost of repairs $ 70 million.

USS *George Washington*

Shipboard fires are difficult to control and difficult to put out
(Smoke, limited access, extreme heat, …)

Current Fire Fighting Issues



LPD17 Well Deck/ Vehicle Stowage area firefighting test

- Fire fighting puts crew at risk for injury
- Optimal manning increases difficulty of fire fighting
- Existing suppression systems on ships
  - Difficulty extinguishing shielded fires
  - Require extensive piping increasing weight
  - Do not efficiently use water to suppress fire

Heat and Humans are incompatible!

- Experienced firefighter can actively work in the fire environment for about **15 minutes**.
- Inexperienced firefighters may be adverse to entering compartment.
- Steam generation from fire hose discharges significantly impacts firefighter

Human Heat Tolerances

Hot Air Exposure	
Temperature	**Effects**
90 °C (200 °F)	Incapacitation in 35 minutes, death in 60 minutes
150 °C (300 °F)	Incapacitation in 5 minutes, death in 30 minutes
190 °C (380 °F)	Immediate incapacitation, death in 15 minutes
200 °C (400 °F)	Irreversible respiratory tract damage

Firefighter Heat Tolerances

Firefighters Experience	Effects on Firefighter in Full Ensembles	
	Successful Entry	**Working Duration**
High	Up to 200 °C	10 minutes maximum
High	Up to 150 °C	15 minutes maximum
Low	Below 150 °C	15 minutes maximum

Shipboard Firefighting

- There are many possible solutions to this problem:
  - Better maintenance / inspection [but pretty good now]
  - Better sensors for quicker identification of fires [exist now]
  - Better fire suppression systems [current research]
  - Better protection for firefighters [current research]
  - …
- Researchers are actively working on all these problems!
- We were asked to work on autonomy, AI, and human robot interaction for a humanoid robot

Our constraints

- Robot constraints:
  - The system must eventually be able to work in shipboard environments, while being able to work in proximity to other sailors
  - Today's robots (and AI systems) probably can not currently do the entire job by themselves, so they will need human partner(s) and interaction
  - These robots should have "Natural interaction" — don't force sailors to learn a new set of commands to work with the robot. Rather, sailors should work with the robot the same way that they would work with other sailors

Natural Interaction

- "Natural interaction" — don't force sailors to learn a new set of commands to work with the robot. Rather, sailors should work with robot the same way that they work with other sailors
- This was one of our first challenges:
  - Why can't the robot just do everything?
    - Maybe eventually, but current robots don't have the capabilities (perception, manipulation, understanding, intelligence, experience, language understanding)
  - Why can't the robots be tele-oped?
    - Limited field of view, slow response, potentially 2+ operators per robot, comms issues, …
  - What interaction? There is almost no interaction or even communication after firefighting starts!
    - Multiple expert firefighters said that almost all the communication occurred during the preparation stage of firefighting, but once firefighting started there was very little communication and the nozzleman just put out the fire. So really, the robot just needed the ability to perceive the fire, have some high level instructions that could be given during the preparation stage (e.g., prioritize the fire closest to the door). Then just let the robot go!

Natural Interaction (?)

- Not surprisingly, we were frustrated by the experts' message (and they were top notch experts!)
- Perhaps we should just work on communication during preparation?

Progress needed!

- This project (like many others), needed results fast! So we needed to show a demonstration of a robot putting out a fire with a strong interaction component
- We went to observe a team of 2 firefighters and informally noticed that there were many redundant cues (e.g., "Put out the fire in the center of the floor [deictic gesture]")
  - Our interaction would thus emphasize redundant cues
  - Our interaction would occur primarily during the preparation stage

System needs

- Through a combination of talking to the experts, task analysis, and trial-and-error, we developed a set of functions and needs that our robot would need to have

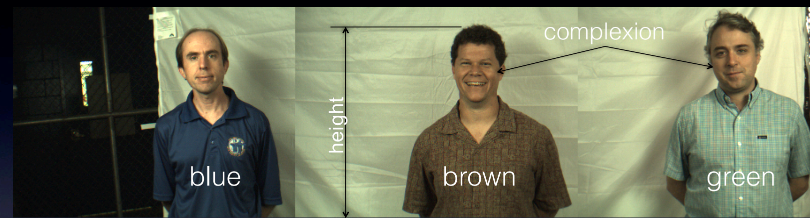
System needs

- Identify people (with understanding of who they are, their preferences, language models, gesture models, etc.)
- Communicate and understand via language
- Communicate and understand via gesture
  - Deictic gestures
  - Iconic gestures (follow me)
  - Symbolic gestures (stop)
- Track and follow people
- Localize the fire
- Suppress the fire
- (Too many systems to go into lots of details on each one, so I'm only going to describe Person Identification here…)

Identifying people

- Identify people
  - Current person identification is quite good! But it focuses on face identification and over-relies on good lighting, good cameras, and lots of training (for some)
  - Because firefighting frequently has poor lighting, hidden faces, noisy cameras, and little training, we took a different approach: Soft biometrics
  - (Details in Martinson, Lawson, & Trafton, 2013)

Soft Biometrics



Soft Biometrics : Lack permanence and distinctiveness for unique identification, but include intuitive features useful for recognition

- When combined they form a powerful way to recognize individuals
- We used 3 very simple modalities: complexion, clothing color, height
- Because each modality has a different failure mode (e.g., height is not a good identifier when someone is sitting down), they can be jointly robust in different environments/situations

Person Identification

- Requires us to be able to
  - Find people
  - Track people

Enrollment

- Find and track a person
- Extract different modalities
  - Complextion
  - Clothing
  - Height
- Build a model of that individual for each modality

Identification

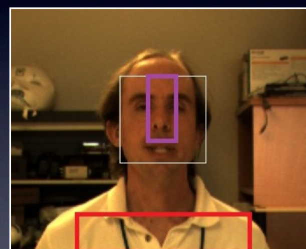
- Find and track a person
- Extract as many modalities as you can
- Compute similarity of extracted modalities to enrollment modalities
- Fuse data and identify

Person Detection and Tracking

- We detect individuals in the environment using an SR4000 IR time of flight camera
  - Connected components are found using depth measurements from the camera
- Silhouettes are tracked in subsequent frames using the Tanimoto distance, which compares silhouette shape similarity

Modeling Complexion and Clothing

- Sampled regions for complexion and shirt are determined relative to the location of the detected face
  - Extract color from the face (purple box) and shirt (red box)
  - Model color using a 4096-bin color histogram
- During enrollment multiple histograms can be associated with a person
  - A new histogram is added to the models if the distance to existing histograms exceeds a threshold



Identification: Color Similarity

- Given a color histogram for a person we want to identify, we can compare the model against the (as yet) unknown person using chi squared distance

$$\chi^2(H_a, H_b) = 2\sum_i \frac{(H_a(i) - H_b(b))^2}{(H_a(i) + H_b(i))}$$

- When multiple models are associated with the same individual, the approach returns the smallest distance to any model

Height

- Height is determined based on the relative position to the robot
  - Height is computed from the center of the detected face (x,y,z) in the ToF intensity image
  - This is converted to height using the joint angles of the head, neck and eyes of the robot
- Height is estimated using a running average of the last 20 detected face locations, providing robustness to individual motion

$$T_{cam2robot} = T_{torsopan} * T_{neckpitch} * T_{headpan} * \ldots$$
$$\ldots * T_{headpitch} * T_{headroll}$$
$$Face(x,y,z) = T_{cam2robot} * [x,y,z,1]^T$$

Identification: Height Similarity

- Given two height measurements (m1 and m2), the similarity is

$$S(m_1,m_2) = \begin{cases} 1 - \frac{|m_1 - m_2|}{0.15}, & |m_1 - m_2| < 0.15 \\ 0, & |m_1 - m_2| \geq 0.15 \end{cases}$$

- A maximum difference of 15cm, with a linear decrease in similarity to that point

Multimodal fusion

- Scores are normalized to sum to 1.0, then the weighted sum is computed, followed by division by the number of *reporting modalities*
- Weighting of Modalities
  - We can learn weights over time using simple statistical methods
- If a modality was unavailable, (e.g., clothing was not visible due to the robot position) the weight of that modality is set to 0

Evaluation?

- Our system seemed to work quite well in the lab
  - (With people who knew to look at the robot, didn't move, and where the lighting was excellent)
  - TPR = .96
- In more difficult environments the system also worked quite well (TPR = .90)

Language

- Communicate and understand via language
  - We used off-the shelf NL system (Berkeley parser)
  - We enhanced it with spatial referencing (exocentric and egocentric referencing capabilities)
  - Integrated language with fire detection doctrine (e.g., fire is more likely to be at a gas juncture than on a metal bunk)
  - We used simple speech output and structured outputs

Gesture

- Communicate and understand via gesture
  - We used an in-house gesture recognition system using 3D point cloud pattern recognition to find arms and motion history images for periodic gestures (Lawson & Martinson)
  - Simplified problem because we only needed to recognize a relatively small subset of gestures
  - Redundant cues with language
  - Made more complicated by varying distances and differences by individuals (e.g., one handed vs two handed "follow")

	Stop	Wave	Follow	Alert	Rally
Stop	18	0	0	0	0
Wave	0	18	0	0	0
Follow	0	0	18	0	0
Alert	0	0	1	17	0
Rally	0	0	0	0	18

Table 1. Closed-world gesture recognition at various distances.

Person Tracking

- Tracking and following people
  - We used a time of flight camera to extract depth discontinuities and then performed shape similarity to track over time
  - Worked well in cluttered environments
  - Worked well even when part of the person was occluded

Localizing and Suppressing the fire

- Localizing and suppressing the fire
  - When we started this project, all of the traditional fire detection sensors and algorithms assumed a static sensor and used background subtraction, so they would not work on a mobile robot
  - Thermal sensors work, of course, but at the time there were proprietary concerns with some of the sensors
  - So we used two IR cameras to triangulate the location of the fire (simple and robust approach for simple situations)
  - We used a traditional Macaw fire extinguisher that sprayed water at the hottest locations in the room
    - The Macaw is used by human firefighters today

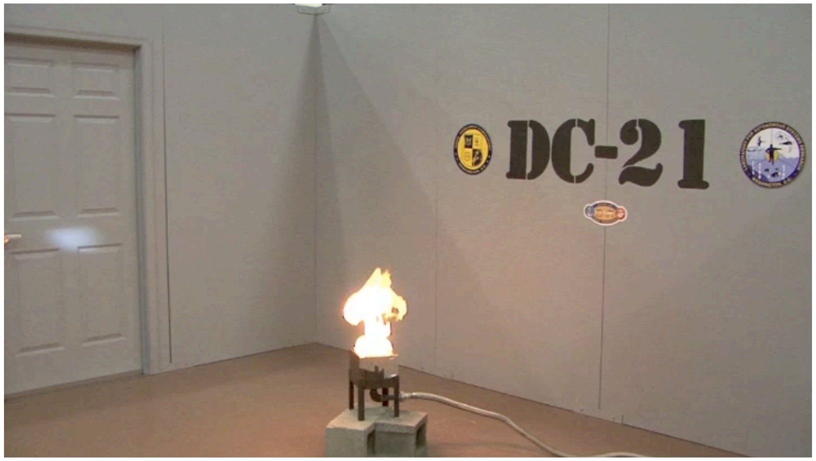
Where to fight fires?

- We also needed an environment where we could safely light fires of moderate size and extinguish them
  - The environment needed to bear some resemblance to a ship environment as well
  - Safety was a substantial concern here

Firefighting Facility

- Built replica of ship passageway and compartments in the prototyping high bay of our Laboratory for Autonomous Systems Research facility
- Can set fires via propane burner (~4')

Putting it all together …



Testing through demos

- System run many times for demos; strict end-to-end success rate was over 80% [success = putting out fire after interaction]

Success!

- Within 16 months, we were able to build a working, interactive firefighting system and demo it to multiple groups

Success?

- Within 16 months, we were able to build a working, interactive firefighting system and demo it to multiple groups
- My group's main focus, however, was interaction
  - We had a set of relatively cheesy, easy-to-identify gestures with a closed-world assumption and an expert operator
  - NL was bare bones
  - (BUT both worked!)
- We had no real understanding of how firefighters really interact with each other
  - What kind of autonomy did the nozzleman (robot) actually have?
  - How did the team leader really give instructions?
  - Was the only place for interaction before firefighting started, or did it occur during the firefight itself?

Interaction mandate with no interaction?

- What interaction? There is almost no interaction or even communication after firefighting starts!
  - Multiple expert firefighters said that almost all the communication occurred during the preparation stage of firefighting, **but once firefighting started there was very little communication and the nozzleman just put out the fire**. So really, the robot just needed the ability to perceive the fire, aim and use the macaw, have some high level instructions that could be given during the preparation stage (e.g., prioritize the fire closest to the door). Then just let the robot go!
- If experts said it, it must be true, right?

Listen to the experts?

- We should definitely listen to the experts!
- But …
- There are many reasons to suspect that experts are not perfect describers of tasks they perform (Ericsson & Simon, 1993; Nisbett & Wilson, 1977; Trickett & Trafton, 2009)
  - Retrospective memory is poor about details
  - Social expectation effects: How it "should" be done
  - Declarative / Procedural transition

Studying Interaction

- I'm a strong believer that in the interaction sciences, we need to understand people so that we can build successful models of people (though sometimes it happens the other way around)
- So, we collected *in vivo* (naturalistic) data of Navy firefighters putting out fires
- I'm going to show highlights here (not science), but we have full data (science)

Ex-USS Shadwell, Mobile AL

Two primary findings from empirical work

1. Did nozzlemen make their own decisions with high level guidance issued before firefighting, or did the team-leader provide a lot of instructions?
   - Nozzlemen made 25% of the decisions of what to do next; team-leaders issued instructions 75% of the time (contrary to what the experts told us). **Interaction is key**
2. How did team-leaders issue instructions while fire-fighting?
   - Speech: 10 utterances / minute
   - Gesture: .2 gestures / minute
   - **Touch: 2.5 touches / minute** [a surprise]
     - Touch information is primarily spatial and adds information not easily available in speech (e.g., metric information about exactly what fire to work on)

"Come now"

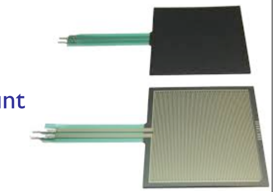
Touch

- Touch gestures are relatively frequent and incredibly important given the limited visibility
- Most current robots can not sense or understand touch made by an interaction partner
- How can we get a robot to understand some of these touch gestures?

Touch Sensing

- Most HRI work on tactile sensing falls into 3 categories (Argall & Billard, 2010):
  - Hard shell
  - Soft skin
  - Joint sensors
- Tactile sensing focuses on touch itself, but understanding what to do with touch can be slow on today's robots (and most of today's robots are slow)
- Our approach is to fuse a tactile sensor with a 3D hand sensor

Touch sensing on a robot



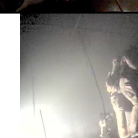
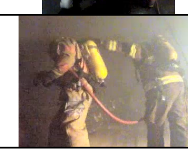
- Multi-sensing approach:
  - Force sensor resistors (Phidgets)
    - Very sensitive, can sense even a small amount of force (up to 20N)
    - They respond very quickly, and have stable measurements (3 microsecond response rate)
    - They can detect approximate location and amount of force, but no orientation
  - 3D hand sensor with our internal SDK
    - Can work with gloves
    - Goes beyond hand velocity and orientation by identifying hand features
    - Features are integrated into a neural network classifier

Touch Sensing on Octavia

- 21 FSRs are mounted along the back, side, and front
- The FSRs are covered with foam to diffuse the force
- These are combined with 3 Leap Motion sensors which can detect hands along the back or sides

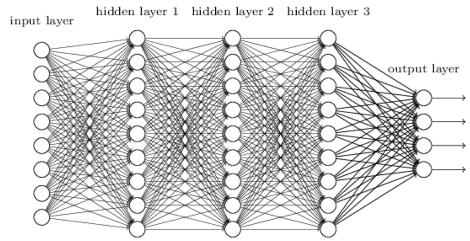
Data collection

- We asked 8 participants to give the robot different instructions based on the gestures the firefighters used
- 100 gestures of each class were collected
- We got 1 and 2 handed gestures, differing levels of force, different hand orientations…

Engage/ Advance	
Turn Left/ Right	
Adjust nozzle	
Exit	

Classifying gestures

- We built two simple deep neural networks to classify gesture based on:
  - 3D sensor (Yaw, Pitch, Roll, hand splay, orientation, velocity, acceleration)
  - 3D sensor + FSR (grid representation of force in N)
- Results were evaluated by using 10-fold cross validation



Recognition Accuracy

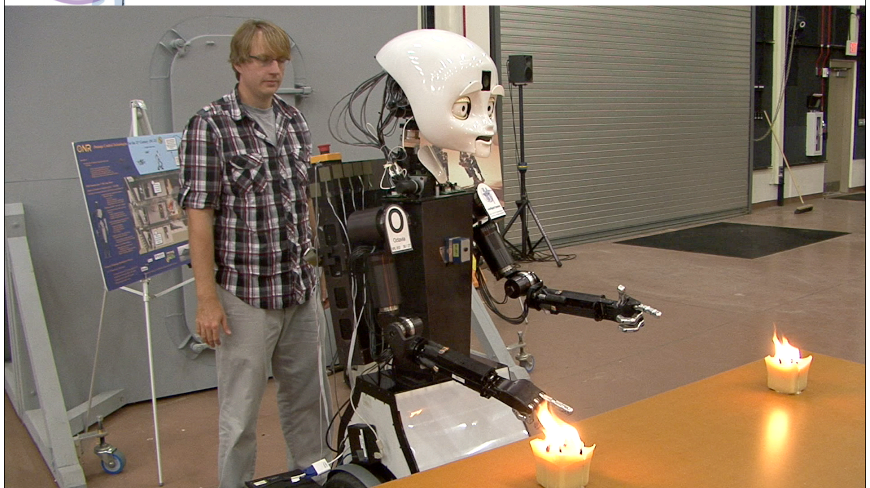
Gesture	True Positive Rate (3D)	True Positive Rate (3D + FSR)
Engage	96%	99%
Turn Left	91%	99%
Turn Right	91%	99%
Exit	93%	99%

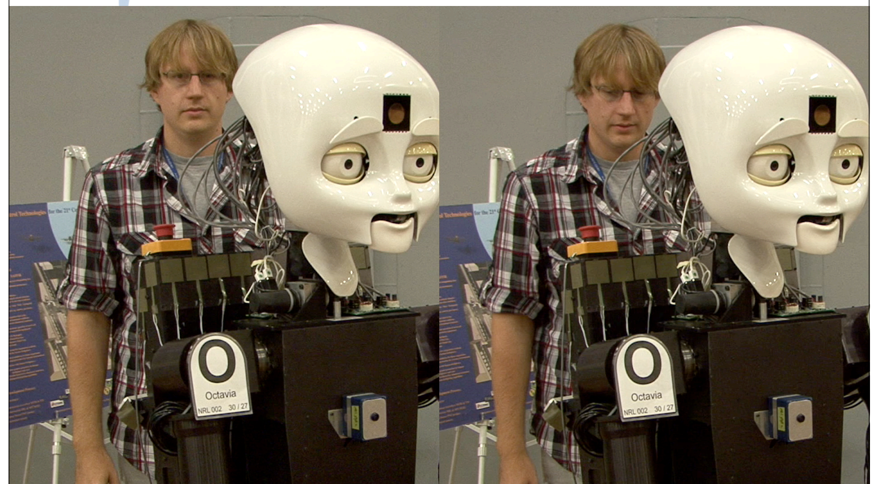
- Our recognition rate is quite good using both sensors
  - 93% (3D alone) vs 99% (3D + force sensor)
- Can we recognize a touch gesture before it occurs?

- Robots are generally slow; if we can use the 3D sensor to predict what the gesture will be *before* the person actually touches the robot (and verify with the FSR), we may be able to speed up the robot
  - A successful prediction should allow us to start the IK system and speed up the overall system
- The current system predicts the gesture using the 3D sensor, starts the IK and then verifies the gesture with the FSRs

Firefighting information

- Sometimes the team leader will have additional information that requires a refocus of the nozzleman's efforts (e.g., a smaller fire is close to a gas leak and has a higher priority than the current fire)
- The robot must have some autonomy (able to see the fire and spray it) but be able to take commands from the team leader
- Octavia can be re-directed to different fires by touch
- Fire detection is done by the IR camera
- We point at the fire here instead of using the full Macaw





- Predicting what gesture will occur before the physical touch saves 200-500ms (especially important in these safety-critical domains)
- After using the anticipatory system, the non-anticipatory system seems slow and clunky

Summary

- Our first firefighting demonstration was a great success and pinpointed many things that needed to be worked on
- Collecting empirical data and performing in-depth analyses allowed us to:
  - Show the importance of interaction during the firefighting phase (and in contrast to what others thought)
  - Make a surprise discovery about the importance of touch during these types of tasks (Many people later came through and talked to us about how important touch was in other domains)
- Predicting human behavior to improve human-robot interaction can improve overall system performance